QUANTUM

JANUARY/FEBRUARY 1999

US $7.50/CAN $10.50

Springer

NSTA

GALLERY Q

Oil on linen, 42 3/4 × 58 1/4 inches, Widener Collection. © 1998 Board of Trustees, National Gallery of Art, Washington

Work (1863) by Pierre Puvis de Chavannes

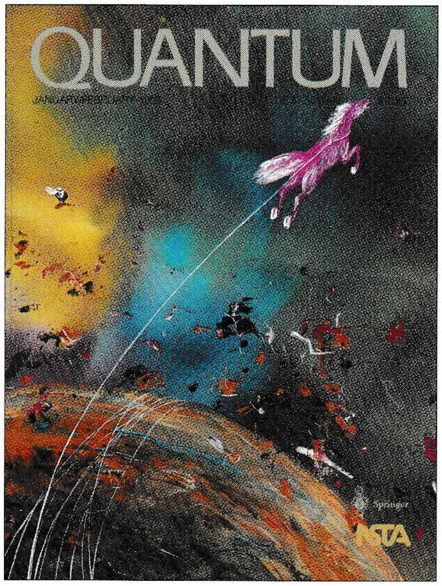
THE CHOREOGRAPHED CLANG OF HAMMERS ON THE anvil, the dull thud of axes sinking into wood—these striking images depict events that, when simplified to their fundamental elements, can be regarded as interactions between two bodies. In physics we know such interactions as collisions, a subject of signal importance in the field.

The "clang" and "thud" of the hammer and the axe present auditory evidence that the character of collisions can vary to a large degree. These variations naturally lead us into the realm of elastic and inelastic collisions.

In this issue's "Collide-o-scope," we examine the forces that govern the chance meetings of everything from atoms to zeppelins. Due to space limitations, we are forced to neglect such famous collisions as the *Titanic* and the iceberg, Comet Shoemaker-Levy 9 and Jupiter, and the heads of The Three Stooges. However, our historical overview does discuss the impact of Galileo's theories on single-body systems as well as Sir James Chadwick's work with particle collisions. Turn to page 28 to begin working your way through the world of collisions.

QUANTUM

JANUARY/FEBRUARY 1999 VOLUME 9, NUMBER 3



Cover art by Pavel Chernusky

Just hold your horses! Before you make any assumptoins about this month's cover, you need to know that it deals with the satellite aerodynamic paradox. How is it possible for a body to enter the rarefied gas of our upper atmosphere and actually speed up rather than slow down?

Being of stable mind, we're sure you'll enjoy the challenge presented by this paradox. Rein in your preconceptions and turn to page 18 for a closer look at this orbital anomaly.

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

FEATURES

DEPARTMENTS

Quantum (ISSN 1048-8820) is published bimonthly by the National Science Teachers Association in cooperation with Springer-Verlag New York, Inc. Volume 9 (6 issues) will be published in 1998–1999. *Quantum* contains authorized English-language translations from *Kvant*, a physics and mathematics magazine published by Quantum Bureau (Moscow, Russia), as well as original material in English. **Editorial offices:** NSTA, 1840 Wilson Boulevard, Arlington VA 22201-3000; telephone: (703) 243-7100; electronic mail: quantum@nsta.org. **Production offices:** Springer-Verlag New York, Inc., 175 Fifth Avenue, New York NY 10010-7858.

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

Subscription Information:

North America: Student rate: $18; Personal rate (nonstudent): $25. This rate is available to individual subscribers for personal use only from Springer-Verlag New York, Inc., when paid by personal check or charge. Subscriptions are entered with prepayment only. Institutional rate: $47. Single Issue Price: $7.50. Rates include postage and handling. (Canadian customers please add 7% GST to subscription price. Springer-Verlag GST registration number is 123394918.) Subscriptions begin with next published issue (backstarts may be requested). Bulk rates for students are available. Mail order and payment to: Springer-Verlag New York, Inc., Journal Fulfillment Services Department, PO Box 2485, Secaucus, NJ 07094-2485, USA. Telephone: 1(800) SPRINGER; fax: (201) 348-4505; e-mail: custserv@springer-ny.com.

Outside North America: Personal rate: Please contact Springer-Verlag Berlin at subscriptions@springer.de. Institutional rate is US$57; airmail delivery is US$18 additional (all rates calculated in DM at the exchange rate current at the time of purchase). SAL (Surface Airmail Listed) is mandatory for Japan, India, Australia, and New Zealand. Customers should ask for the appropriate price list. Orders may be placed through your bookseller or directly through Springer-Verlag, Postfach 31 13 40, D-10643 Berlin, Germany.

Advertising:

Representatives: (Washington) Paul Kuntzler (703) 243-7100; (New York) Jay Feinman (212) 460-1682; and G. Probst, Springer-Verlag GmbH & Co. KG, D-14191 Berlin, Germany, telephone 49 (0) 30-827 87-0, telex 185 411.

Printed on acid-free paper.

Visit us on the internet at www.nsta.org/quantum/.

QUANTUM

THE MAGAZINE OF MATH AND SCIENCE

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

The mission of the National Science Teachers Association is to promote excellence and innovation in science teaching and learning for all.

BRAINTEASERS

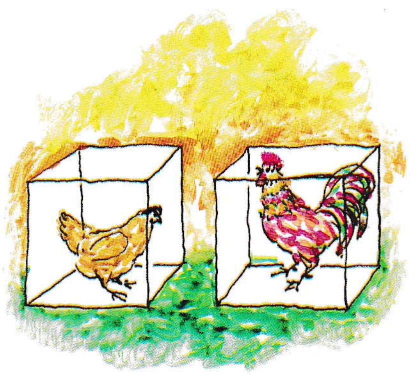
Just for the fun of it!

B251

Sequential thinking. The sequence a_n is defined as follows:

$$a_1 = 1776, a_2 = 1999, \ldots, a_{n+2} = \frac{a_{n+1} + 1}{a_n}.$$

Find a_{2002}. (A. P. Savin)

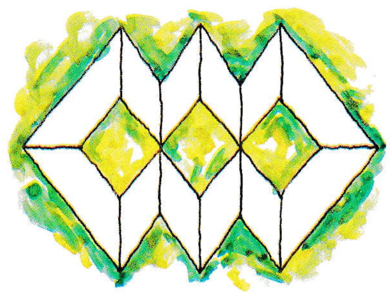


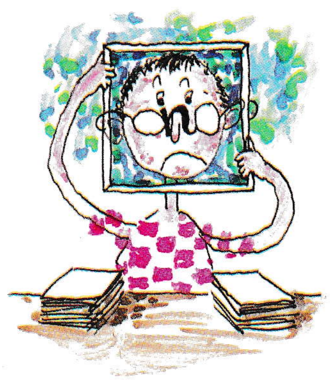
B252

Cubic coincidence. Five vertices on each of two equal cubes are colored black, and the remaining vertices are colored white. Prove that the cubes can be superimposed in such a way that at least four of the black vertices coincide. (S. G. Volchenkov)

B253

Walk this way! Twenty-two towns are connected by roads as shown in the figure at right (the towns are located at the intersections of the roads). Is it possible to walk to every town, yet visit each of them only once? (S. G. Volchenkov)





B254

Short stack. A total of n cards numbered 1 through n is divided into two stacks. What is the minimum value of n such that at least one stack will include a pair of cards whose numbers add up to an exact square?

B255

Water in a paper box. Water is poured into a paper box, which is then placed over the flame of a candle. The water boils but the paper doesn't burn. Why?

Art by Pavel Chernusky

ANSWERS, HINTS & SOLUTIONS ON PAGE 50

The anthropic principle

A key to deciphering the Universe?

by A. Kuzin

THE DISCOVERY OF THE VASTness of the Universe has led to a fundamental problem: Does a human being mean anything in this immense Universe? What is a human being—the aim of the Universe's development, the crown of creation, or one of nature's negligible byproducts that isn't much different from the other creatures and processes in the Universe?

Is the appearance of the human creature the culmination of the macro- and microcosm, or just a whimsical turn in the development of the Universe? Has biological life adapted itself to surrounding conditions that "knew" nothing about it, or were these external conditions "tuned" to make life possible?

For a long time science had no data to consider this problem rationally (although there were plenty of irrational and mystical answers). The available scientific knowledge about the Universe was too fragmentary to provide an integrated picture of the Universe's development. However, the organized efforts of science during almost four centuries has yielded some fruits. Now we have a scientific vista from which to form an integrated view of the Universe.

Of course, how little we know! Therefore, modern hypotheses should not be considered as the final verdict, which "sentences" humans to be the crown of creation. The skeptic can always justifiably point to the fact that scientific knowledge is limited. The topics covered here are among those "eternal questions" that will never be answered once and for all.

Stating the problem

The anthropic principle is the child of a mental experiment. In this experiment we assume some change in the natural laws and then see whether or not a human could exist in the modified world. Of course, we should consider the very essence of a human being, because we cannot expect a human to preserve all the same features in the changed surroundings. There are, perhaps, two basic qualities inherent to human beings—intelligence and freedom.

Freedom is the ability to preserve one's ego, and not to depend entirely on what is going on at a particular moment in the surrounding world. In other words, it is the ability to act according to inner motivation.

Intelligence is a prerequisite condition of freedom, because an active response to the outside world is impossible without it (in this sense animals also have intelligence). At present, the natural sciences are not ready to treat the problems of intelligence and freedom in general. We must find the starting rung of the ladder to make the first steps in solving these problems.

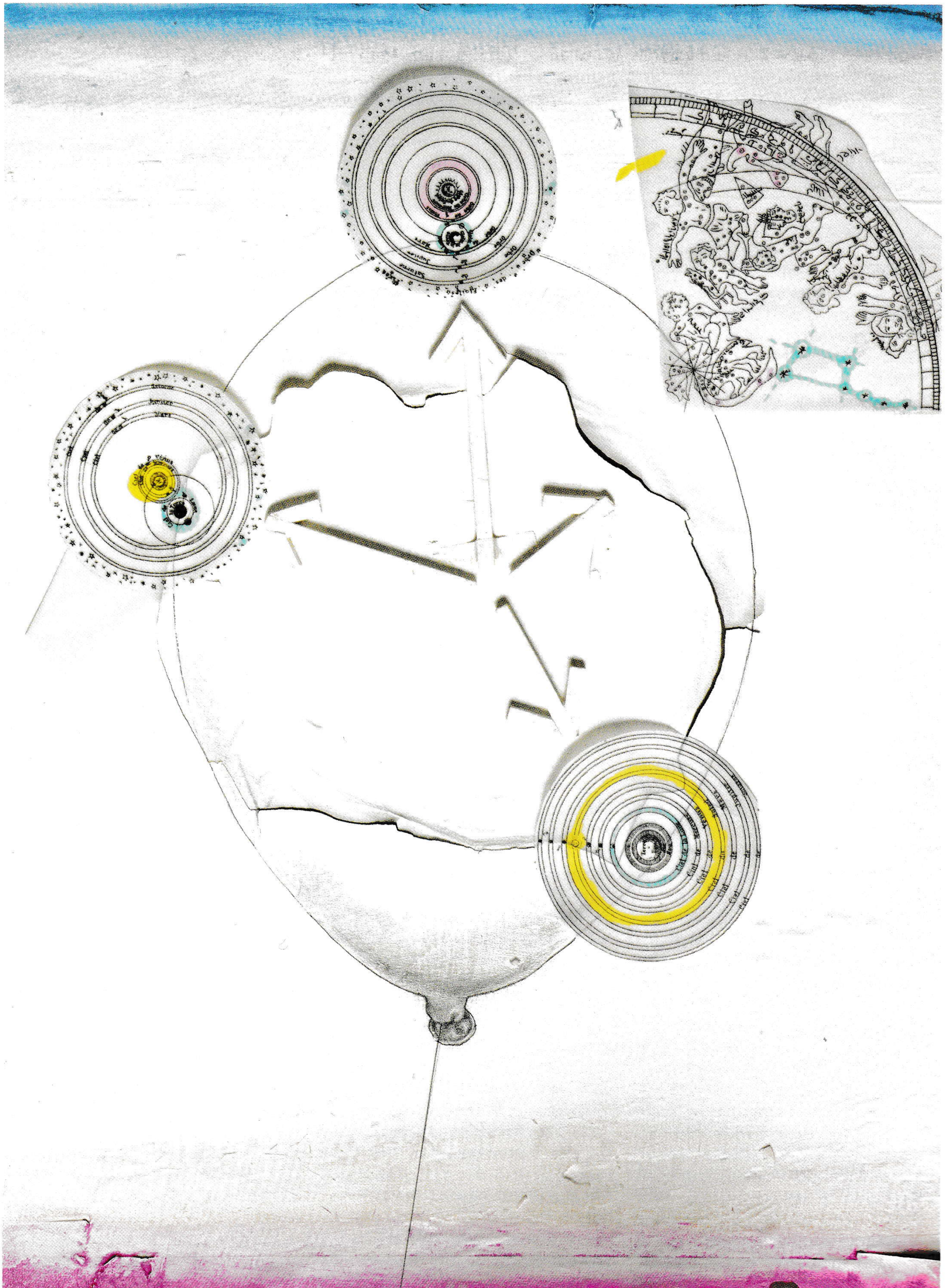
Only a complicated creature can actively withstand its surroundings by understanding and modifying them according to the needs of its own existence. The degree of complexity can be qualitatively described—and this is what the natural sciences deal with.

The anthropic principle provides an answer to the question, Does any conceivable world order imply the appearance of more and more complex structures? It should be noted that the appearance of such structures is only a necessary, but not the sufficient condition to create a human being. However, even an analysis of only this condition yields many fruits. Thus, let's consider the evolution of the Universe from this viewpoint.

Evolution of the universe

According to modern views, the Universe is limited in time and

Art by Yury Vashchenko



space. That is, its age is not infinite (about 15 billion years), and its volume is finite. The number of particles in the Universe is enormous, but it is also a finite value: $N \cong 10^{80}$. The Universe "started" from an unimaginable compressed and hot state, confined in a volume with a radius of curvature of about 10^{-34} cm. From that time (known as the Big Bang) on, the Universe continuously expanded like an inflating balloon while its galaxies, stars, nebulas, and other kinds of matter moved away from each other like drawings on the balloon's surface.

Matter, space, and time are interdependent; they were born simultaneously—so the questions, What was before the Universe? or What can one see at the edge of the Universe? are not correctly formulated. There was no time "before" the Big Bang, and it is impossible to reach the limits of the Universe in the way that we can reach the end of Earth's surface.

In the first tiny moments after the "start," the temperature was so huge that no stable structure could be formed. Even the elementary particles were continuously converted into each other. However, the expansion of matter was accompanied by a decrease in temperature, and at some moment stable particles were formed—electrons, protons, and neutrons. They were the first objects with structure, and they point to the first enigma in the life of the Universe.

The primary boiling pot contained equal numbers of particles and antiparticles, but the symmetry was broken for some unknown reason during cooling. Therefore, the number of particles N_p exceeded the number of antiparticles N_a. As a result, the annihilation was not complete—not all of the matter was converted to light. The scale of this process is a miracle in itself: $(N_p - N_a)/N_p \cong 10^{-9}$, which means that only one billionth of all the particles was preserved to form the Universe (we will see the same number in another context).

What happened next? Further cooling yielded the generation of the simplest elements, hydrogen and helium, which need a stable proton for their existence. The masses of a proton m_p and that of a neutron m_n are known to be almost equal: $m_n - m_p \cong 2.5m_e$, where m_e is the electron's mass. The heavier neutron lives only 16 minutes on the average and disintegrates into a proton, an electron, and an electron antineutrino:

$$n \to p + e^- + \tilde{\nu}_e.$$

Inside nuclei, where the density of nuclear matter and the kinetic energies of the particles are large, this reaction also proceeds in the reverse direction.

Thus, the neutrons exist in a state of dynamic equilibrium inside the nuclei. Were the neutron lighter than the proton, the latter would disintegrate via the following reaction:

$$p \to e^+ + n + \nu_e,$$

where e^+ is a positron and ν_e is an electron neutrino. In this case, the protomatter would have no hydrogen, which is the main fuel of stars and the basic element of water, the cornerstone of life.

Let's assume that in one way or another the existence of atoms is guaranteed. Note that we consider only the two simplest atoms—hydrogen and helium. These atoms were synthesized everywhere almost immediately, yet the formation of heavier atoms was delayed, because it required special conditions.

Indeed, to produce heavier atoms, one needs the simplest stable nuclei as raw material and a very high temperature, at which these atoms can be fused. This high temperature must be maintained for billions of years, because thermonuclear fusion is quite a slow reaction. The Universe marched this part of its route too quickly: It needed not billions, but only hundreds of thousands of years to reach the more or less present status of low density and low temperature. Such conditions are favorable for the existence of atoms (at higher temperatures electrons could not be held by nuclei), but they are not suitable for synthesizing atoms.

Thus, the conditions for synthesis and stability of nuclei are incompatible. However, this incompatibility is not the last to be overcome. To take the next steps along the road of increasingly complicated structures, the Universe rejected the property of space homogeneity. This is the stage where gravitation was given the key role. In the gaseous Universe, condensed regions were formed—the seminal galaxies. In turn, they divided into even smaller parts—the protostars. Note that gravitation needed enough time to finish this work before the Universe expanded too much. While a protostar becomes more dense, its temperature grows until it ignites a thermonuclear reaction, which produces huge amounts of kinetic energy and stops further compression of the star.

It is noteworthy that the first stars were composed mostly of hydrogen (the hydrogen to helium ratio in the young Universe was about 3:1). This was possible because protons are lighter and more numerous than neutrons, so not all of the protons are used to form helium. In an alternative scenario stars are composed mostly of helium. Such helium stars would be too hot and they could not live long enough to support biological evolution on the planets.

In the interior of stars "alchemical" transmutations occur continuously: The nuclei of light elements collide, fuse together, and turn into nuclei of heavier elements. This process of increasing complexity is also based on rather fine tuning of the nuclear and electromagnetic forces. Without this tuning the reaction chain leading from helium via carbon and oxygen to iron and heavier elements would stop at the initial stages.

Here is a striking example illustrating this fact. At temperatures

$T \cong 10^8$ K, which characterize the stellar interior, helium is converted into carbon:

$$3\ ^4\text{He} \rightarrow {}^{12}\text{C} + 2\gamma \qquad (1)$$

The letter γ signifies a gamma ray. The probability of a triple collision of ^{4}He nuclei in the rarefied stellar plasma is very small (the reaction occurs in less than 10^{-21} s). Therefore, reaction (1) should proceed very slowly, that is, at a rate too slow to produce the amount of carbon necessary to make a planet like Earth.

However, reaction (1) also occurs via another route! It has two stages. First, two ^{4}He nuclei form one nucleus of ^{8}Be:

$$2\ ^4\text{He} + (99 \pm 6)\ \text{keV} \rightarrow {}^8\text{Be} \qquad (2)$$

The energy value in parentheses[1] means that this reaction needs energy. In other words, the nucleus of ^{8}Be is unstable and "wants" to disintegrate. If this nucleus were stable, the second stage of reaction (1) (which gives ^{12}C), would become less and less probable over time, because all the ^{4}He is increasingly depleted in the production of ^{8}Be. In reality, the unstable but still integrated nucleus ^{8}Be captures a nucleus of ^{4}He to produce an atom of carbon:

$$^8\text{Be} + {}^4\text{He} \rightarrow {}^{12}\text{C} + 2\gamma. \qquad (3)$$

The probability of the two-stage reaction (2)–(3) is larger by far than the triple collision of reaction (1), because the unstable ^{8}Be still "lives" 10,000 times longer than helium nuclei. But this is only one reason. Here we again meet a striking phenomenon. The rates of nuclear reactions do not vary monotonically with the energy E of the colliding particles. At some energies (E_1 and E_2 in fig. 1), a drastic increase in the reaction rate takes place. This phenomenon is called resonance, and the corresponding energy values are known as "resonant energies." These energies are determined entirely by the structure of the nucleus produced in the reaction. The critical reaction (3) is resonant: The resonance energy of a ^{12}C nucleus (7.656 ± 0.008 MeV) is almost the same (and a little bit larger) than the sum of the energies of ^{8}Be and ^{4}He nuclei (7.3667 MeV). The necessary energy is supplied by the high temperatures deep within stars.

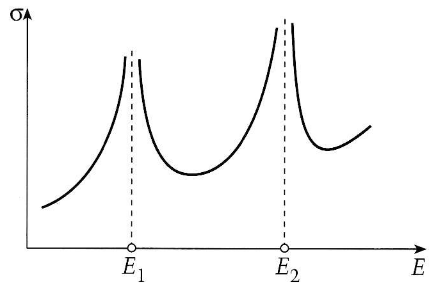


Figure 1

However, can the carbon be burnt quickly? Indeed, there is such a reaction:

$$^{12}\text{C} + {}^4\text{He} \rightarrow {}^{16}\text{O}. \qquad (4)$$

No, it will not occur quickly! Reaction (4) is nonresonant, so it occurs very slowly (the resonance energy of an ^{16}O nucleus is 7.1187 MeV, which is less than the sum of the rest energies of ^{16}C and ^{4}He nuclei: 7.1616 MeV). The high temperature increases this sum and thus aggravates the "mismatch" of the energy values and detunes the resonance.

Now we see that due to a long chain of "coincidences"—the instability of the ^{8}Be nucleus together with its comparatively long lifetime, the resonant character of reaction (3), and the nonresonant nature of its sibling (4)—the reaction chain is not stopped, and much carbon is produced, which is so important for life in the Universe.

When the nucleus of a star is enriched enough by the heavy elements, the star explodes with a drastic increase in luminosity (it becomes a supernova). In doing so, it jettisons some of its mass. New stars are formed from these remains; these stars are even more enriched with heavy elements: and so it goes on and on.

Cosmologists have found that our Sun is fourth-generation star. As the age of the Universe is about 15 billion years, it appears that the mean lifetime of a stellar generation is about 3 billion years. However, this period is far from being sufficient for biological evolution, because the age of Earth is about 5 billion years. This means that our Sun belongs to a rather rare type of comparatively long-lived stars. Since most stars don't live as long (the lifetime of a star depends on its mass), the "meaning of their existence" lies in providing stars like our Sun with heavy elements.

The ancients believed in the harmony of the celestial spheres, and they considered the cosmological processes to be interconnected and understandable by a philosophical mind, to which they sound like an orchestra. The life of a star is similar to a concert by an orchestra whose members are the forces of Nature, each with its own instrument and part. The cello (nuclear forces) and the violin (electromagnetic forces) lead the entire performance; a flute (the strong force) squeaks sometimes, while the contrabass (gravitation) plays in the background, determining the rhythm. The concert is finished by the weak interactions. Putting aside the flute, they take French horn and play solo: This is the rushing flow of neutrinos from the nucleus of a supernova, which carries away its outer mantle.

Thus, the Mendeleev's table is "produced." The development of complexity proceeded with the formation of molecules, and the infinite character of its process is underlain by the existence of a wonderful element, carbon, which can form molecular chains of immense length, because it conforms ideally to this task.[2]

Carbon has four valence bonds with angles between them of about 90°. Linking to each other, the carbon atoms form something like a line while two free valences per carbon atom bind other atoms and molecules. In this way something like an inscription is formed. Before the

[1]In nuclear physics energy is measured in electron-volts: 1 eV = $1.6 \cdot 10^{-19}$ J.

[2]In this short paper I cannot discuss the problems why other quadrivalent elements (silicon, for example) are not quite suitable for this task.

formation of carbon chains, the entire cosmological process produced letters. When organic molecules were formed, the letters (atoms) began to unite to form words (organic molecules). However, arbitrary words do not yield a text. A random set of words (molecules) and a reasonable text of a living creature are separated by an immense gap. Before touching the problem of life, some aspects of inorganic evolution should be mentioned.

The further the process of increasing structural complexity proceeds, the more restrictive are the conditions imposed on this process. Nuclei do not disintegrate until they reach temperatures in the billions of degrees, yet atoms cannot withstand even a few thousand degrees. As for molecules, they decay at temperatures as low as hundreds of degrees. Temperature is the measure of chaos, and chaos is the antagonist of structure. However, the process of organization cannot go on without chaos, because organization is a permanent process that needs motion. The necessary motion is provided by heat.

Therefore, evolution is confined within two extremes: heat and cold. Too much heat means destruction; too much cooling means solidification and death. When evolution reaches the stage at which life becomes possible, this principle contradiction of the organization process becomes so acute that only a very narrow interval of temperatures is left for life based on organic molecules.

At present we don't know the precise range of this interval—perhaps only one or two dozen degrees. When this value is compared with the huge range of temperatures in the Universe, it becomes clear that the existence of such a temperature interval is far from being self-evident: It results from quite a fine tuning of natural laws. Moreover, the possibilities of the existence of the necessary temperature interval are not enough for life—there must be a place somewhere in the Universe where this possibility becomes reality.

The important point here is that the "life" temperature ($+20 \pm 15°C$) in such a place must be provided for billions of years! In addition, there must be free water, because no other medium is suitable for cooking the organic soup. The cradle of organic life must be protected from the murderous ultraviolet radiation of the stars, and so on. It is clear that the particular character of natural laws must be supplemented with the existence of an event of extremely small probability: the coincidence of all the necessary conditions.

The natural conditions on Earth are very well suited for life. However, even a tiny displacement of Earth relative to the Sun would be disastrous for life. Indeed, an increase of the distance between Earth and the Sun would cause a drop in temperature. The growth of the polar caps would increase the reflection of light from Earth's surface, which would induce a further decrease in temperature, and so on. Complete freezing is quite possible in this scenario. On the contrary, a small drift toward the Sun and a corresponding small increase in temperature may well trigger further increase, and the climate on Earth could become like that on Venus. The unique location of Earth precludes any hopes of finding a similar planet somewhere in a reasonable proximity.

I will not discuss the origin of life and its evolution, because these are very difficult and unclear problems about which we know too little. One point should be noted, however. The hypothesis of a probabilistic origin of life via random mixing of molecules by the "trial and error" method should be rejected for two reasons. First, to provide self-reproduction, it is necessary to have a vast amount of molecules of a definite sort. However, the number of variants is so huge (about 10^{120}) that a tiny living cell cannot indulge itself in the play of chance, to make a single correct choice. Second, life on Earth appeared very quickly, almost immediately after the planet cooled. By the way, a famous Russian cosmologist V. I. Vernadsky considered this fact as an indication that life existed forever, which is at odds with the modern theory of the hot Universe.

Thus, the origin of life is still a mystery, but the laws of its development are more or less understood. The same can be said about the origin of intelligence. Similar to the appearance of life among inorganic matter, the advent of intelligence to the non-intelligent Universe looks like a miracle. In this process, as in any other transition from one level of organization to a more complicated level, we see the same principal feature: At the beginning of its development, the more complex structure needs special conditions. It must be "cherished," "fostered," and "nursed" before it gains full strength. A grown man is more clever than a horse and more adapted to ever-changing conditions of life, but an infant is certainly much weaker than a colt in a meadow. The "coincidences" that set the stage for the play of structural complication can be formulated in mathematical terms. We'll consider this approach to the problem in the following section.

The anthropic principle as a system of equations

The mathematical formulation of almost every natural law includes some numerical parameter that is a given that cannot be modified. Here is the simplest example: Any electric charge is described by the formula $Q = Ne$, where N is an integer. In this example, the parameter e is an electron's charge. Another example: The force of gravitational attraction of two masses m_1 and m_2 is given by the law

$$F = G\frac{m_1 m_2}{r_{12}^2}.$$

The parameter in this formula is the gravitational constant G. The laws of modern physics are similarly formulated. For example, the energy of a photon is proportional to its frequency ω, so $E = \hbar\omega$. The parameter $\hbar$ in this formula is Planck's constant.

In physics, values like e, $\hbar$, G, c (the speed of light) and so on are called *universal constants*. They mean nothing by themselves, because their numerical values depend on the chosen system of units. However, their dimensionless combinations are of universal importance.

For example, there is only one dimensionless combination composed of the constants e, $\hbar$, and c: $\alpha = e^2/\hbar c \cong 1/137$. This value is known as the fine structure constant, and it characterizes the force of the electromagnetic interaction. The characteristic energy of interactions of free electrons and those in atoms is only a small correction (of about α to some degree) to their rest energy $m_e c^2$. For example, the energy of electrons in an atom with atomic number Z is

$$E \cong m_e c^2 (Z\alpha)^2.$$

Since α is a small value, the probability of the conversion of electrons into other particles is negligible. The small value of α underlies the stability of electron-proton structures (that is, atoms and solid bodies). On the contrary, heavy atoms with $Z \gtrsim \alpha^{-1} = 137$ are unstable: The electromagnetic field at the surface of the nucleus becomes so strong that it begins to produce electron-positron pairs, which screen the "surplus" nuclear charge and effectively "decrease" the charge to a value smaller than $e\alpha^{-1}$.

Similar dimensionless constants can be composed for any other interaction—strong (S), weak (W), and the gravitational (G). In the latter case it looks like

$$\alpha_G = \frac{Gm_p^2}{\hbar c} \cong 10^{-39},$$

where m_p is the mass of a proton. It is known that $\alpha_S \cong 15$ and $\alpha_W \cong 10^{-5}$. Both of these interactions are effective only at distances of about the nuclear radius.

Thus, the four types of interactions correspond to four arbitrary (free) parameters α_k ($k - \alpha$, G, S, and W). They can be furnished with the dimension of our space $d = 3$, the number of particles in the Universe $N \cong 10^{80}$, the ratio of the number of photons to the number of particles $S \cong 10^9$, and the ratio of an electron's mass to a proton's $m_e/m_p \cong 1/1830$. Theoretically, the latter value must be expressed via α, α_S, and α_W, but presently we don't know how to do this. In general, it is a cherished hope of physicists to deduce all the above cornerstone numbers from a single one. However, this remains only a dream.

A number of relationships must exist among the free parameters. To illustrate this, let's consider some examples.

The optimal expansion rate of the Universe, at which the stars have time to be born, is provided by the relationship

$$N\alpha_G^2 \cong 1.$$

The very existence of Sunlike stars (neither too hot, nor too cold) is based on the relationship

$$\alpha_G \cong \alpha^{12}\left(\frac{m_e}{m_p}\right)^4.$$

To produce a new generation of stars enriched with heavy atoms, the flux of neutrinos (the only particles that participate only in weak interactions) should jettison the mantel of a supernova. This is possible only if

$$\alpha_W^4 \cong \alpha_G\left(\frac{m_p}{m_e}\right)^6.$$

For atoms to appear in the course of the evolution of matter, they must be formed before the development of gravitational instability, which makes the Universe heterogeneous. To meet this requirement, the following relationships are necessary:

$$S \cong \alpha^{-2}\left(\frac{m_p}{m_e}\right),$$

$$S \le \alpha^{10}\left(\frac{m_p}{m_e}\right)\alpha_G^{-1}.$$

In a similar way we can consider the nuclear reactions within stars discussed above, which also have some "coincidences" that can be formulated as necessary conditions imposed on the universal constants.

The principle fact is thus: The number of restrictions (equations) that must be satisfied by the universal constants exceeds by far the number of these parameters themselves. This is a very important feature of the world we live in. Not every system of equations where the number of equations is greater than the number of unknowns has a solution. To have a solution, the superfluous equations must be deducible from the rest of the equations.

Therefore, the very structure of the natural laws hides some extremely important principle. At present we don't know how to describe this in mathematical language. It seems to be some kind of symmetry imposed on the equations describing the basic world parameters. Such symmetry would illustrate the "idea" of the existence of the Universe, the main principle of its development in time and space, and the law of evolution of its component structures. Everything we presently know is just the consequence of this main principle. All the particular laws and regularities in nature are united in a single principle law: Somewhere in the Universe a human being must appear. ◘

Quantum articles about the Universe:

Y. Zeldovich, "A Universe of Questions," January/February 1992, pp. 6–11.

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

Y. Solovyov, "The Universe Discovered," May/June 1992, pp. 12–18.

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

G. Myakishev, "The 'Most Inertial' Reference Frame," March/April 1995, p. 48.

S. Silich, "Interstellar Bubbles," November/December 1997, pp. 14–19.

I. D. Novikov, "The Thermodynamic Universe," March/April 1998, pp. 10–14.

Prime time

These numerical oddities exhibit a lack of factors

by G. A. Galperin

NOBODY WOULD DISAGREE THAT THE MAjority of natural numbers can be factored: $10 = 2 \cdot 5$, $60 = 3 \cdot 4 \cdot 5$, $111 = 3 \cdot 37$, $144 = 3 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \cdot 2$, and so on. Such numbers are called *composite*.

But numbers exist that cannot be represented in this way. For example, 11 cannot be represented as the product of two smaller natural numbers both greater than 1. For this reason, 11 is called a prime number. In general, prime numbers are those that cannot be represented as the product of two factors both greater than 1 (the number 1 is not considered a prime number). Here are the first few prime numbers: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, ... There is exactly one even number among them; all the others are odd.

Problem 1. Find all pairs of primes that differ: (a) by 1; (b) by 17.

We immediately see that the sequence of prime numbers is rather peculiar. We cannot perceive any simple law that governs the formation of the sequence.

Is this sequence finite? This question was raised in book IX of Euclid's *Elements*. The answer is also given in that book: "For any given prime number, one can find a greater prime number; that is, the sequence of prime numbers is infinite."

The proof of this proposition given by Euclid is extremely ingenious. Here is his reasoning. Assume that the sequence of prime numbers is finite and that p is the largest of them. Then, the number $N = 2 \cdot 3 \cdot 5 \cdot ... \cdot p + 1$ is not prime, since it is greater than p. Thus, it is divisible by a prime number in the range from 2 to p (by our assumption, there are no other prime numbers). However, N is not divisible by any of these prime numbers, because the remainder upon division by any of them is 1. This contradiction proves that the set of prime numbers is infinite.

Euclid's argument is an indirect proof and does not suggest any method for constructing a prime number greater than p. However, it is not difficult to suggest such a method: It is sufficient to check, for each number in the range from $p + 1$ to N, whether it is prime or not. There must be a prime number among them. Indeed, if N itself is not prime, then it is divisible by a prime number greater than p but smaller than N. That is, this number must be in the interval $[p + 1, N]$.

Intervals containing prime numbers

There is another method for partitioning the number line into intervals containing one prime number each. First, let's prove the following: The smallest divisor of the number $N = n! + 1$ (where $n! = 1 \cdot 2 \cdot 3 \cdot ... \cdot n$) is a prime number greater than n.

We'll call this smallest divisor p. We have $p > n$, because $n! + 1$ is not divisible by any of the numbers 2, 3, 4, ..., n. On the other hand, if we assume that p is a composite number (that is, if p is divisible by any number less than p), then p is not the smallest divisor of $n! + 1$, which contradicts our assumption. Thus, p is a prime number greater than n.

It follows from this proof that each interval $[n, n! + 1]$ contains at least one prime number. Thus, the numbers 2, $2! + 1$, $(2! + 1)! + 1$, $((2! + 1)! + 1)! + 1$, and so on partition the number axis into an infinite number of intervals that each contain at least one prime

Art by Vasily Vlasov


41
48
-7030
115
1997
125
109
41
119
109
49
264-79-98
123
85
10
3
54
40999
28.8
11, 13, 17, 19, 23.......
2, 3, 5, 7
B.B.

number. We have again proved the infinity of prime numbers.

Problem 2. Prove that if p divides $(p-1)! + 1$, then p is a prime number. (Hint: Use the proof of the proposition above). The converse of this statement is known as Wilson's theorem, and is also true.

There is yet another way to partition the number line into intervals each containing a single prime. It turns out that each of the intervals [2, 4], [4, 8], [8, 16], [16, 32], ... contains at least one prime number. However, this statement is difficult to prove. It follows from a theorem known as "Bertrand's postulate" (although it was actually proved in 1852 by the prominent Russian mathematician Chebyshev (1821–1894)).

This theorem is formulated as follows: For $n > 7$, there exists at least one prime number between n and $2n - 2$.

Problem 3. Using Bertrand's postulate, prove that for any natural n, (a) at least one n-digit prime number exists; (b) at least three n-digit prime numbers exist. (Hint: 10^{n-1}, $2 \cdot 10^{n-1}$, $4 \cdot 10^{n-1}$, and $8 \cdot 10^{n-1}$ are n-digit numbers.)

Note that Euclid's proof does not give the nearest prime number following p, but usually gives a number rather far from p. For example, instead of 13, this proof suggests 2311 as the prime number that is greater than 11; for 13, it gives not 17, but 59, which is a prime divisor of 30,031.

Intervals not containing prime numbers

To demonstrate the complexity of the structure of the set of prime numbers, we prove that there are arbitrary long intervals that do not contain prime numbers. For example, we can find a million successive numbers of which none is prime. Indeed, let N = 1,000,000 and consider 1,000,000 of the following numbers:

$$(N+1)! + 2, (N+1)! + 3, ..., (N+1)! + (N+1).$$

The first of these numbers is divisible by 2, the second by three, the third by four, and so on. The kth number $(N + 1)! + k$ is divisible by k, because both terms divide by k. Thus, all 1 million of these numbers are composite. This method makes it possible to find arbitrary long gaps in the sequence of prime numbers.

Interestingly enough, the problem of arbitrary long gaps in the sequence of prime numbers, which is very close to the problem of the infinity of the set of prime numbers, both in its statement and its proof, cannot be found in the works of mathematicians of ancient Greece. In the next section, we consider another problem addressed in mathematics of the modern era.

Arithmetic progressions and prime numbers

Consider all natural numbers that give the remainder 2 when divided by 3: 2, 5, 8, 11, 14, ... These numbers can be represented by the formula $3n + 2$. Let's prove that there are infinitely many prime numbers among them. For this purpose, we modify Euclid's proof a little. Instead of the number $N = 2 \cdot 3 \cdot 5 \cdot ... \cdot p + 1$, we consider the number $M = 2 \cdot 3 \cdot 5 \cdot ... \cdot p - 1$, which belongs to the sequence 2, 5, 8, 11, 14, ..., $3n + 2$, ..., because its remainder upon division by 3 is 2.

Problem 4. Give the complete proof of the fact that M can be represented as $3n + 2$.

The number M, like N in the previous proof, is not divisible by any of the numbers 2, 3, 5, ..., p. Whether M is prime itself, or has smaller prime factors, each of the factors of M is greater than p. We need to find out whether there is a number of the form $3n + 2$ among

Table 1.

A fragment of the distribution of prime numbers. The table shows the variation of the number of primes on the interval from 8,900,000 to 9,000,000 partitioned into 1000 hundreds. The first row shows the number of primes, and the second row shows the number of groups of 100 numbers that contain the corresponding number of primes. For example, 1 group of 100 does not contain primes at all, while 117 hundreds contain 4 primes each and 130 hundreds contain 8 primes each.

0	1	2	3	4	5	6	7	8	9	10	11	12
1	4	13	53	117	177	196	175	130	72	47	11	4

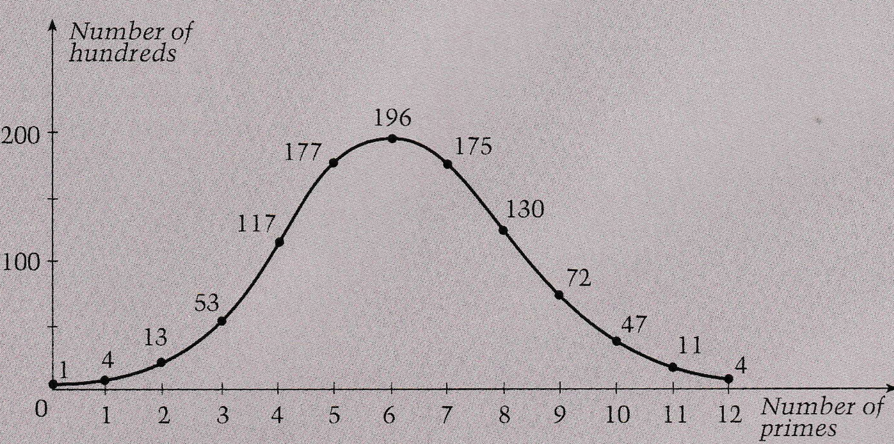


Table 2.
The distribution of prime numbers in the set of integers. A_n is the number of primes among the first n integers. As n increases, the ratio A_n/n becomes closer to $1/\ln n$ ($A_n/n : 1/\ln n$ is almost equal to 1 for $n = 10^9$).

n	A_n/n	$1/\ln n$	$A_n/n : 1/\ln n$
1 000	0.168	0.145	1.159
1 000 000	0.078498	0.072382	1.084
1 000 000 000	0.050847478	0.048254942	1.053

these prime factors—that is, a factor belonging to the sequence 2, 5, 8, 11, ... Assume that it is not true that all prime factors of M have the form $3k + 1$. However, in this case, their product also has the form $3k + 1$ (see problem 5a), and this contradicts the fact that M has the form $3n + 2$ (see problem 4). Therefore, our assumption cannot hold, and at least one prime factor of M has the form $3n + 2$. Thus, there are infinitely many prime numbers of the form $3n + 2$.

Problem 5. (a) Prove that the product of numbers of the form $3k + 1$ also has the form $3k + 1$. (b) Prove a similar proposition for numbers of the form $4k + 1$; (c) Prove a similar proposition for numbers of the form $6k + 1$.

This reasoning (with certain modifications) provides a tool for proving the infinity of prime numbers of the form $4k + 3$ and $6k + 5$. We suggest thinking about the following problem first.

Problem 6. Prove that any prime number greater than three (a) has either the form $4k + 1$ or the form $4k + 3$; (b) has either the form $6k + 1$ or $6k + 5$.

In this article we prove the infinity of the set of prime numbers of the form $6k + 5$. Our argument, like Euclid's, will be a proof by contradiction. Assume that there is only a finite number of primes of the form $6k + 5$: $p_1, p_2, \ldots, p_n$. Consider the number

$$K = 6p_1p_2 \ldots p_n - 1 = 6(p_1p_2 \ldots p_n - 1) + 5.$$

The number K is either prime or has prime factors different from $p_1, p_2, \ldots, p_n$ (why?). Not all of these prime factors has the form $6k + 1$, because K itself does not have this form (see problem 5b). Thus, at least one of the prime factors of K has the form $6k + 5$ and is different from $p_1, p_2, \ldots, p_n$. This contradicts our assumption, as it proves that there are infinitely many prime numbers of the form $6k + 5$.

Problem 7. Prove the infinitude of the set of prime numbers of the form $6k + 5$, giving an explicit method for their construction.

Problem 8. Give a detailed proof of the infinitude of the set of prime numbers of the form $4k + 3$. (Hint: Multiply the product of the numbers of this type by four and subtract 1 from this product.)

Problem 9. Prove that the set of primes whose decimal numeral do not end in 1 (that is, that end in 3, 7, or 9) is infinite. (Hint: Consider all primes of the form $10k + a$, where $a \neq 1$, and then follow the above reasoning.)

The following theorem, formulated by the French mathematician Legendre in 1788 and proved by the German mathematician Dirichlet in 1837, is a generalization of the propositions we have considered.

Theorem. Any infinite arithmetic progression a, $a + d$, $a + 2d$, $a + 3d$, ... in which the first element, a, is coprime to the difference d contains infinitely many prime numbers. In other words, the function $y = dx + a$, where a and d are coprime integers, takes infinitely many prime values when x runs through the set of natural numbers.

Dirichlet's proof is not elementary, and for a long time no elementary proof of this remarkable theorem was found. An elementary proof was first obtained in 1949 (161 years after Legendre formulated his theorem) by the Danish mathematician A. Selberg, who found elementary proofs of many difficult theorems of number theory.

Twins

Recall the first problem formulated at the beginning of the paper. As you have certainly guessed, if two prime numbers differ by an odd number p (by 1 or 17 as in problem 1), one of them is even and thus equals 2. The other prime number of the pair, q, differs from p by 2. If p is also prime, as is the case in problem 1, in which $p = 17$, the prime numbers p and q are called *twins*. In problem 1, these are 17 and 19.

Problem 10. Using Dirichlet's theorem, prove that there are infinitely many prime numbers that do not belong to any pair of twins. (Hint: These prime numbers can be taken from the arithmetic progression $15k + 7$.)

We can raise the problem: How many pairs of twins exist? For example, there are 1225 pairs of twins in the range from 0 to 100,000, and only 518 pairs of twins in the range from 8,000,000 to 8,100,000. Is the number of the pairs of twins infinite? Neither this question nor a more general one that was stated by the great German mathematician David Hilbert at the 2nd International Congress of Mathematicians in Paris in 1900 has yet received an answer. Hilbert's problem is formulated as follows: Is the linear equation $ax + by = c$ with integer coefficients a, b, and c, where a and b are coprimes solvable in the set of prime numbers?

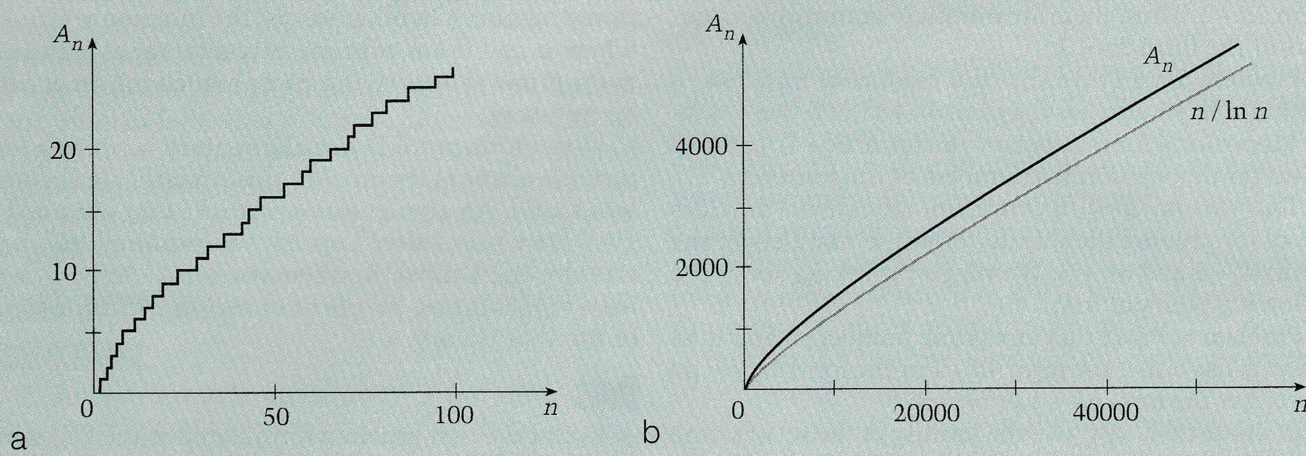
Table 3.

Prime numbers and prime twins in eight intervals of the length of 150000.

Figures (a) and (b) show the plots of the number of primes, A_n, on the segment [1, n]. Plot (a) shows that A_n increases rather regularly, though certain local variations exist. However, if we increase n up to 50,000 (fig. (b)), the regularity of the plot becomes quite clear. The smoothness of this curve is one of the most remarkable facts of mathematics. Note that for very large n, A_n is approximately equal to

$$\frac{n}{\ln n - 1.08366}.$$

interval	*number of primes*	*number of prime twins*
$10^8 \div 10^8 + 150\ 000$	8154	601
$10^9 \div 10^9 + 150\ 000$	7242	466
$10^{10} \div 10^{10} + 150\ 000$	6511	389
$10^{11} \div 10^{11} + 150\ 000$	5974	276
$10^{12} \div 10^{12} + 150\ 000$	5433	276
$10^{13} \div 10^{13} + 150\ 000$	5065	208
$10^{14} \div 10^{14} + 150\ 000$	4643	186
$10^{15} \div 10^{15} + 150\ 000$	4251	161



The fundamental theorem of arithmetic

Theorem. Any natural number greater than 1 admits a unique factorization into prime factors (apart from the order of the factors).

Proof. If there is at least one number that admits two different factorizations, then there is a smallest such number N.

The number $N = p_1p_2 \ldots p_n = q_1q_2 \ldots q_m$, where p and q are primes. We can assume that $p_1 \le p_2 \le \ldots \le p_n$ and $q_1 \le q_2 \le \ldots \le q_m$ (if this is not the case, we can change the order of the factors). Note that $p_1 \ne q_1$, because otherwise, the number $N/p_1 = N/q_1$ that is less than N would have two different factorizations into prime factors. This would contradict our assumption that N is the smallest such number. Assume that $p_1 < q_1$ and consider the number $N' = N - p_1q_2 \ldots q_m$. We can see that the number

$$N' = p_1(p_2p_3 \ldots p_n - q_2q_3 \ldots q_m) = (q_1 - p_1)q_2q_3 \ldots q_m$$

is positive and less than N. Thus, by our assumption, N' has a unique factorization into prime factors. Since the prime number p_1 is a factor in the factorization of N', it either coincides with one of the factors $q_2, q_3, \ldots, q_m$ or divides $(q_1 - p_1)$. The inequalities $p_1 < q_1 \le q_2 \le \ldots \le q_m$ show that the first case is impossible. Therefore, p_1 divides $q_1 - p_1$. However, in this case, p_1 divides q_1, which contradicts the fact that q_1 is a prime. Thus, our assumption is wrong, which completes the proof of the theorem.

Note. The proof of the fundamental theorem of arithmetic explains why the number 1 is not considered a prime number. If we included 1 in the set of prime numbers, any integer could be factored into prime factors in a number of different ways, because an arbitrary number of 1s could be included in any factorization.

Here is one important consequence of the fundamental theorem of arithmetic: If a prime number p is a factor of the product ab, then it is either a factor of a or a factor of b. Indeed, if p were a factor neither of a nor of b, then we would obtain a factorization of ab not containing the factor p by multiplying the factorizations of a and b. On the other hand, $ab = pt$, where t is an integer. Multiplying p and the factorization of t, we would obtain another factorization of ab that contained p as one of the factors. Thus, we would obtain two different factorizations of ab, which would contradict the fundamental theorem.

The fundamental theorem implies that any number N can be represented as

$$N = p_1^{k_1} p_2^{k_2} \ldots p_s^{k_s},$$

where $k_1, k_2, \ldots, k_s$ are the exponents of the prime factors $p_1, p_2, \ldots, p_s$ in the factorization of N. All divisors of N have the form $a = p_1^{r_1} p_2^{r_2} \ldots p_s^{k_s}$, where $0 \le r_1 \le k_1$, $0 \le r_2 \le k_2$, $0 \le r_s \le k_s$.

Yet another proof that the number of primes is infinite

The uniqueness of the factorization into prime factors makes it possible to give another proof that there are infinitely many primes. This proof belongs to Leonhard Euler.

Assume that 2, 3, 5, ..., p is a list of all the prime numbers that exist. The formula for the sum of a geometric progression with a ratio less than 1 implies that for any n,

$$1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^n}<\frac{1}{1-\frac{1}{2}},$$

$$1+\frac{1}{3}+\frac{1}{3^2}+\ldots+\frac{1}{3^n}<\frac{1}{1-\frac{1}{3}},$$

$$\vdots$$

$$1+\frac{1}{p}+\frac{1}{p^2}+\ldots+\frac{1}{p^n}<\frac{1}{1-\frac{1}{p}}.$$

Multiplying these inequalities term by term, we obtain

$$\left(1+\frac{1}{2}+\ldots+\frac{1}{2^n}\right)\left(1+\frac{1}{3}+\ldots+\frac{1}{3^n}\right)\ldots\left(1+\frac{1}{p}+\ldots+\frac{1}{p^n}\right)$$
$$<\frac{1}{\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\ldots\left(1-\frac{1}{p}\right)}.$$

We'll call the number on the right side of this inequality A. If we remove the parentheses on the left side of the inequality, we obtain the sum S of all the numbers that are reciprocals of the divisors of $N = 2^n 3^n 5^n \ldots p^n$ (it is here that we use the fundamental theorem of arithmetic). Therefore, the left side of the inequality is greater than $A_n = 1 + 1/2 + 1/3 + 1/4 + 1/5 + \ldots + 1/2^n$, and this sum includes only a part of the terms of S. Thus, for any n, $A_n < A$. However,

$$A_n = 1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right)+\ldots+\left(\frac{1}{2^{n-1}+1}+\ldots+\frac{1}{2^n}\right)$$
$$>1+\frac{1}{2}+2\cdot\frac{1}{4}+4\cdot\frac{1}{8}+\ldots+2^{n-1}\cdot\frac{1}{2^n}=1+\frac{n}{2}.$$

We have arrived at a contradiction: $A > 1 + n/2$ for all n. Therefore, the set of prime numbers is infinite.

Verifying primality

When we factor a number N or check its primality, we must check whether N is divisible by the sequential prime numbers 2, 3, 5, 7, ... It is sufficient to check the prime divisors that do not exceed $\sqrt{N}$. Indeed, if $N = ab$, then the smaller of the numbers a and b does not exceed $\sqrt{N}$ (if both of them were greater than $\sqrt{N}$, then the product would be greater than N). The divisibility of N by a immediately implies that N is divisible by N/a, so we do not need to verify the divisibility by N/a. Fibonacci (Leonardo of Pisa) was the first to note this fact.

Examples. (a) If $N = 91$, then $\sqrt{91} < 10$, and checking the primes 2, 3, 5, and 7, we find that $91 = 7 \cdot 13$; (b) if $N = 1987$, then $\sqrt{N} < 45$, and since N is not divisible by any of the primes up to 43, 1987 is a prime number.

In some cases, the primality of a number N can be established without performing the divisions. The following simple proposition formulated by Euler as early as the eighteenth century makes it possible to establish the primality of a number N in a quite different way.

Euler's first criterion. If an odd number $N > 1$ can be represented as a difference of the squares of two natural numbers in more than two different ways, then N is composite; if such a representation is unique, then N is prime.

Proof. We can assume that N is not a perfect square, because a perfect square is a composite number. (Likewise, we can assume N is odd.) Let

$$N = m^2 - n^2 = (m - n)(m + n).$$

Therefore, $m - n$ and $m + n$ are divisors of N. If N is prime, then $m - n = 1$ and $m + n = N$. Therefore, $m = (N + 1)/2$ and $n = (N - 1)/2$ are uniquely determined by N, and N cannot be represented as the difference of two squares in a different way.

If N is composite (that is, if $N = ab$, where $a > b > 1$ are odd), then the numbers $x = (a + b)/2$ and $y = (a - b)/2$ provide another representation of N as the difference of two squares: since $a = x + y$ and $b = x - y$, we have $N = ab = x^2 - y^2$.

We see that if N can be represented as the difference of two squares in more than one way, N cannot be prime: Prime numbers have a unique representation of this kind. Conversely, if N can be represented as the difference of two squares in a unique way, then N cannot be composite (as we have just proved), and thus, it is prime.

This criterion makes it possible to use a table of squares to verify the primality of numbers. We successively add the squares of the numbers $n < (N-1)/2$ to N and check to see if the sum is a perfect square.

For example, let's factor 3,551 by using this method. Successively adding $1^2, 2^2, 3^2, \ldots$ to 3,551, we check whether the sum obtained is a perfect square. The verification (using the table of squares) shows that $3{,}551 = 60^2 - 7^2 = 53 \cdot 67$.

Problem 11. Use the method described to factor the following numbers: 6,557, 19,019, and 209,209.

It is not difficult to prove the following criterion.

Euler's second criterion. If a natural number N can be represented as a sum of two squares in more than one way, then N is composite (changing the order of the addends does not create another representation).

It follows from Euler's second criterion that if a prime number can be represented as the sum of two squares, then this representation is unique. It is interesting to find out which prime numbers can be represented in this form.

Problem 12. Prove that the numbers of the form $4k + 3$ cannot be represented as a sum of two squares. (Hint: The square of any even number is divisible by 4, and the square of any odd number gives a remainder of 1 when divided by 4.)

Thus, only prime numbers of the form $4k + 1$ are candidates for this representation. Fermat proved that all such primes can be represented as the sum of two squares. His result allows us to answer questions such as which of the three prime numbers 1973, 1979, and 1987 can be represented as the sum of two squares.

Problem 13. (a) Prove that for any integer n, $N = n^4 + 4$ is composite (Germain's theorem). (b) Prove that for any integers m and n, $N = n^4 + 4m^4$ is composite. (Hint: $N = (n^2 - 2m^2)^2 + (2mn)^2$; then, apply Euler's second criterion.

Polynomials that generate primes

It would not be difficult to deal with prime numbers if a simple formula existed that made it possible to find them. Attempts to find such a formula have been made for a long time. For example, Euler found a remarkable trinomial: $n^2 + n + 41$ that takes prime values for $n = 0, 1, \ldots, 39$. However, for $n = 40$, its value is 41^2, which is certainly not a prime. It is not difficult to prove that *no polynomial of a single variable can take only prime values.*

Quite recently, a polynomial has been found such that all its positive values at integer points coincide with the set of all prime numbers. The polynomial involves 25 variables, so one must use all but one of the letters of the alphabet to write it down:

$$
\begin{aligned}
P = (k+2)\Big\{1 &- [wz+h+j-q]^2 - [(gk+2g+k+1)(h+j)+h-z]^2 \\
&- [2n+p+q+z-e^2] - \left[16(k+1)^3(k+2)(n+1)^2+1-f^2\right]^2 \\
&- \left[e^3(e+2)(a+1)^2+1-o^2\right]^2 - \left[(a^2-1)y^2+1-x^2\right]^2 - \left[16r^2y^4(a^2-1)+1-u^2\right]^2 \\
&- \left[\left(\left(a+u^2(u^2-a^2)\right)^2-1\right)(n+4dy)^2+1-(x+cu)^2\right]^2 - [n+l+v-y]^2 \\
&- \left[(a^2-1)l^2+1-m^2\right]^2 - [ai+k+1-l-i]^2 \\
&- \left[p+l(a-n-1)+s(2ap+2a-p^2-2p-2)-x\right]^2 \\
&- \left[z+pl(a-p)+t(2a-p^2-1)-pm\right]^2\Big\}.
\end{aligned}
$$

This polynomial was found as a result of the study of Diophantine equations. It is related to the solution to Hilbert's 10th problem found by the Russian mathematician Y. A. Matiyasevich.

In conclusion, we suggest a few more problems.

Problems

14. Find all prime numbers that are simultaneously the sum and difference of two prime numbers.

15. Prove that the square of any prime number $p > 3$ gives the remainder 1 when divided by 12.

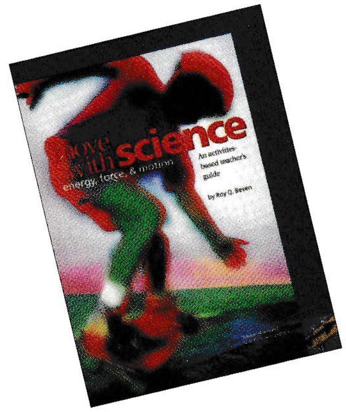
16. Prove that if p and $p^2 + 2$ are prime numbers, then $p^3 + 2$ is also a prime number.

17. Which prime numbers can be represented as the sum of two cubes of integers?

18. What is n if $n + 1$, $n + 3$, $n + 7$, $n + 9$, $n + 13$, and $n + 15$ are prime numbers?

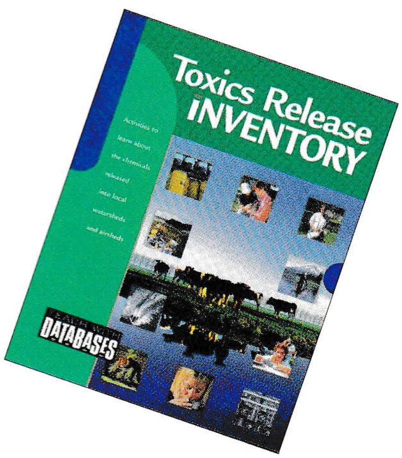
New NSTA Publications...
The Latest in Science Education

Move With Science: Energy, Force & Motion



This resource uses methods of transportation that are most familiar to high school students to connect the basic concepts of physics and human biology to the concrete sights, sounds, and physical sensations that students experience nearly everyday. *Move with Science* brings real transportation situations, representing such concepts as inertia, stability, and the relationships between mass, energy, and motion, into the classroom through analogous hands-on activities and background reading sections. This book also focuses on the physics and biology of transportation safety and safety devices, providing an understanding of physics which will allow students to make informed safety decisions. (grades 9–12, 1998, 160 pp.)

#PB144X $21.95 Member Price $19.76

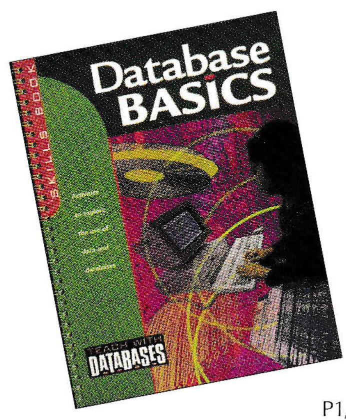


Teach With Databases: Toxics Release Inventory

This resource was designed by NSTA and the Environmental Protection Agency to provide teachers and students with a community-based context for chemistry, biology, and environmental science classes. The TRI database is a record of chemicals whose release—into air, surface water, or groundwater—is monitored by the EPA. This package offers TRI as a tool for student investigation, with a range of labs covering sampling and analysis methods, computer-based explorations, and guidance for using this real data from the public and private sectors to investigate the chemical history of your local watershed. Includes *Toxics Release Inventory Teacher's Guide*; TRI CD-ROM and users guide; *Database Basics*, a guide to using databases; *Getting Started*, a primer to environmental education; EPA's *Guide to Environmental Issues*. (grades 9–12, 1998)

#PB143X01 $35.00 Member Price $31.50

Database Basics



Database Basics leads students through the process of gathering data, organizing it, describing it, identifying trends, and then using data to make decisions. Each activity includes teaching tips, data collection methods, sample data sets, procedures, student worksheets, and assessment suggestions. Suggestions are provided for incorporating these data skills into the science curriculum along with the *National Science Education Standards*. (grades 9–12, 1998, 60 pp.)

#PB142X $12.95 Member Price $11.66

P1/299Q1

Advance classroom learning and inspire discovery...

Call 1–800–722–NSTA to order.

www.nsta.org/scistore

Satellite aerodynamic paradox

Examining the forces affecting objects in near-Earth orbits

by A. Mitrofanov

HOW DO YOU THINK A BODY can increase its velocity while moving in a resistant medium? At first glance such an event seems no more probable than the great Baron Munchausen's famous proposal to lift himself up by pulling his hair with his own hands.

It seems conventional that the velocity of a moving body decreases in a viscous medium. Still, do not jump to conclusions. In reality, such an event is quite possible and is regularly observed when satellites or meteors move about Earth in the outer atmosphere, which is a rarefied gas. This is the famous satellite aerodynamic paradox: When entering the upper atmosphere, the spacecraft is slowed by the rarefied gas, but it nevertheless manages to increase its velocity.

Before cracking this orbital nut, let's consider a simple example from classical mechanics. A small bob fastened to the end of an elastic cord moves with a constant speed along a circle in the horizontal plane. The cord obeys Hooke's law—that is, its restoring force is proportional to its extension or compression. If the bob is somehow slowed, its motion will be modified. For example, if the bob is instantaneously stopped and then

Art by Pavel Chernusky

set free, the stretched cord will give the bob some velocity toward the center of rotation. The maximum speed acquired by the bob can be found by applying conservation of energy to the system consisting of the bob and elastic cord. In the case when the stretch of the cord is far greater than its initial length, the maximum final velocity of the bob will be almost equal to its initial rotational velocity.

Our example is a very simplified analogy of what happens with satellites during aerobraking in the atmosphere. The bob plays the role of a satellite, and the cord mimics the gravitational attraction of Earth. Of course, this analogy is a far cry from reality, because Hooke's law has nothing to do with gravitational forces, which, similar to electrostatic forces, vary inversely with the square of the distance between the attracting bodies (in our case, to the square of the radius of the satellite's orbit). The motion of a satellite in the outer atmosphere is much more interesting and complicated than the motion of a bob on a rubber cord.

To study the orbit of a satellite moving in a rarefied gas, we need some formulas. Let's consider a satellite of mass m in a circular orbit of radius R around the Earth of mass M. In high orbits the major force affecting a satellite is Earth's gravitational attraction, so the velocity of a satellite is determined by the equation

$$v^2 = \frac{GM}{R}, \text{ or } v = v_0\sqrt{\frac{R_0}{R}}, \qquad (1)$$

where G is the gravitational constant; R_0 is Earth's radius, which is approximately equal to 6400 km; and $v_0 = \sqrt{g_0 R_0}$ is the orbital velocity just above the surface. Since the value of acceleration due to gravity near Earth's surface is $g_0 = 9.8 \text{ m/s}^2$, the orbital velocity is 7.9 km/s.

The resistive force acting on a satellite due to the rarefied gas of the outer atmosphere is given by the formula

$$F_{\text{res}} = C_x \frac{\rho v^2}{2} S_x. \qquad (2)$$

Here ρ is the density of the orbital atmosphere, which strongly depends on the satellite's altitude; S_x is the cross-sectional area of the satellite (more precisely, the area of maximum cross-section of the satellite perpendicular to the velocity vector $\mathbf{v}$ of the satellite's flight relative to the medium (this value is referred to as the midsection); and C_x is the drag coefficient, which, strictly speaking, depends on velocity, although for a high-altitude flight of a satellite, it is about 2.

This value means that collisions of gas molecules with the heat shield of a satellite are inelastic, so in a unit time, the satellite is given a momentum ρv^2 per unit area of its midsection. Recall that the orbital speed of a satellite is far greater than the mean speed of thermal motion for atmospheric molecules and atoms. (Otherwise Earth would lose its atmosphere very quickly!) Therefore, when calculating the aerobraking forces in the following examples, we neglect the thermal motion of the particles that compose the surrounding medium.

What results from the existence of rarefied gas at the altitude of the satellite's orbit? In the case of high orbits, the resistive forces are small disturbances that cause slight variations in the orbital parameters. During gradual braking in a rarefied gas, a satellite descends to a lower orbit. However, formula (1) says that at smaller values of R the orbital velocity should be greater. Thus, the resistive force acting opposite the satellite's velocity can accelerate the satellite in the direction of its motion!

Moreover, we will see that the tangential acceleration—the acceleration along the trajectory—is exactly equal to the resistive force divided by the satellite's mass. This interesting phenomenon is called the satellite aerodynamic paradox, and now we'll scrutinize it more closely. It is curious that this seemingly difficult problem can be tackled using only conservation laws and elementary calculations.

The increase of a satellite's speed during aerobraking in the outer atmosphere has a very simple explanation. To put it bluntly, when a satellite loses its initial tangential speed, it falls in Earth's gravitational field, because the attraction force $F = GMm/R^2$ becomes larger than the force mv^2/R, which is necessary to keep the satellite in its initial orbit. However, the satellite falls not in the vertical direction (say, like a brick from a high building), but along a gradually decreasing spiral, nearing Earth's surface with every turn. Each turn is almost a circle.

As we know, when a body falls in a gravitational field, its velocity increases. In the case of a decelerating satellite, a decrease in its potential energy not only compensates for the work of frictional forces in the orbit (the resistance of the medium) but also increases the satellite's speed v and its kinetic energy $mv^2/2$. Therefore, it is not the aerial friction that accelerates a falling space vehicle, but the attraction of our planet. The resistive forces only help to transfer a satellite from a high orbit into a low one—just as in the simple mechanical analogy with a bob fastened to an elastic cord.

Now look at fig. 1, which shows the trajectory of an artificial satellite in the upper atmosphere and the forces affecting the satellite. This motion occurs in a plane and is characterized by a slow decrease in the orbital radius. In other words, the trajectory is a spiral that gradually

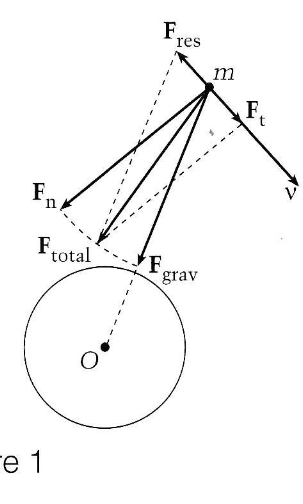


Figure 1

approaches Earth. The decrease in altitude for each turn of the spiral is small compared to the satellite's altitude $h = R - R_0$. In this figure, $\mathbf{F}_{grav}$ is the force of gravitational attraction to Earth, $\mathbf{F}_{res}$ is the force of air resistance, $\mathbf{F}_{total}$ is the vector sum of $\mathbf{F}_{grav}$ and $\mathbf{F}_{res}$. Since the satellite's trajectory is a spiral, every turn of which differs from a circle (although by a very small value!), the force $\mathbf{F}_{total}$ can be decomposed onto two constituent parts: $\mathbf{F}_n$ and $\mathbf{F}_t$—that is, the normal and tangential (to the satellite's trajectory) components. The force $\mathbf{F}_t$ that acts along the satellite's trajectory increases its speed such that at a given point on the trajectory the instantaneous acceleration in the direction of vector $\mathbf{v}$ has a magnitude F_t/m. Now we show that $F_t = F_{res}$.

Let the force of air resistance F_{res} determined by formula (2) act on a satellite revolving at some orbit of radius R. The density $\rho(R)$ is assumed to be constant and small along the entire orbital turn. We are to find the increase in the satellite's velocity Δv and the decrease in its orbital radius ΔR in a single turn. We will do this with the help of conservation of energy, taking into account the work performed by the force of air resistance. Recall that the potential energy of a satellite in an orbit is

$$E_1 = -\frac{GMm}{R} = -mv^2,$$

and its kinetic energy is

$$E_2 = \frac{mv^2}{2} = -\frac{E_1}{2}.$$

Therefore, the total energy of the satellite is

$$E_1 + E_2 = -\frac{mv^2}{2}.$$

The balance of energy of the satellite at the start and finish of an orbital turn is described as follows:

$$-\frac{mv^2}{2} - 2\pi R F_{res} = -\frac{m(v+\Delta v)^2}{2}. \quad (3)$$

As $\Delta v << v$, this equation yields

$$\Delta v = \frac{2\pi F_{res} R}{mv} = \frac{2\pi F_{res}\sqrt{R}}{m\sqrt{g}},$$

or

$$\frac{\Delta v}{v} = \frac{2\pi F_{res}}{mg}. \quad (4)$$

The period of the satellite is $\Delta t = 2\pi R/v$, so the tangential acceleration of a satellite moving in the upper atmosphere is

$$a_t = \frac{\Delta v}{\Delta t} = \frac{2\pi F_{res} R}{mv}\frac{v}{2\pi R} = \frac{F_{res}}{m}, \quad (5)$$

from which we obtain

$$F_t = ma_t = m\frac{F_{res}}{m} = F_{res},$$

which was to be proved.

Therefore, the larger the force of air resistance, the greater the increase in the satellite's speed. Can you imagine such a thing when riding a toboggan? Certainly not. Why, by the way? After all, a satellite and toboggan move in the same gravitational field, don't they?

Let's find the decrease in the orbital radius R in a single turn. The relationship between ΔR and Δv is easily obtained from formula (1):

$$\Delta v = -\frac{v}{2R}\Delta R,$$

where Δv was found above. Therefore,

$$\frac{\Delta R}{R} = -\frac{4\pi F_{res}}{mg}. \quad (6)$$

Note that the relative decrease in altitude is exactly two times the respective increase in the relative speed of the satellite.

Look at formula (6). Imagine unrolling the circle of radius R and a single turn of the satellite's trajectory into the line segments AB and AC (fig. 2). Arrange them such that segment BC assumes the value

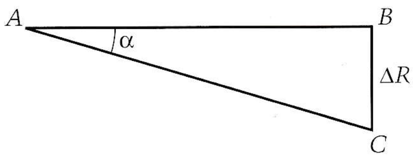


Figure 2

$$\Delta R = \frac{4\pi F_{res} R}{mg}.$$

Now we can see that a tangent line drawn at any point of the satellite's trajectory in the rarefied atmosphere is deflected from the local horizontal by a certain small angle:

$$\alpha \cong \arctan\frac{4\pi F_{res} R}{mg \cdot 2\pi R} \cong \frac{2F_{res}}{mg}.$$

This angle is not constant. It depends on the resistive force and therefore on the altitude of the satellite's trajectory. The larger the braking force, the greater this angle. The satellite moves like a body sliding along an inclined plane, where the component of the gravitational force driving the body along this plane is $mg \sin\alpha \cong 2F_{res}$, about twice that of the resistance force. The vector sum of the driving and resistance forces is equal in magnitude to F_{res} and directed forward. Therefore, the resulting force accelerates the satellite, which explains the nature of the aerodynamic paradox.

Perhaps, an experienced reader may have noted, the aerodynamic paradox in the given formulation results from a specific feature characterizing the gravitational and electrostatic fields, in which the total energy of a body is equal to the negative of its kinetic energy. For example, if the force of gravitational attraction to Earth varies as $1/R^3$, the tangential acceleration of a satellite in the rarefied atmosphere would be $F_{res}/3m$. Problem 4 (at the end of the article) considers the general case, when the radial dependence of the attractive force in the rarefied atmosphere is described by a power law.

We calculated the acceleration of a satellite on the basis of the balance of energy in the gravitational field, taking into account the work performed by the external resistive force. The same result could be obtained in another way without conservation of energy. To this end, we use the equation describing the rate of change of the satellite's angular momentum $L = mvR$ in the circular orbit:

$$\frac{\Delta L}{\Delta t} = \tau, \qquad (7)$$

where $\tau = -F_{res}R$ is due to the external force. This torque decreases the angular momentum of the satellite during its braking in the atmosphere and causes the satellite to descend from a high to a low orbit. Let ΔL be the decrease in angular momentum during a single orbital turn. As previously, the gas density at the satellite's orbit is assumed to be so small that the force F_{res} produces only a small disturbance to the orbit during a revolution of the satellite. Thus, $\Delta L = mv\Delta R + mR\Delta v$. Previously we obtained the change in speed in the form

$$\Delta v = -\frac{v\Delta R}{2R},$$

which is true for gravity. This formula and equation (7) yield the same formulas for Δv, ΔR, and a_t as we obtained earlier.

Equation (7) describes the evolution of angular momentum and helps simplify many problems of a satellite's movement in the centrally symmetric gravitational field, because in this approach we need not take into account the torque of the gravitational force: The force acts exactly through the center of masses of the satellite and Earth, so its torque is zero.

The aerodynamic paradox and related problems are important in applications. Here are some examples.

Case 1. *Density of the atmosphere at high altitudes.*

Observations of the aerobraking of satellites made it possible to determine the profile of atmospheric density at altitudes so high that neither airplanes nor balloons can fly.

If a single force that changes the angular momentum of a satellite is the force of resistance

$$F_{res} = \frac{C_x \rho v^2 S_x}{2},$$

where $\rho = \rho(R)$ is the unknown function of the density dependence on the orbital radius R (or altitude of the flight $h = R - R_0$), then simple calculations using the angular momentum equation (you may do them on your own) result in equations that describe the function $\rho(R)$ using the data for the rate of decrease either of the orbital radius dR/dt or the satellite's period of revolution dT/dt obtained at various altitudes:

$$\rho(R) = -\frac{1}{2CvR}\frac{dR}{dt}, \qquad (8)$$

$$\rho(R) = -\frac{1}{6\pi CR}\frac{dT}{dt}, \qquad (9)$$

Here $C = C_x S_x/2m$ is a constant factor known as the satellite ballistic coefficient (its units are $m^2 \cdot kg^{-1}$). Equations (8) and (9) are valid for high-altitude circular orbits where collisions of a satellite with molecules in the rarefied atmosphere produce only minor changes in its orbit.

In the times of yore when artificial satellites did not dance in a ring about Earth, data on the upper atmosphere were obtained through astronomical observations and radiolocation of the flight of meteors and meteorological rockets. The modern navigational devices placed aboard satellites and the radio transmitters working hand in hand with land-based computers made it possible to track satellites and detect their orbital parameters with very high precision. Due to the many observations of satellite flights at various altitudes, a vast amount of information is now available on the gas density in the upper atmosphere, as well as on its dependence on the season, time of day, latitude, solar activity, and so on.

Experiments for determining gas density in the upper atmosphere can be more easily conducted with ball-shaped satellites, in which the cross-sectional area S_x and thus the ballistic coefficient C do not depend on the orientation of the satellite. The American *Explorer* satellites had such a spherical form. In addition, they were hollow, a feature that enhanced the efficiency of aerobraking as they probed Earth's atmosphere in a broad range of altitudes up to 1000 km, where the atmospheric density is about 10^{-13} to 10^{-15} kg/m^3.

Case 2. *The last orbital turn.*

Let's evaluate the decrease in a satellite's altitude during a single turn in the upper atmosphere. We assume the mass of the satellite to be 10^3 kg and its midsection to be 1 m^2. The mean air density at an altitude of 200 km is $4 \cdot 10^{-10}$ kg/m^3. Formula (6) yields

$$\Delta R = \frac{4\pi\rho v^2 S_x R}{mg} \cong 2 \text{ km}.$$

At first glance, ΔR seems to be a small value: At every point of the satellite's trajectory at this altitude, the velocity vector is deflected from the local horizontal by a negligible angle

$$\alpha \cong \frac{2\rho v^2 S_x}{mg} \cong 5 \cdot 10^{-6} \text{ rad} \cong 1''.$$

However, a satellite makes 16 revolutions per day, and due to a continuous decrease in altitude, it sinks into the layers where the atmospheric density sharply increases. At such altitudes the satellite plunges into the atmosphere more and more steeply. At an altitude of 150 km, where $\rho \cong 4 \cdot 10^{-9}$ kg/m^3, the same satellite loses 20 km in altitude per turn! One or two additional turns, and it encounters such dense air that it cannot finish a turn, and instead of continuing its spiral trajectory, it begins to fall almost vertically. Such a fall is characterized by huge mechanical and thermal loads. The end of orbital flight is nigh! Hollow, light satellites fall more rapidly, because

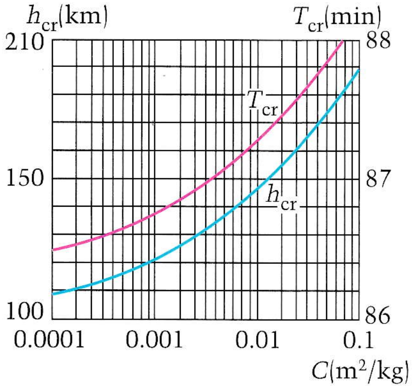


Figure 3

they are forced out of orbit at higher altitudes, while their heavy siblings can revolve around Earth at lower altitudes.

Figure 3 shows how the critical altitude and the corresponding critical period of a satellite's revolution around Earth depend on its ballistic coefficient C. For example, consider a satellite with a mass of 2.4 t and a diameter of 2.3 m, so its ballistic coefficient is $C = 1.7 \cdot 10^{-3}\ \mathrm{m}^2 \cdot \mathrm{kg}^{-1}$. The plot shows that the critical altitude is $h_{cr} \cong 130$ km, and the critical period of revolution $T_{cr} \cong 86$ min 54 s. The satellite considered in this example has approximately the same ratio S_x/m as the spacecraft *Vostok* and similar critical orbital parameters. Specifically, the critical altitude of the satellite's flight is about 125 km.

Note that the critical altitude of an ice ball 1 cm in diameter is higher than 200 km, and this parameter is even greater for smaller objects! Therefore, Earth and its atmosphere work as a huge vacuum cleaner, which eliminates small litter from near-Earth orbits.

Another point of interest: When a descending satellite approaches the critical altitude, the force of resistance is still not as large as the force of gravity. According to formula (6), these forces relate to each other as (approximately) the effective thickness of the atmosphere and Earth's radius multiplied by 4π. Thus, the force of resistance at the critical altitude is only about 1/10,000 of the force of gravity. Isn't that a tiny value? Perhaps, but it is quite enough to destroy a satellite in the near future.

Case 3. *Lifetime of a satellite in orbital flight.*

The upper atmosphere shortens the lifetime of a satellite. The formulas derived in this article help evaluate this value, provided the altitude profile of atmospheric density and the initial altitude of the orbital flight are known. Although precise calculation of a satellite's lifetime is a laborious task, a simplified estimate assumes that the majority of a satellite's life occurs at the highest altitudes, where the air density is minimal. The estimates depend on the type of satellite, or more specifically, on its ballistic coefficient C. We do not discuss the solution to this problem, but we show the results in Table 1, which provides lifetime estimates for conventional scientific satellites in orbits with various initial altitudes.

Table 1

altitude (km)	*lifetime*
150	1 day
190	2 days
210	1 week
230	1 month
400	1 year
500	10 years
650	100 years
850	1,000 years
1300	10,000 years
2000	100,000 years

First of all, this table shows how drastically the air density decreases at high altitudes. It also explains why a scientific satellite tightly packed with expensive devices designed for many years of active work is launched into an orbit with an altitude of no less than 500 km.

Problem 1. The first launches of satellites showed a curious phenomenon. After the satellite separated from the final stage of the carrier rocket, the rocket outran the satellite and flew ahead of it even when the rocket engines were stopped. What is the reason for this strange phenomenon? Assume that at the moment of separation the velocities of the satellite and the carrier rocket were identical.

Problem 2. Why does aerobraking of a satellite in the outer atmosphere transform its initial elliptical orbit into a circular one?

Problem 3. Show that if the air density decreases with altitude as $\rho \propto R^{-1/2}$, where R is the distance from the planet's center, the rate of decrease in the orbital radius of a satellite is constant.

Problem 4. A satellite revolves in a circular orbit about a planet with a rarefied atmosphere. Let's assume that the gravitational attraction to the planet obeys the law $F \propto R^{-n}$, where n is an arbitrary positive number ($n = 2$ corresponds to the usual gravitational force). For what range of n is the satellite aerodynamic paradox possible?

Problem 5. Does the braking of a satellite at high altitudes depend on air temperature?

Problem 6. Assume that a strong heating led to the evaporation of all the water in the Earth's oceans and the formation of a thick and hot atmosphere of water vapor. How would the motion of the Moon and existing artificial satellites be affected by such a metamorphosis?

Quantum articles on aeromechanics and aerobraking:

A. Eisenkraft and L. D. Kirkpatrick, "A physics souflée," July/August, 1997, pp. 30–33.

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

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

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

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

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

Challenges

Math

M251

Pick two. Six different numbers are given. Prove that one can choose two of them, say x and y, such that the following inequalities hold:

$$0 \le \frac{x-y}{1+xy} \le \frac{1}{\sqrt{3}}.$$ (I. N. Sergeyev)

M252

Triangle construction. Two nonintersecting circles with radii R and r are each tangent to both sides of the same angle. Construct an isosceles triangle such that its base lies on one side of the angle, the vertex is on the other side, and each leg touches one of the circles. Express the length of the altitude to the base of this triangle in terms of R and r. (I. F. Sharygin)

M253

Nailing it down. An equilateral triangle made of a piece of cardboard lies on a plane. Three nails are driven at points K, L, and M at its sides in such a way that the triangle cannot move (fig. 1). It is given that points K and L divide their corresponding sides in the proportion of 2:1 and 3:2 as in figure 1. In what proportion does point M divide its side of the triangle? (A. Shen)

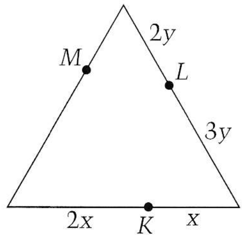


Figure 1

M254

Slicing a cube. A plane intersects a unit cube and divides it into two polyhedrons. It is known that the distance between any two points of one polyhedron does not exceed 3/2. What value can the area of this section take? (N. P. Dolbilin)

M255

Convention glad-handing. Delegates from 100 countries arrived at an international conference. Each delegation consisted of two persons—the President and the Prime Minister. Before the beginning of the conference, some of the participants shook hands, but none of the Presidents shook hands with his or her own Prime Minister. During the adjournment, the President of Illyria asked all the other participants how many handshakes they gave. All the answers were different. With how many conferees did the Prime Minister of Illyria shake hands? (A. Andzhans)

Physics

P251

Railroad robber. On a flat section of railroad tracks there was a flatcar with a load. One night a robber sneaked up to it carrying a light, elastic rubber cord. He tied one end of the cord to his belt and the other to the flatcar. Then the robber ran along the tracks with a constant speed of 5 m/s. Then something happened . . . when he came to he was lying on the flatcar, which was moving with a speed of 9 m/s. By how much did the flatcar's mass exceed that of the robber? What happened there, after all? Assume that the robber's boots did not slip, and neglect rolling friction. (A. Vargin)

P252

Wheel on an incline. A wheel consists of a thin rim of mass M, very light spokes, and an axle of mass m. The wheel is put on an incline that makes an angle α with the horizontal, and then it is set free. What speed will the wheel acquire at the time it has covered a distance L if it rolls without slipping? For what minimal coefficient of friction is motion without slipping possible? (A. Zilberman)

P253

A holey pail. A round hole with diameter d = 10 cm is drilled in the bottom of a cylindrical vessel to let the water drain out. When the hole is open, some water nonetheless remains in the vessel. Estimate the mass of this residual water if the water does not wet the bottom (that is, the force of adhesion is zero). The diameter of the vessel is D = 50 cm, and the surface tension of water is ρ = 0.07 N/m. (V. Mozhayev)

P254

An unknown lamp. Figure 2 shows the dependence of the current on the applied voltage for a lamp of unknown construction.

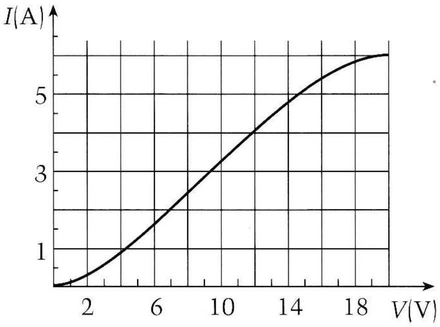


Figure 2

This lamp is connected to a source via a 10 Ω series resistor. At what source voltage will the power dissipated by the lamp be 1/4 of the power produced by the source? (B. Grinchenko)

P255

Interference after reflection. In an interference scheme with a Lloyd mirror, a point source S is located at a horizontal distance b = 20 cm from a flat mirror, and at a height a = 10 cm above the mirror as shown in figure 3. The

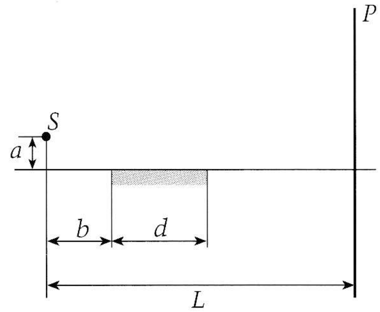


Figure 3

mirror has a length d = 10 cm. A screen P is placed at a distance L = 1 m from the point source. Determine the vertical height of the interference picture that is formed on the screen. (N. Beryulova)

ANSWERS, HINTS & SOLUTIONS ON PAGE 47

AT THE
BLACKBOARD I

The problem of eight points

by N. B. Vasiliev

SOMETIMES A PROBLEM WITH a simple statement has a complicated solution. On these occasions, the challenge is often to pull out a solution that is orderly, and from which one can learn. The following problem is of this type. It was suggested to students of the ninth form at the All-Soviet mathematical olympiad in Ashkhabad in 1983. It is given here with some minor modifications.

Problem 1. A circle is circumscribed around a triangle ABC. Lines AP, BP, and CP are drawn in the triangle's plane through an arbitrary point P (not on the circle), and the second points of intersection of these lines with the circle are marked. Prove that there are no more than eight points P for which the marked points do not coincide with any of the triangle's vertices A, B, and C and which are the vertices of a triangle congruent to the original triangle ABC.

This problem turned out to be rather difficult. Some students solved it by analyzing various positions of point P relative to lines AB, BC, CA, and the circle. In this article, we give a more instructive solution, motivated by the idea of "moving" figures about the plane. That is, we imagine that certain elements of our configuration rotate. When they reach a particular position, the desired figure appears.

First, we solve the following inverse problem.

Rotation and intersection of lines

Problem 2. Two (not necessarily congruent) triangles ABC and $A_1B_1C_1$ are inscribed in a circle. Triangle ABC is fixed, and triangle $A_1B_1C_1$ rotates about the center of the circle. In what positions of $A_1B_1C_1$ do the lines AA_1, BB_1, and CC_1 pass through the same point P? How many such positions are there?

The answer to the last question is as follows. Such a position is unique. That is, as triangle $A_1B_1C_1$ makes a complete rotation, lines AA_1, BB_1, and CC_1 meet at a point only once (and in a certain degenerate case, such a position does not exist).

This problem can be solved by using the method of loci.

Lemma. Let chord AB of the circle be fixed, and let the ends of chord A_1B_1 slide along the circle. Then the angle ϕ between lines AA_1 and BB_1 remains fixed, and their point M of intersection (if $\phi \neq 0$) describes a circle passing through points A and B (fig. 1).

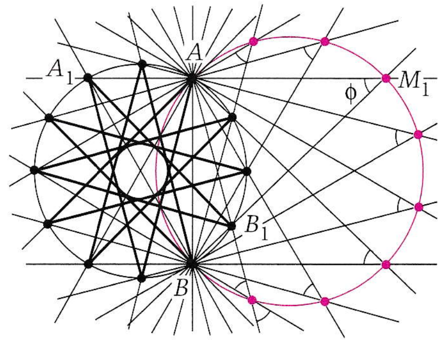


Figure 1.

If points A *and* B *are fixed and points* A_1 *and* B_1 *move uniformly with the same angular speed* ω *along the circle, lines* AA_1 *and* BB_1 *rotate uniformly with the angular speed* $\omega/2$*, and the point of their intersection,* M_1*, moves along the red circle (with the angular speed* ω*).*

Here is a way of thinking about the proof of this lemma. Assume that points A_1 and B_1 are moving uniformly with the same speed along the circle. Then lines AA_1 and BB_1 rotate uniformly with the same speed about points A and B, respectively. Therefore, the angle between them does not change (in fig. 1, for points M on one side of AB, angle AMB equals ϕ, and for points on the other side, this angle equals $\pi - \phi$). If, at the initial moment, lines AA_1 and BB_1 intersect at a point M_1, the circle circumscribed about triangle ABM_1 is the desired trajectory of point M. Point M moves uniformly along this circle (the angular speed of the rotation of the lines equals half the angular speed of the rotation of points A_1, B_1, and M along their respective circles).

A formal proof of this lemma would involve several applications of the inscribed angle theorem. The reader is invited to construct such an argument.[1]

We must note two special cases of the situation in the lemma. They will be useful in future considerations.

1. The special case $\phi = 0$ occurs when chords AB and A_1B_1 are equal, and at a certain initial moment, point A_1 coincides with B and point B_1 coincides with A. In this case, lines AA_1 and BB_1 initially coincide, and then, as chord A_1B_1 moves, they become parallel.

2. When A_1 coincides with A (or B_1 coincides with B), the line AA_1 (or BB_1) must be considered tangent to the circle (if we do not make this assumption, the two corresponding

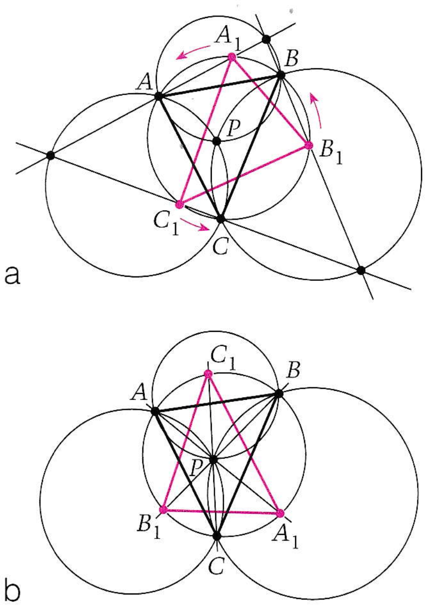


Figure 2.

Point P *is at the intersection of two loci: When the line reaches the position* BP, *lines* AA_1 *and* CC_1 *coincide with* AP *and* CP.

points must be excluded from the locus).

Let's return to problem 2. Using the lemma, we can construct two circles that are the loci of the points of intersection of line AA_1 with BB_1 and BB_1 with CC_1 (fig. 2). The first circle passes through points A and B, and the second passes through points B and C. Only one point can play the role of P—the point of intersection of these two circles that is different from B. We needn't construct the third locus of intersection of AA_1 and CC_1—the third circle. This third circle will pass through the same point P. A proof of this fact is left to the reader.

Thus, in general, a unique position of triangle $A_1B_1C_1$ exists that satisfies the condition required. In any case, taking into account the refinements we have made, we can say that there

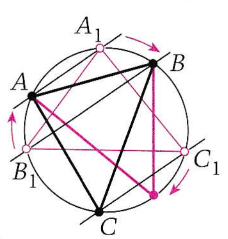


Figure 3.

A particular case $\phi = 0$*: Point* P *does not exist (lines* AA_1, BB_1, *and* CC_1 *are parallel).*

can be no more than one point P (the special case is shown in fig. 3). Thus, problem 2 is solved.

Now we can turn to problem 1. For every point P not on the circle, denote by A_1, B_1, and C_1 the second points of intersection of lines AP, BP, and CP with the circle. The statement of the problem concerns points P such that triangle $A_1B_1C_1$ is congruent to triangle ABC. At first glance, it may seem that by taking the congruent triangles in problem 2, we can find a single desired triangle $A_1B_1C_1$ that is symmetric to triangle ABC with respect to the center of the circle. However, there is a fine point in the reasoning that is more logical than geometrical, which we will discuss after a brief excursion.

Permutation of vertices and symmetries

Congruent triangles have, by definition, congruent angles and sides, and they can be superimposed. It is common practice to write the congruence of triangles so that the corresponding vertices are listed in the same order. So, for example, if $\triangle ABC \equiv \triangle DEF$, then $\angle A = \angle D$, $\angle B = \angle E$, $\angle C = \angle F$, $AB = DE$, and so on.

Looking back at the statement of problem 1 with this in mind, we notice that triangle $A_1B_1C_1$ is not necessarily congruent to triangle ABC. If we take into account the correspondence of vertices, it can be congruent to any of the six triangles ABC, BCA, CAB, BAC, ACB, and CBA. And the number of variants is double this, as we will see, for purely geometrical reasons.

Triangle $A_1B_1C_1$ can be *directly* congruent to triangle ABC. That is, these triangles can be superimposed by a continuous motion on the plane (by a rotation R in our problem), or *inversely* congruent. In the latter case, in order to superimpose the triangles, we must "flip" one of them (reflect it in a line). In our problem, it suffices to reflect one of the triangles with respect to a certain line. All the triangles $A_1B_1C_1$ that are inversely congruent to triangle ABC can be obtained from each other by rotations. To differentiate between

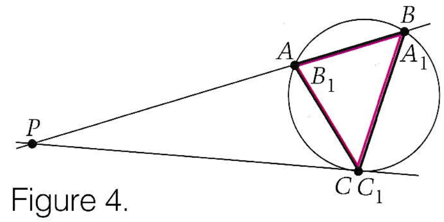


Figure 4.

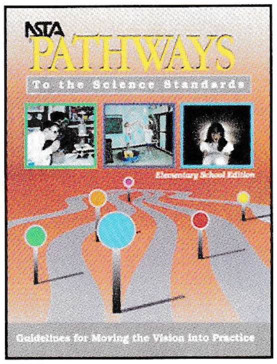
A particular case: Point P *appears when the triangles coincide.*

these cases, we will write the letter R or S above the equality sign. For each of $2 \cdot 6 = 12$ alternatives, we can use problem 2 and construct at most one desired point P.

This reasoning can be explained as follows. We take a triangle T made of cardboard (with the same circumradius as $\triangle ABC$), place it on the plane on one of its sides or another, mark the vertices A_1, B_1, and C_1 (6 different alternatives), place its vertices on the circle, and find point P for each of the $2 \cdot 6 = 12$ alternatives by rotating this triangle.

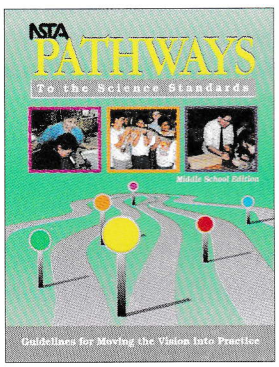
To finish solving problem 1, we must explain why four alternatives are excluded in the case $T = \triangle ABC$. One of them ($\triangle ABC \overset{S}{=} \triangle A_1B_1C_1$) can be eliminated at once, because in this case, for each of the three sides, special case (1) of the lemma occurs (fig. 3)—lines AA_1, BB_1, and CC_1 never meet at a point as triangle $A_1B_1C_1$ rotates. Also, the condition that none of the points A_1, B_1, and C_1 coincide with the corresponding points A, B, and C eliminates the case $\triangle BAC \overset{R}{=} \triangle A_1B_1C_1$ (fig. 4) and two similar cases: $\triangle ACB \overset{R}{=} \triangle A_1B_1C_1$, $\triangle CBA \overset{R}{=} \triangle A_1B_1C_1$. In these cases, point P appears at the moment when the triangles coincide (in this case, special cases (1) and (2) of the lemma both occur (fig. 4)). Problem 1 is solved.

For those who are patient enough to finish the analyses, we recommend thinking about the following questions: (1) Is it true that in the general case, all 12 alternatives (and in the case $T = \triangle ABC$, all 8 alternatives) are realized and give (as a rule) different points P? (Try experimenting with straightedge and compass.) (2) How much does the number 12 (or the number 8 in the particular case) decrease for an isosceles or equilateral triangle T?



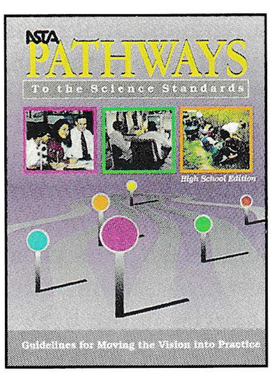
"Pathways is a roadmap, a model. It could get us closer to a strengthened science literate citizenry than any other guidelines currently available"

Dr. John R. Pancella
Science Education Consultant



"I am delighted to report that this book may be a means of rescuing many practitioners from the overwhelming nature of the task of reform before them."

Dr. Dennis W. Cheek
Rhode Island Department of Education, Director



Bring the *Science Standards* to life with the NSTA Pathways Series

Guidelines for Moving the Vision into Practice

Each practical guidebook demonstrates how you can carry the vision of the *Standards*—for teaching, professional development, assessment, content, program, and system—into the real world of the classroom and school. *Pathways* is also a tool for you to use in collaborating with administrators, school boards, and other stakeholders in science education. Filled with specific suggestions and clear examples that guide teachers on implementing each of the standards, *Pathways* is a valuable resource for everyone involved in science education. (Elementary School, Middle School, and High School Editions available)

EACH PATHWAYS BOOK ADDRESSES:

Science Teaching Standards
Professional Development Standards
Assessment Standards
Content Standards
Program Standards
System Standards

PRICING FOR NSTA PATHWAYS

Each Edition of NSTA Pathways
$34.95

▲ Elementary School #PB124X
(grades K–6, 1997, 152 pp.)

▲ Middle School #PB125X
(grades 5–9, 1998, 156 pp.)

▲ High School #PB126X
(grades 9–12, 1996, 196 pp.)

Resources for the Road CD-ROM (MAC/WIN)
#PB146X $19.95

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

To order, call 1–800–722–NSTA

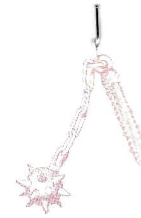
Resources for the Road CD-ROM

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

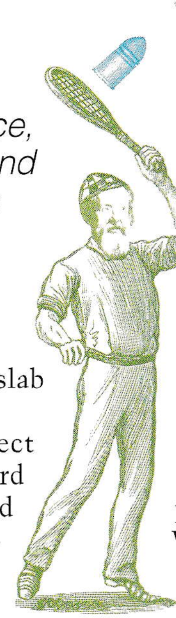
P1/299Q2

COLLID
SCOP

Interacting



"I concluded that the question on the force of collision looks very obscure, and nobody who previously considered this problem could penetrate into its essence, that is full of darkness and far from usual human ideas."
—Galileo Galilei



IT IS NO WONDER THAT THE problem of collisions was very difficult for Galileo. After all, he was a founder of the dynamics of only a single body. On the contrary, Christiaan Huygens, who modestly said that he merely "supported and generalized" Galileo's theory, greatly advanced mechanics, because he started to formulate the dynamics of several interacting bodies. Having elucidated such a complex phenomenon (by the way, it was a disaster for Cartesian mechanics), Huygens formulated the law of conservation of momentum more accurately and virtually discovered the law of conservation of mechanical energy. His achievements opened the door to the further study of collisions, which was done by such celebrities as Edme Mariotte, Thomas Young, Siméon-Denis Poisson, and Heinrich Hertz.

When beginning to study different collisions, you must learn how to distinguish between elastic and inelastic impacts, to reveal the connections between the concepts of deformation and wave propagation, and to feel the difference between the mechanical collision of macroscopic bodies and the interaction of atoms or elementary particles.

All this is prerequisite if you want to understand how a ball is kicked or a nail hammered, how piles are driven into soil or how iron swords are forged. There are also such phenomena as "reversed" collisions (gunshots and explosions). Collisions are very important because they underlie the behavior of molecules of gas and the scattering of atoms that pierce a thin metal foil. Ionization of atoms and the interaction of a quantum of light with an electron also provide examples of collisions.

We hope that your own "collision" with this broad theme will not be absolutely elastic and that the despair of Galileo will be replaced by the insight of Huygens.

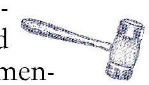
Questions and Problems

1. Why does a steel ball bounce much higher from a marble slab than from asphalt?

2. Why does a fragile object break when dropped on a hard floor, but lands safe and sound when the floor has a soft covering?

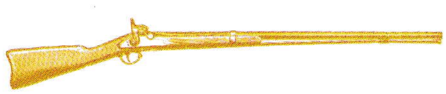
3. Several people can move a stationary bus, but the bus doesn't move when an anti-tank missile pierces it through. How can this happen if the force acting on the bus is much larger than the force acting in the first case?

4. In a circus performance an athlete lies under a heavy anvil. A colleague strikes the anvil with a hammer. Is this trick really dangerous for the athlete?



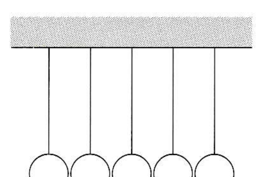
5. When a hammer strikes a piece of steel, it bounces away, but when it strikes a piece of lead, it recoils much less. Which piece of metal was given more energy?

6. In which case would a rifle fire a bullet a longer distance—when it is firmly fixed in a vise or suspended by strings?



7. Why doesn't a soldier firing a bazooka (grenade launcher) feel a recoil?

8. In Newton's collision toy, the five identical steel balls are suspended by strings so that they lie along a line and touch each other. How will this set of balls behave if the rightmost ball is pulled aside and then released? What will happen if the same procedure is performed simultaneously with two or three balls?



9. A ball falls vertically onto a smooth wedge that forms a 45° angle with the horizontal. What will its trajectory be after an elastic collision with the wedge if the wedge is initially at rest?

10. Why is it difficult to kick an underinflated soccer ball a long way?

11. When an experienced basketball player catches a fast moving ball, he relaxes his hands and moves slightly backward with the ball. Why?

12. What is the principal difference between the force of a rocket and the force of an ordinary engine?

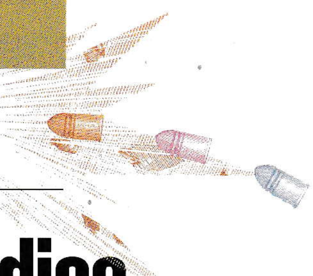
13. A projectile fired from a gun at some angle to the horizon explodes at the top of its trajectory into two fragments of equal mass. One fragment returns to the gun along the projectile's trajectory. Where will the other fragment land?

14. Under normal conditions, gas molecules have speeds of hundreds of meters per second. Why does the diffusion of gases proceed rather slowly?

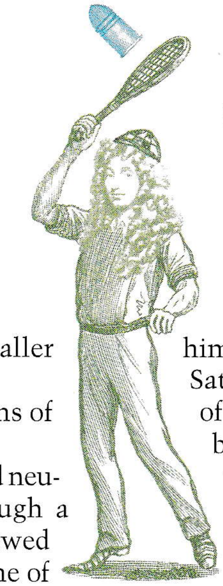
15. Why is the Brownian movement of small suspended particles more pronounced than similar motion of larger suspended particles?

16. Why do electric discharges in

ng bodies



"When an immobile body is stricken by a similar moving body, than the latter comes to a standstill, while the former starts to move with the velocity of the incident body."
—Christiaan Huygens

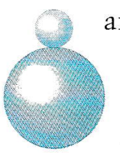


rarefied air occur at smaller voltages?

17. How can the atoms of a gas be excited?

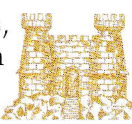
18. Why do high-speed neutrons easily pass through a block of lead but are slowed down in the same volume of paraffin, water, or other substances containing hydrogen atoms?

Microexperiment



Hold a small rubber ball and a large one, setting the small ball atop the large one, and drop them both. How will they behave after colliding with the floor? Why?

It's interesting that ...

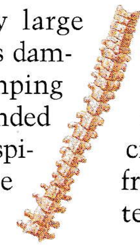


... in the Middle Ages, castles were assaulted with the help of a battering ram made of a log with a mass of several hundred kilograms. The warriors rushed to the gate of a castle, holding the log on their shoulders. At the gate they stopped abruptly and released the log, which continued its motion by sliding on the warriors' leather shoulder plates.

... even before Huygens, the Czech scientist Johannes Marci studied collisions and classified bodies as soft, crumbly, and hard, while René Descartes, who distinguished between hard and soft materials, could not see any difference between elastic and inelastic bodies.

... the variety of interests of Christiaan Huygens is attested to not only by his invention of the free pendulum clock and by constructing an excellent telescope that helped him discover a satellite and a ring of Saturn, but also by his attainment of a doctoral degree in law. Shortly before his death, he wrote one of the first textbooks on astronomy, *Cosmotheoros*, the Russian translation of which was performed at the behest of Peter the Great.

... Sir Isaac Newton initially formulated his third law only as a working hypothesis needed to construct mechanics. He thoroughly checked it in his experiments with collisions of pendulums.



... collisions evoke very large forces that can inflict serious damage. For example, when jumping onto hard ground with extended legs, one can traumatize the spinal cord, even when the height of the jump is just slightly more than 1 m, because too great a load is exerted on the cord at the moment of impact.

... under normal conditions, an oxygen molecule travels only 1/20,000 mm between collisions. However, this distance is rather large when compared to the size of the molecule: The proportion is the same as if a billiard ball traveled 10 m before hitting its target.

... to extract fillings from almost inaccessible places in teeth, dentists once used a clever contrivance. A filling was hooked by a rod, which had a sliding load. The load was lifted and then released. When the load struck the support, the resulting strong jerk pulled out the filling.

... although alpha particles approaching stationary atomic nuclei do not contact them, the model of completely elastic collisions nevertheless accurately describes the scattering of these particles by the nuclei.

... at room temperature, collisions between atoms are mostly elastic: They start to be excited in the collisions at temperatures of tens of thousands of degrees. On the contrary, almost all mutual collisions of elementary particles are elastic in the known temperature (energy) range. This is why these particles are considered to have no internal structure.

... when in 1932 Sir James Chadwick investigated the properties of uncharged particles emanating from a piece of beryllium, he could not detect them directly. However, using collisions of these particles with nuclei of other elements, he found all of the parameters of the unknown particle. This is how the neutron was discovered.

... when any moving object disintegrates—a projectile, rocket, atomic nucleus—, the center of mass of its fragments moves along the same trajectory that the intact body would have taken. This is why nuclear physicists prefer to study collisions in the frame of reference fixed to the center of mass of the colliding particles.

—A. Leonovich

ANSWERS, HINTS & SOLUTIONS ON PAGE 50

Quantum **articles about collisions**

A. Eisenkraft, L. D. Kirkpatrick, "Click, Click, Click...," September/October 1990, pp. 41–42.

M. Digilov, "A Strange Box and a Stubborn Brit," March/April 1991, pp. 26–27.

S. R. Filonovich, "The Power of Likeness," September/October 1991, pp. 23–27.

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

R. Vinokur, "The Science of the Jump-Shot," January/February 1993, pp. 46–50.

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

Sportin' life

by Arthur Eisenkraft and Larry D. Kirkpatrick

> *Lots of play*
> *in the way things work,*
> *in the way things are.*
> *History is made of mistakes.*
> *Yet—on the surface—*
> *the world looks OK,*
> *lots of play.*
> *—Gary Snyder*

MICHAEL JORDAN MAKES it all look so easy. The ball tossed at an angle θ with an initial velocity v from a height h gracefully glides in its arc and swishes through the net. All that polish from years of practice and no formal physics.

What are we to do? Can our mathematical approach help us to replicate Jordan's skills? Definitely not. But our analysis can help us appreciate the skill of someone who can sink the jump shot. In fact, our analysis can then be used to mimic the work of the broad jumper, the javelin and discus thrower, the volleyball spiker, the football punter, the soccer midfielder, and the baseball batter.

Trajectories without air resistance follow the simple equations of kinematics for horizontal motion with no acceleration and vertical motion with the acceleration due to gravity:

$$x = v_0 t \cos\theta$$
$$y = -\frac{1}{2}gt^2 + v_0 t \sin\theta.$$

From these equations we can derive an equation for the range of a trajectory thrown from the ground and returning to the ground. For this special case, the vertical displacement y is zero. By eliminating the time t, we obtain

$$x = \frac{2v_0^2}{g}\cos\theta\sin\theta = \frac{v_0^2}{g}\sin 2\theta.$$

The maximum range of a trajectory is now proven to be 45°, since $\sin 2\theta$ is equal to 1 for this value.

If we do not restrict the vertical displacement to zero, we find that

$$y = \frac{-gx}{2v_0^2\cos^2\theta} + \frac{x\sin\theta}{\cos\theta}.$$

This now demonstrates that for any given initial velocity and any angle, the path of a thrown object must be a parabola!

Using the identity relations

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

and

$$\tan^2\theta + 1 = \sec^2\theta,$$

we can derive a new equation that will help us find the angle a ball must be thrown to reach a specific point in space:

$$y = \frac{-gx^2}{2v_0^2}\tan^2\theta + x\tan\theta - \frac{gx^2}{2v_0^2}.$$

Since the equation is quadratic, we see that we can reach any position with two different angles.

The trajectory equations can provide us with some insights to many of our sports dilemmas. The field-goal kicker can certainly use the equations to determine the range of angles that will provide his team with three points. The punter has a different job requirement. He wants to kick the ball as far as possible. But if he kicks it so far that his players can't get downfield, then the return will negate his good punt. He must maximize that distance and maximize the hang time. Maximizing the distance requires a kick at 45°, yet maximizing the hang time requires a vertical kick (with no down field component). What is a punter to do?

These decisions must be very difficult, and experience must guide the punter. Physics can provide some help. If we know the speed of the defenders as they run down the field and we know the initial velocity of the punt, we can suggest adjusting the angle of the kick so that the ball arrives when the defenders do. We will analyze this while ignoring air resistance.

The range of the punt traveling at an initial velocity v_0 and angle θ and the corresponding time of flight are given by the following equations:

$$R = \frac{2v_0^2}{g}\sin 2\theta,$$
$$t = \frac{2v_0\sin\theta}{g}.$$

Art by Tomas Bunk


VIVA LA EVOLUTION
START
ONE WAY
SPQR
TOMAS BUNK

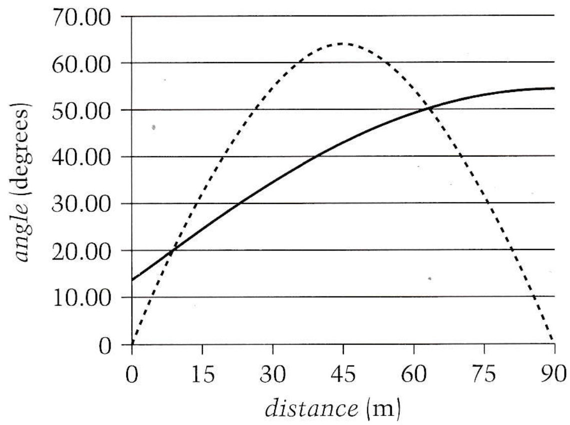


Figure 1

If the offensive team can run downfield at a speed v_r, they can travel a distance x downfield in the time t, where $x = v_r t$. The punter kicks the ball from a position 15 yards behind the line of scrimmage. If we minimize the difference between the range of the ball R and the distance the runners travel, taking into account the 15-yard lead the runners get, we will determine the optimum angle for the punt.

The corresponding equations can most readily be solved using a solver on a calculator, a spreadsheet, or graphing the equations for R and $(x + 13.7 \text{ m})$. Figure 1 shows typical values of 25 m/s for the ball and 8 m/s for the runners. The runners will arrive as the ball arrives when the ball is kicked at an angle of 65° and travels a distance somewhat less than its maximum range.

There are some interesting trajectory problems where the physics can also provide assistance. These will make up the contest problems for this month.

1. (a) A free kick is being set up in soccer. The defenders form a wall with their bodies between the kicker and the goal. The ball must clear the players. The defenders are 1.8 m tall, and they set up the wall 15 m from the kicker, who kicks the ball at 35 m/s. At what locations can the ball *not* land?

(b) How does this shadow region change as the wall moves in relation to the kicker?

2. (a) A basketball player shoots a jump shot with an initial velocity v_0 at an angle θ_0. For a given basket that is h meters above the release point of the ball and L meters horizontally from the basket, determine the relationship between v_0 and θ_0.

(b) Since the ball must enter the basket during its descent, describe this constraint on the initial angle mathematically.

(c) At what angle is a minimum speed required to sink the shot?

Doppler beats

In the July/August 1998 issue we posed a problem combining beats and Doppler shifts. The problem was designed by Leaf Turner, one of the coaches of the U.S. Physics Team, to help select members of the 1998 team. This problem was successfully solved by Zach Frazier, who graduated last spring from Ferris High School in Spokane, Washington, and by Stephen Hanzely from Youngstown State University in Youngstown, Ohio.

Two sirens are located along the x-axis with frequencies f_L and f_R for the left and right sirens, respectively. An observer moving with speed v_o along the x-axis hears frequencies 0.99 Hz, 0 Hz, and 1.01 Hz on the left-hand side of the sirens, between the sirens, and on the right-hand side of the sirens, respectively.

A. In the absence of any motion, the beat frequency Δf is a frequency source, just like any other type of source. Therefore, we see that the frequency is red-shifted on the left and blue-shifted on the right. This means that the observer is moving from right to left.

B. On the left-hand side, the observed frequency $\Delta f_L'$ is given by

$$\Delta f_L' = \Delta f\left(1 - \frac{v_o}{v_s}\right) = 0.99 \text{ Hz},$$

where v_s is the speed of sound. Likewise, on the right-hand side, the observed frequency $\Delta f_R'$ is given by

$$\Delta f_R' = \Delta f\left(1 + \frac{v_o}{v_s}\right) = 1.01 \text{ Hz}.$$

Dividing these two equations yields

$$\frac{\Delta f_R'}{\Delta f_L'} = \frac{1 + \frac{v_o}{v_s}}{1 - \frac{v_o}{v_s}} = \frac{1.01}{99},$$

from which we get $v_o = 0.01 v_s$.

C. When the observer is between the two sources, there is no beat frequency, and the observer measures the same frequency from both sources. However, the source on the left is blue-shifted, and the source on the right is red-shifted. Therefore, $f_R > f_L$.

D. Numerically, we have

$$f_L' = f_L\left(1 + \frac{v_o}{v_s}\right)$$

and

$$f_R' = f_R\left(1 - \frac{v_o}{v_s}\right).$$

Setting the two shifted frequencies equal to each other, we get

$$\frac{f_R}{f_L} = \frac{1 + \frac{v_o}{v_s}}{1 - \frac{v_o}{v_s}} = \frac{1.01}{99}.$$

From either of the first two equations in part B and using $v_o = 0.01 v_s$, we obtain

$$f_R - f_L = 1 \text{Hz}.$$

Solving these two simultaneous equations provides us with the numerical values $f_R = 50.5$ Hz and $f_L = 49.5$ Hz.

Cold boiling

by S. Krotov and A. Chernoutsan

EVERYBODY KNOWS THAT to boil water, we must heat it. But could it be possible to boil water by cooling it? At first glance this seems impossible. However, do not hurry with your answer. Carry out the following simple experiment and think about how to explain it.

You will need a 30–40 mL test tube with a tight plug, a Bunsen burner, and a test-tube holder. Also prepare a bottle of room-temperature water and a bottle of ice-cold water.

Wearing gloves and goggles, pour the room-temperature water into the test tube until it is a little more than half full. Start to warm the test tube over the burner, making sure to hold the test tube at an angle (and pointing away from any people) and to heat the upper part of the water column. *If the water is warmed at the bottom of the tube, the expanding vapor can shoot the contents of the tube into the air.*

Wait until the water reaches a steady boil, then quickly and firmly plug the tube, simultaneously removing it from the flame. As expected, the boiling stops immediately. Now turn the tube upside down and pour the room-temperature water over the upper (empty) part of the tube. What do you think will occur? The water in the tube starts boiling again! Of course, after a while, the boiling stops, but pouring room-temperature water over the test tube again will cause boiling to start again. When pouring room-temperature water over the tube no longer causes boiling (at this time the tube is cold enough to hold in your hand), pour ice-cold water over the empty part of the test tube. The water in the cool tube will boil once more! Can you provide a reasonable explanation of this so-called "cold boiling"?

ANSWERS, HINTS & SOLUTIONS ON PAGE 51

Art by Sergey Ivanov

As easy as (a, b, c)?

Searching for Pythagorean triples

by S. M. Voronin and A. G. Kulagin

HOW CAN WE FIND POSITIVE integer solutions to the following famous equation:

$$a^2 + b^2 = c^2? \tag{1}$$

Specific solutions to this equation were known long before mathematics became a science. Ancient Babylonians knew the solution $(a, b, c) = (3, 4, 5)$ and a number of other solutions to equation (1), including some that are difficult to find, such as (105, 36, 111) or (12709, 13500, 18541).

Although various viewpoints of pre-Greek mathematics exist, it is not likely that the Babylonians used the deductive methods of mathematics. Mathematics as a deductive science first appeared in ancient Greece in the sixth century B.C. Tradition ascribes the first statements of mathematical problems to Pythagoras. The mathematicians of ancient Egypt and Babylon could calculate the number $\sqrt{2}$ very accurately, but the statement of the problem that $\sqrt{2}$ is irrational may still have been alien to them. Equally alien to them would be the statement of the problem of the description of all solutions to equation (1). This problem was set and solved by the Pythagorean school.

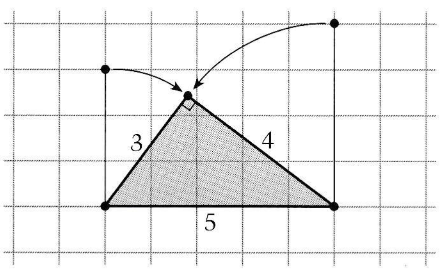


Figure 1

For this reason, and possibly because of its clear relationship to the Pythagorean theorem as well, equation (1) is called the *Pythagorean problem*, and the triples of natural numbers satisfying this equation are called *Pythagorean triples*. It follows from the Pythagorean theorem that a Pythagorean triple $0 < a < b < c$ can be assigned to a right triangle with integer legs a, b, and hypotenuse c. Conversely, each right triangle with integer sides supplies a solution to equation (1). Thus, the Pythagorean problem has a clear geometrical interpretation (fig. 1).

Arithmetic method

If a triple of natural numbers (a, b, c) is Pythagorean, the triple (ka, kb, kc) is also Pythagorean for any positive integer k. Therefore, to solve the Pythagorean problem, it is sufficient to find all primitive triples (a, b, c)—that is, Pythagorean triples such that the numbers a, b, and c have no common divisor.

It turns out that it is rather easy to enumerate all primitive Pythagorean triples. The Pythagoreans knew a simple method based on the following proposition: *If* p *and* q *are coprime and of different parity and* p > q, *then the triple of numbers*

$$a = p^2 - q^2,\ b = 2pq,\ c = p^2 + q^2 \tag{*}$$

is a primitive Pythagorean triple.

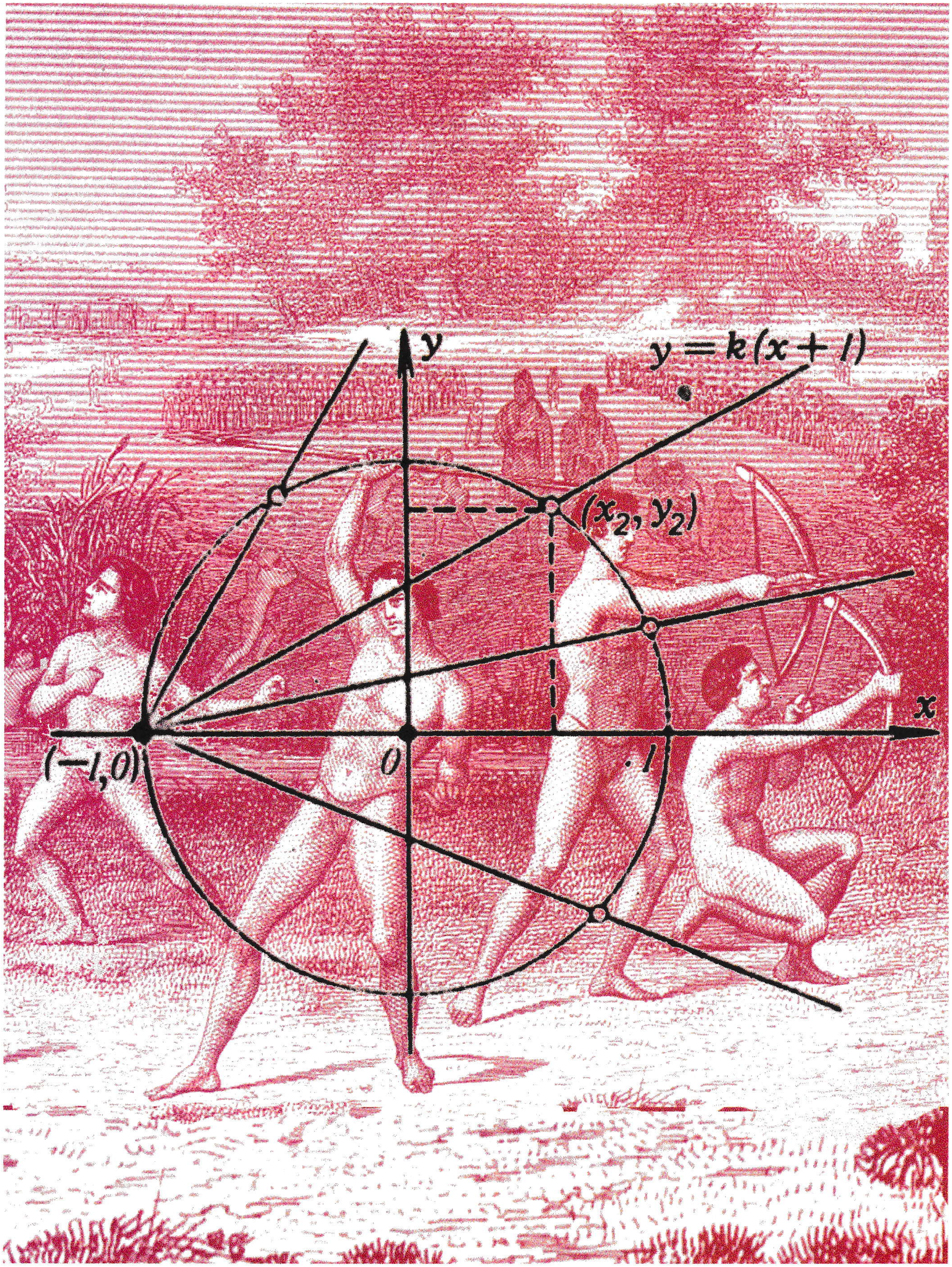
For example, if $q = 2 \cdot 3 \cdot 7 = 42$ and $p = 163$, we obtain the following Pythagorean triple: (24805, 13692, 28333).

The proof of the proposition is clear. Indeed, by virtue of (*),

$$\begin{aligned} a^2 + b^2 &= (p^2 - q^2)^2 + 4p^2q^2 \\ &= (p^2 + q^2)^2 = c^2. \end{aligned}$$

That is, the triple (a, b, c) is Pythagorean. In addition, it is primitive: If a, b, and c had a common divisor, this would also be a common divisor of the numbers $c + a = 2p^2$ and $c - a = 2q^2$, which contradicts the fact that p and q are relatively prime.

Art by Vera Khlebnikova


y
$y = k(x+1)$
(x_2, y_2)
x
$(-1,0)$
0
1

We have proved that the method described produces a primitive triple. It turns out that *all Pythagorean triples can be obtained by using method* (*).

Next we give (in brief) an arithmetical proof of this remarkable fact. The proof explains how it would be possible to guess formulas (*). (Those readers who do not like arithmetical reasoning can skip this proof and go to the next section.)

Let a primitive triple (a, b, c) be given. The condition of primitivity can be written as

$$\mathrm{GCD}(a, b, c) = 1, \qquad (2)$$

where GCD denotes the greatest common divisor of the numbers in parentheses. Note that we must consider the divisors that are common to all three numbers a, b, and c and not pairwise common divisors. Equality (2) does not imply that

$$\mathrm{GCD}(a, b) = \mathrm{GCD}(b, c) = \mathrm{GCD}(c, a) = 1. \qquad (3)$$

However, for any primitive Pythagorean triple, relations (3) do hold. Indeed, suppose, for example, $\mathrm{GCD}(a, b) = k > 1$. Then, by virtue of equation (1), c^2 is divisible by k^2, and thus, c is divisible by k, which is impossible because of (2).

For any primitive triple (a, b, c), the numbers a and b are of different parity. Indeed, equation (3) tells us that they cannot both be even. If they are both odd, say $a = 2k + 1$ and $b = 2l + 1$, then

$$a^2 + b^2 = 4(k^2 + l^2 + k + l) + 2.$$

Thus, $a^2 + b^2$ is divisible by 2 but not by 4. By equation (1), c has the same property. That is, c is even, so $c = 2s$ for some integer x. Consequently, $c^2 = 4s^2$ is divisible by 4. Thus, we have a contradiction!

We have established that a and b are of different parity. For definiteness, let a be odd, b be even, and c be odd.

Now the expressions $c + a$ and $c - a$ are both even, so we can write $c + a = 2m$, $c - a = 2n$ for some integers m and n. The equation can be rewritten as

$$b^2 = c^2 - a^2 = (c + a)(c - a) = (2m)(2n) = 4mn. \qquad (4)$$

(Expressions $(c + a)$ and $(c - a)$ can be written as $2m$ and $2n$ because they are even.)

Lemma. The numbers m and n are perfect squares,

$$m = p^2, n = q^2,$$

of two coprime numbers p and q of different parity.

We invite the reader to prove this lemma. (Hint: Prove that m and n are relatively prime using (3), factor them into prime factors, and substitute into (4).)

Recalling that $2m = c + a$ and $2n = c - a$, we obtain

$$2p^2 = c + a, 2q^2 = c - a,$$

from which we get $c = p^2 + q^2$ and $a = p^2 - q^2$. Using (1), we obtain $b = 2pq$. Thus, if a, b, and c are a primitive Pythagorean triple, then it can be obtained from formulas (*).

Geometric method

Using coordinates we can give a geometric interpretation of equation (1). All the solutions to (1) can be obtained from the primitive solutions by multiplying by a natural number k. That is, these solutions have the form (ka, kb, kc), where (a, b, c) is a primitive Pythagorean triple. It follows from (1) that

$$\left(\frac{a}{c}\right)^2 + \left(\frac{b}{c}\right)^2 = 1. \qquad (5)$$

Since $a, b, c > 0$ are integers, the numbers a/c and b/c are rational. The equation of the unit circle is $x^2 + y^2 = 1$, and thus there is a point on the unit circle with rational coordinates corresponding to every primitive solution (a, b, c) of equation (1). Points with rational coordinates are called *rational points*. Conversely, if we have a rational point (x, y) with coordinates $x = m_1/n_1$ and $y = m_2/n_2$ on the unit circle, we can obtain equation (5) by reducing the fractions m_1/n_1 and m_2/n_2 to the least common denominator $c > 0$. This equation gives a primitive triple. Thus, *there is a one-to-one correspondence between the rational points on a unit circle and the primitive Pythagorean triples.*

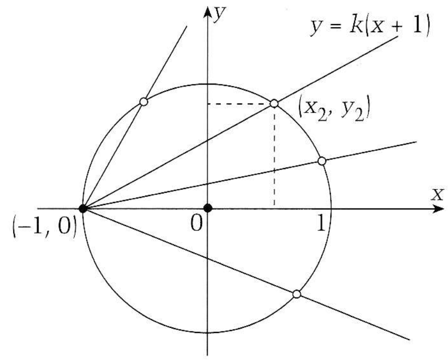


Figure 2

We have obtained a geometric formulation of the Pythagorean problem: *Find all rational points on the unit circle* $x^2 + y^2 = 1$.

Let's try to find all such points. Draw lines through the point $(x_1, y_1) = (-1, 0)$. Any line that passes through the point $(-1, 0)$ and is not tangent to the circle intersects the circle at one more point, point (x_2, y_2), for example (fig. 2). The equation of such a line is

$$y = k(x + 1),$$

where k is the slope of the line. Thus, the coordinates of the point (x_2, y_2) satisfy the system of equations

$$\begin{cases} x^2 + y^2 = 1, \\ y = k(x + 1). \end{cases}$$

Solving this system for x and y and substituting the value $y = k(x + 1)$ from the second equation into the first one, we obtain

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

or

$$(1 + k^2)x^2 + 2k^2x + k^2 - 1 = 0.$$

If x satisfies the last equation, then x is the abscissa of the intersection point of the line $y = k(x + 1)$

with the unit circle. In other words, either $x = x_1 = -1$ or $x = x_2$. Using the usual expression for the sum of the roots of a quadratic equation, we have

$$x_1 + x_2 = -\frac{2k^2}{1+k^2},$$

from which we obtain

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

Since the point (x_2, y_2) lies on the line $y = k(x + 1)$, we have

$$y_2 = \frac{2k}{1+k^2}.$$

The last two formulas assign a point

$$(x_2, y_2) = \left(\frac{1-k^2}{1+k^2}, \frac{2k}{1+k^2}\right), \qquad (6)$$

of the circle $x^2 + y^2 = 1$ to each number k. Conversely, a unique value of the slope, k, corresponds to any point $(x, y) \neq (-1, 0)$ of the unit circle, namely, the value

$$k = \frac{y}{x+1}. \qquad (7)$$

Note that if k is rational, the point (x_2, y_2) defined by (6) has rational coordinates, and, conversely, if x_2 and y_2 are rational, k is also rational by virtue of (7). Thus, we have the following proposition: *A one-to-one correspondence exists between the points with rational coordinates on the unit circle (except for the point (–1, 0)) and rational numbers.*

Making k run through all values in the interval from $-\infty$ to $+\infty$, we can enumerate all points with rational coordinates on the unit circle (except for the point (–1, 0)) and thus, find all primitive solutions to the Pythagorean problem.

Let's write down the corresponding formulas. Let $k = p/q$, $q > 0$ and $\mathrm{GCD}(p, q) = 1$. Then, it follows from (6) that

$$(x_2, y_2) = \left(\frac{q^2-p^2}{q^2+p^2}, \frac{2pq}{q^2+p^2}\right). \qquad (**)$$

In essence, these formulas are equivalent to (*)! To obtain exactly the same result, we must analyze the parity of p and q.

If p and q are of different parity, formula (**) corresponds to the primitive Pythagorean triple $(a, b, c) = (q^2 - p^2, 2pq, q^2 + p^2)$. If p and q are both odd, then, passing to new variables $q_1 = (q + p)/2$, $p_1 = (q - p)/2$, $\mathrm{GCD}(p, q) = 1$, we obtain

$$q^2 + p^2 = 2(q_1^2 + p_1^2),$$
$$q^2 - p^2 = 4p_1q_1,$$
$$2pq = (q_1^2 - p_1^2).$$

Reducing the fractions in (**) by 2, we obtain

$$(x_2, y_2) = \left(\frac{2q_1p_1}{q_1^2+p_1^2}, \frac{q_1^2-p_1^2}{q_1^2+p_1^2}\right). \qquad (8)$$

Note that (8) differs from (**) only by the permutation of x_2 and y_2. If it turns out that p_1 and q_1 are both odd, then we pass to the variables

$$q_2 = (q_1 + p_1)/2,$$
$$p_2 = (q_1 - p_1)/2,$$
$$\mathrm{GCD}(p_2, q_2) = 1,$$

and so on. As a result, we obtain either a formula of type (**) or of type (8), where p_n and q_n are of different parity and $\mathrm{GCD}(p_n, q_n) = 1$. Thus, the primitive solution that corresponds to the rational point $(x_2, y_2) \neq (-1, 0)$ is given either by formula (**) or (8) in which p and q are coprime numbers of different parity. By virtue of the one-to-one correspondence between primitive solutions to the Pythagorean equation and rational points of the unit circle (except for the point (–1, 0)), formulas (**) and (8) imply formulas (*).

Rational parameterization of conics

We have obtained formula (6), which can be used to find all rational points on the unit circle $x^2 + y^2 - 1 = 0$. This formula establishes a one-to-one correspondence between the parameter k that takes all real values and the points of the circle (except for the point (–1, 0)); the coordinates of points (x, y) of the circle are rational functions of k. A natural question arises: Can this method be used to produce points lying on other curves, for example, on an ellipse, a parabola, or a hyperbola? (Ellipses, parabolas, and hyperbolas are called conic sections.)

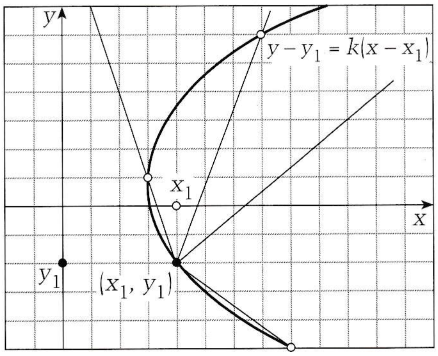


Figure 3

To answer this question, consider a curve on the plane given by equation $K(x, y) = 0$, where $K(x, y)$ is a quadratic polynomial in x and y. Ellipses, parabolas, and hyperbolas are determined by such equations. Let (x_1, y_1) be a fixed point on a curve of this type, and with rational coordinates. Draw a line with a slope k through this point (fig. 3). We seek intersection points, (x, y), of this line with the curve. The coordinates of these points satisfy the following system of equations:

$$\begin{cases} K(x,y) = 0, \\ y - y_1 = k(x - x_1). \end{cases}$$

Solving this system for (x, y), as we have done for the circle, we represent the coordinates (x_2, y_2) of the second point of intersection of the line $y = y_1 + k(x - x_1)$ with curve K in terms of the parameter k. It is easy to verify that

$$(x_2, y_2) = \left(\frac{A(k)}{C(k)}, \frac{B(k)}{C(k)}\right), \qquad (***)$$

where $A(k)$, $B(k)$, and $C(k)$ are polynomials of parameter k of the order not exceeding 2. This formula gives *a rational parameterization* of the curve K. That is, it represents the coordinates of every point of the curve in terms of rational functions of a single parameter k.

Formula (***) makes it possible to find integer points on K. However, now we are interested in another, quite unexpected, application of this formula.

Calculating integrals

It turns out that integrals that look hopeless at first glance, such as

$$I = \int \frac{dx}{\sqrt{x^2 + 3x - 4}}$$

can be calculated by using a rational parameterization of an appropriate curve. In this particular case, we can use the curve

$$K(x, y) = y^2 - (x^2 + 3x - 4).$$

We note that the point $(x_1, y_1) = (-4, 0)$ lies on the curve

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

Consider all lines

$$y = k(x + 4)$$

to obtain the following parameterization (carry out all the manipulations yourself):

$$(x, y) = \left(\frac{1 + 4k^2}{1 - k^2}, \frac{5k}{1 - k^2}\right).$$

Then,

$$dx = \frac{10k}{\left(1 - k^2\right)^2} dk,$$

and thus, we obtain

$$I = \int \frac{dx}{\sqrt{x^2 + 3x - 4}} = \int \frac{dx}{y(x)} = \int \frac{2dk}{1 - k^2}.$$

The last integral is relatively easy to calculate:

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

$$= \ln|1 + k| - \ln|1 - k| + C = \ln\left|\frac{1 + k}{1 - k}\right| + C.$$

Recall that

$$k = \frac{y}{x + 4} = \frac{\sqrt{x^2 + 3x - 4}}{x - 4}$$

to obtain the final result (carry out all the manipulations):

$$I = \ln\left|\frac{x + 4 + \sqrt{x^2 + 3x - 4}}{x + 4 - \sqrt{x^2 + 3x - 4}}\right| + C.$$

This method can be used to calculate many other integrals. We invite the reader to think about which integrals can be thus calculated.

In this article we used the same method to solve Diophantine equations and calculate integrals, thus solving problems from different fields of mathematics. Pythagorean discreteness and Archimedean continuity turn out to be related. ◘

Another perpetual motion project?

by A. Stasenko

THE FIRST ATTEMPTS AT building perpetual motion devices appeared in eighteenth-century France. However, beginning in 1775, the French Academy of Sciences refused to consider perpetual motion projects.

Indeed, projects to create perpetual motion devices are usually scrapped to save wasted time and effort, because such engines defy the laws of nature. However, sometimes it is instructive to think about whether a certain machine should be considered a perpetual motion device.

One thoughtful student heard that molecules can lose momentum in collisions with walls. In completely elastic collisions they bounce off with the same (and opposite) velocity, but in completely inelastic collisions they lose the normal component of their velocity and end up sliding along the surface of the wall. This is how the student conceived of an idea for her own perpetual motion device.

She decided to devise a plate in which completely elastic molecular collisions would occur on one side, and completely inelastic collisions on the other side. She made two such plates and attached them to a weightless rod so that their matching surfaces faced in opposite directions. Then she fixed the rod onto a frictionless vertical axis. Proud of her work, she drew a diagram of it. Figure 1 shows the overhead view: The area of each disk is S, the mean thermal velocity of the surrounding gas molecules is v, and the speed of each disk is u. The circular arrow marks the expected direction of rotation of the device. The student set the left surface of the upper disk to reflect molecules elastically and the right surface to reflect them inelastically.

Our student knew that gas molecules move in all directions with equal probability, but she drew only the molecules moving to and from the disks. According to the estimates given in physics class, there are $n/6$ such molecules per unit volume (where n is the number density of molecules and 6 is the number of faces on a cube). Therefore, the molecular flux (that is, the number of molecules hitting a unit area per unit time) on the left side of the upper disk equals $(n/6)(v - u)$. By the way, sometimes it's useful to check even very simple formulas like this one. Indeed, at $u = v$ the molecules will never catch up with the disk or collide with it, making the molecular flux on the left side of the upper disk zero.

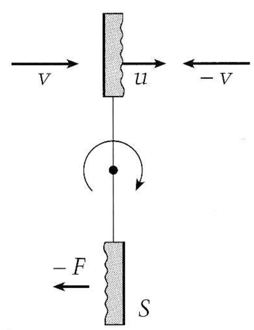


Figure 1

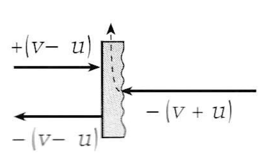


Figure 2

Every molecule hits the left side of the disk with velocity $+(v - u)$ in the disk's frame of reference, and according to the elastic nature of the impact, will bounce off with a velocity $-(v - u)$ of the same value and opposite direction (fig. 2). Therefore, in the laboratory frame of reference, the molecules' velocity after collision is $-(v - u) + u = -v + 2u$, and the respective change in momentum of a single molecule is $m(-v + 2u) - mv$. The upper disk will acquire the same momentum (in the other direction, of course). Multiplying this value by the corresponding molecular flux and by the area of the disk, we arrive at the force acting on the disk from the left side:

$$\frac{n}{6}(v-u)m(2v-2u)S = \frac{mnS}{6}2(v-u)^2.$$

Now let's make similar calculations for the right side of the upper disk. In the disk's frame of reference, the molecular velocity prior to impact is $-(v + u)$, and after the collision it is zero. In the stationary frame of reference, the postimpact velocity equals u, so the change in the molecules' velocity is $u - (-v) = u + v$. The molecular flux on the right side of the upper disk is $(n/6)(v + u)$, so the total force acting on the disk from this side is

$$-\frac{mnS}{6}(v+u)^2.$$

Let's take into account that the product mn equals the density of the gas ρ so that we can write the force affecting the upper disk as

$$F = \frac{\rho S}{6}\left(2(v-u)^2 - (v+u)^2\right)$$
$$= \frac{\rho S}{6}\left(v^2 + u^2 - 6vu\right).$$

An equal and opposite force acts on the lower disk, so the system will rotate in the direction shown in figure 1.

Clearly, each of the above forces will be zero (the system will have zero angular acceleration) when

$$u^2 - 6vu + v^2 = 0.$$

Our young investigator knew how to solve such a quadratic equation, so she found the steady-state speed:

$$u_\infty = 3v \pm \sqrt{9v^2 - v^2} = v\left(3 \pm 2\sqrt{2}\right).$$

Because the disk cannot move faster than the driving molecules, she chose only the solution with the negative sign.

Thus, the steady-state speed of the disks is

$$u_\infty = v(3 - 2\sqrt{2}) = 0.172v,$$

which is appreciably less than the thermal speed of the molecules, so our smart inventor knew to neglect the square of u_∞ in the formula for F. As a result, the equation (which states Newton's second law for the disk) is transformed to a linear differential equation:

$$\frac{du}{dt} = \rho\frac{vS}{m}\left(u - \frac{v}{6}\right) = \frac{u - v/6}{\tau}.$$

We can see some features of the solution even without solving this equation. It's clear that the acceleration decreases with increasing speed and becomes zero when $u_\infty = v/6 = 0.167v$. (Note that this value is close to the one we obtained earlier.) The constant value $\tau = m/\rho vS$ in the denominator on the right side of the equation is called the relaxation time. For example, when taking m = 1 g, $\rho = 10^{-5}$ kg/m^3, $v \cong 300$ m/s, and $S = 1$ cm^2, we have $\tau \cong 3 \cdot 10^3$ s $\cong 1$ h. This is the time to reach 63 percent of the maximum speed (fig. 3).

It should be noted here that our model is based on free molecular flow around the disks. For this reason, our inventor assumed the density of the gas to be five orders of magnitude less than that of air under normal conditions. In this case, the mean free path of each molecule increases by five orders of magnitude, and instead of 10^{-7} m becomes just 1 cm, which is comparable to the assumed size of the disk.

Thus, the device should rotate forever. It's even possible to supply it with a gear to perform useful work. However, this device has nothing to do with the kind of perpetual motion device that was rejected by the French Academy of Sciences. Our device doesn't try to produce energy "from nothing." The disks receive energy from air molecules, and the molecules acquire it in collisions with the walls of a vessel kept at a constant temperature. Therefore, our thermal motor does not violate conservation of energy.

There is an abundance of thermal energy in the world, so at first glance,

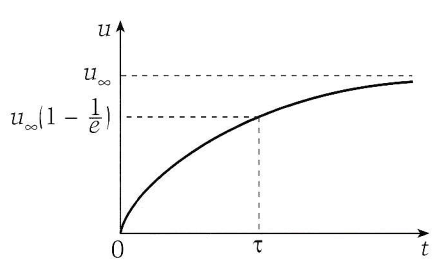


Figure 3

this "thermal motor" could take energy, say, from the ocean and use it to propel ships. But wait—there are still a few design issues to address before we get to that point. How do we obtain disks that reflect molecules elastically at one side and inelastically at the other? This is where our inventor had to proceed with caution. It's easy to take the wrong path and attempt to construct an impossible device, just as many before tried to produce energy from nothing.

Let's consider our "thermal motor" from a technical viewpoint. The elastic surface's mirrorlike reflectivity keeps it in thermal equilibrium with the surrounding gas, so that the molecules bounce off with the same speeds with which they hit the disk. The problem lies with the other side of the disk, which must be inelastic.

How do we make the surface inelastic with respect to the incident gas molecules? There are two approaches to the problem. In the first (mechanical) way, the surface is made porous: The molecules approaching the pores hit their mirrorlike walls and move deeper into the pores. Every pore curves gradually to a right angle, eventually ejecting the molecules out the side of the disk (fig. 2).

The alternative method could be called thermal. A certain liquid (say, liquid nitrogen or helium) keeps the inelastic surface cool. Since the mean energy of molecules bouncing off a surface is determined by the surface's temperature, here the mean energy of the departing molecules will be much less than that of the incident molecules.

Which method is better? At first glance, the second one is more simple and economical. We'd have no trouble replenishing the coolant. To maintain the device's rotation, we would need only to make sure the temperature of the vessel's walls doesn't fall too low (to maintain a temperature difference between the device and its surroundings).

We've just described a device that could supply humanity with a huge amount of cheap energy. Indeed, we could sink a billion of these devices into the ocean and pump energy from it. The energy loss from the ocean would be compensated by solar radiation. It's not clear why such devices (called Maxwell's demons) cannot work. However, the scientific answer to such projects is always an unequivocal "no."

Long ago, scientists understood that such "thermal motors," which are designed to work with only one thermal reservoir, do not in principle differ from classical perpetual motion; they were even termed the second type of perpetual motion. There were perhaps no fewer attempts to construct such devices than to invent the first type of perpetual motion device, but all efforts were fruitless. Of course, this futility wasn't accidental: The second law of thermodynamics stood in the way of numerous inventors. One formulation of this law, given by Kelvin and Plank, states that "no possible process can have as a sole external effect the work performed at the expense of heat taken from a reservoir at constant temperature." In other words, no possible cyclic engine can work using energy from a single thermal reservoir. Therefore, the mechanical approach to constructing the inelastic surface cannot be successful in principle. (Why, by the way?)

The thermal approach, on the other hand, doesn't violate the second law of thermodynamics. An engine designed with this approach contains not only a "heater" (the vessel's walls) but also a low-temperature sink (liquid nitrogen). The efficiency of such a device would probably not be high, though. Alas, there is no such thing as eternal and free energy.

Quantum articles on thermodynamic laws:

Kaleidoscope: "What a commotion," May 1990, 32–33.

A. Buzdin, "Keeping cool and staying put," May/June 1993, 17–20.

A. Savin, "Mathematics in perpetual motion," July/August 1994, 5–8.

The Lorentz/FitzGerald diet

by David Arns

When looking in the mirror one day, dismayed by what I'd seen,
"Svelte" was not the word that leapt to mind;
I seemed distinctly—thicker—than I think I'd ever been,
As I scanned both sides, the front, and the behind.

As I stood, annoyed at just how far I'd let it go,
I recalled an image seen some years before:
Something in an illustration drawn to let us know
That folks get *thin* as through deep space they soar!

"Yes!" I cried, "that's it! The little man inside the rocket
Got wondrous thin as past his friend he flew!"
With breakneck speed, my stopwatch in my hand, wherewith to clock it;
I'd drop off fifty pounds or more, I knew.

I hurried to my bookshelf, quickly found the physics section,
And felt the thrill that great discovery warrants;
The index had "FitzGerald," and additional reflection
Brought associated names like "Hendrik Lorentz."

In the early days of physics, back when "ether" was a fad,
Both Lorentz and FitzGerald did some math
To show why Michelson and Morley's famed experiment went bad,
And where they had departed from the path.

I read in fascination how FitzGerald's new equation
Would show how thin a person could become.
Then, my size would be no longer just a topic for evasion,
For I'd be thin! (At least, as seen by some.)

His equation's easy: just divide velocity by c
(Where c's the speed of light, I'm sure you know)
Then square it, and subtract from 1, and finally—here's the key—
Square root the difference and, well, there you go.

Let's see, now: if I wanted to arrive at half my size,
And revel in the thinness I'd attain,
To five hundred eighty million MPH, my speed would rise,
Or else from social gatherings I'd abstain.

You can bet I was excited, one hair's breadth from going out
To buy an ion-powered ship to fly,
When a paragraph whose subject matter also talked about
The Lorentz/FitzGerald contraction caught my eye.

Oh, why I read that paragraph, I'll never, ever know!
My plans were dashed in shards upon the floor!
The next thing that I learned is *time itself* begins to slow,
As velocity increases more and more.

This means, of course, my normal, laid-back, easy-going style
Would slow yet more, and I could ill afford
To talk much slower still, or 'twould be such a numbing trial
That folks would doze right off, completely bored!

And that was not the worst of it! That selfsame paragraph
Described how high-speed things tend to get *heavy*!
So even though I'd *look* like I was thin (don't make me laugh),
My mass would tend toward that of a Chevy!

Well, I came to the conclusion that a "diet" such as this
Is impractical, and thus, with grim defiance,
I admit that if I just eat less, my weight won't be amiss—
For, after all, this isn't rocket science!

Bohr's quantum leap

by A. Korzhuyev

THE TWENTIETH CENTURY is coming to a close. How will it be remembered—the age of electronics, aviation, or computers? The answer is not clear, but looking to the past, physicists of the twenty-first century will certainly praise the achievements of their predecessors in quantum theory, which is a child of discoveries related to the structure of the atom and to the principles of atomic "life."

Embryonic atomic theory

Although the atomic structure of matter was guessed at by ancient philosophers, the real experimental foundation of this concept was elaborated not so long ago. Let's begin our story with spectroscopy. In 1859 Gustav Kirchhoff and Robert Bunsen developed the method of spectral analysis and explained, among other phenomena, the origin of four dark absorption lines in the solar spectrum. These lines were discovered as far back as 1814 by Joseph von Fraunhofer, and now 45 years later they were shown to coincide rather closely with the bright lines in the light emitted by vapors and heated gases of different substances under normal laboratory conditions.

In 1885 Johann Balmer published a paper in which he found that the wavelengths of these lines could be described with good accuracy by the following formula:

$$\lambda = k\frac{m^2}{m^2 - 2^2},$$

where m = 3, 4, 5, and 6, and k is a constant. He also found that these lines were related to hydrogen. Soon another five lines were found for hydrogen, but now they were located in the ultraviolet range of the solar absorption spectrum. The wavelengths of these new lines were also described with good accuracy by Balmer's formula.

In 1890 Johannes Rydberg proposed another form of this formula:

$$\frac{1}{\lambda} = \frac{\nu}{c} = \frac{4}{k}\left(\frac{1}{2^2} - \frac{1}{m^2}\right).$$

The coefficient $4/k$ was named the Rydberg constant R, and according to modern data, $R = 10973731.77\ \mathrm{m}^{-1}$.

Another three series of lines were subsequently discovered, which were in the infrared region of the hydrogen spectrum and which also obeyed the same law. Soon it was clear that all five series of spectral lines could be described by a single Balmer-Rydberg formula:

$$\frac{1}{\lambda} = R\left(\frac{1}{n^2} - \frac{1}{m^2}\right),$$

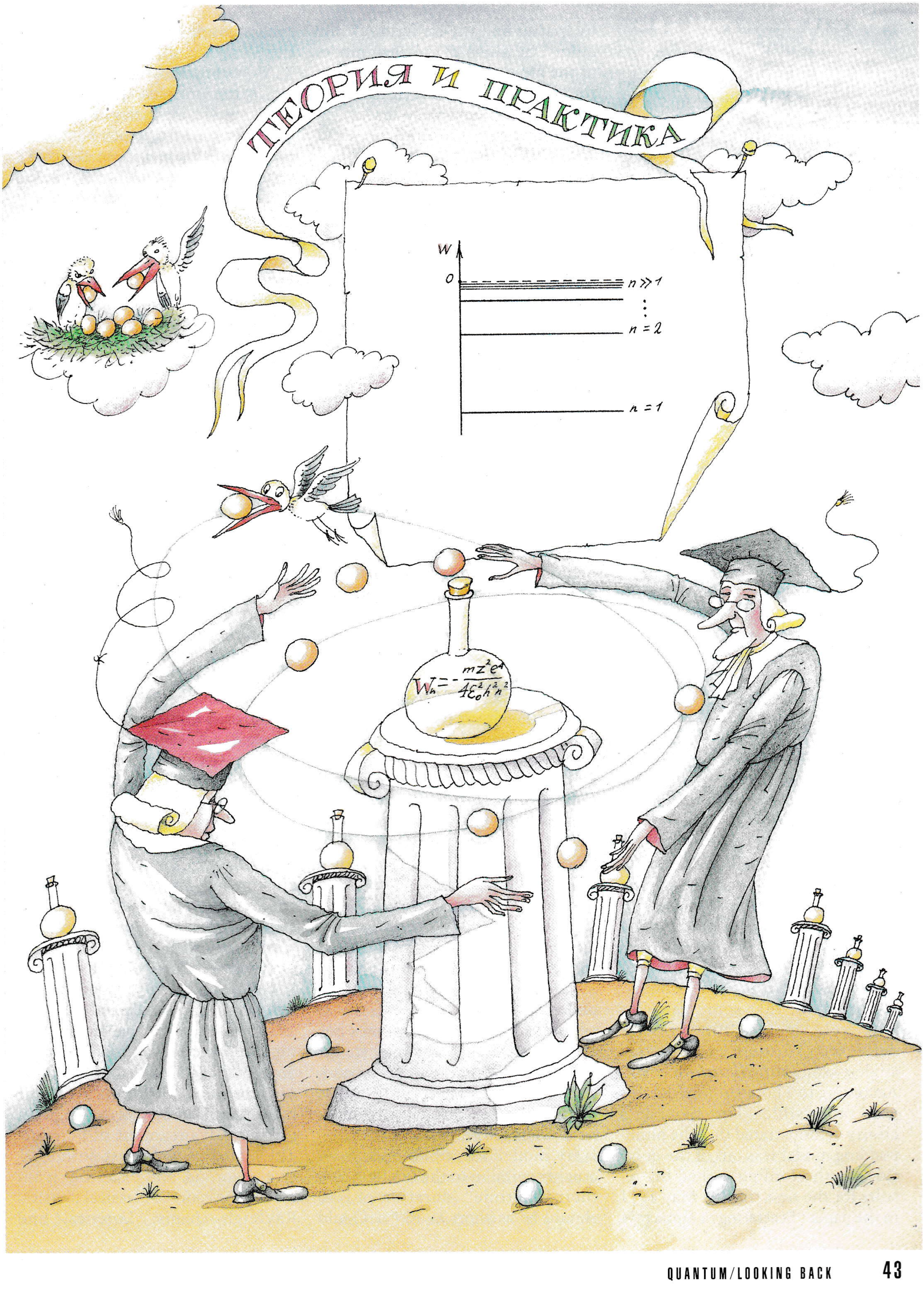
where the integer n = 1, 2, 3, 4, and 5 corresponds to a particular series, and in each series the integer m assumes values starting from $n + 1$.

However, triumph in the mathematical description did not mean the creation of a fundamental physical theory of spectral lines. In particular, the dominating atomic model of Sir J. J. Thomson, which considered matter to be a positively charged fluid in which the negative electrons were arbitrarily distributed like "plums in pudding," did not explain these results.

At the beginning of the twentieth century, there were other indications that pointed to a complicated structure of matter. In 1900 Max Planck advanced the idea of the quantum (discrete) nature of radiation and propagation of light. He needed such a strange concept to explain the regularities of thermal radiation. By the way, his hypothesis did have some experimental basis: As far back as 1887, Heinrich Hertz observed photoemission, an essentially "quantum" phenomenon. Other enigmatic facts were the discovery of the electron by Sir J. J. Thomson, radioactivity by Antoine Becquerel, and thermal electron emission by Sir Owen Willans Richardson.

What was a common feature in all these phenomena? They could not be explained on the basis of the old concepts of atomic structure. However, the history of physics shows that the accumulation of such strange data can go on for a long time until some qualitative leap occurs, a historical event in science that formulates the final verdict either for the accumulated data or for the theory that at first glance contradicts them. In atomic physics such a leap was made by the experiments of Rutherford, which laid the cornerstone for a new theory of atomic design.

Art by V. Ivanyuk


ТЕОРИЯ И ПРАКТИКА
W
0
n ≫ 1
n = 2
n = 1
$W_n = -\frac{mZ^2e^4}{4\varepsilon_0^2 h^2 n^2}$

Rutherford's experiments

As far back as 1906, Ernest Rutherford studied the passage of alpha particles through various substances. In December 1910 he deduced the formula that described the scattering of these particles. It showed that the number of particles emitted by a given source (characterized by the specified flux density and kinetic energy of the radiated particles) into the solid angle θ is related to the angle of scattering by the formula

$$\frac{\Delta N}{\Delta \Omega} \propto \frac{1}{\sin^4 \frac{\theta}{2}}.$$

The plot of this function is shown in fig. 1, where the coefficient of proportionality is taken to be one.

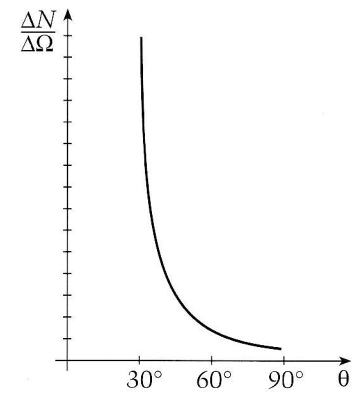


Figure 1

Rutherford's assistants Ernest Marsden and Hans Geiger spent many weeks in complete darkness recording the scintillations on luminescent screens that showed the locations of the scattered alpha particles. They detected and characterized about two million individual collisions!

The results were revolutionary. It turned out that some particles (relatively small in number) were deflected through very large angles—sometimes larger than 90°. According to Thomson's model this was impossible. The new data clearly demonstrated that the plum-pudding model was out of the question. Rutherford published his results for the first time in May 1911 in the paper "Scattering of alpha rays by matter and the structure of the atom," in which the nuclear model of the atom was born and in which its drastic contradictions with Thomson's model were discussed. According to Rutherford, an atom was similar to a planetary system, and it contained a heavy and positively charged nucleus (its own "Sun") as well as negatively charged electrons (the "planets") orbiting around it.

It is instructive to ask if Rutherford was the first physicist who rejected the plum-pudding model. Didn't anybody see all the complications and contradictions in the problem of atomic structure and try to modify the current views?

The history of physics shows that such attempts were made long before 1911. For example, as early as 1901, the French physicist Jean Perrin lectured on the probable nuclear-planetary structure of atoms. In 1904 a Saturn-like atomic model was advanced by the Japanese physicist Hantaro Nagaoka, in which the central positively charged nucleus was surrounded by a ring of electrons revolving with the same angular velocity. History doesn't know whether Rutherford ever met Nagaoka, who traveled in Europe and even visited Manchester, but there is a reference to his model in Rutherford's paper. A similar and quite interesting model was suggested by the English astrophysicist John Nicholson in 1911 or 1912, who constructed it to explain a number of lines of unknown origin in the spectra of nebulae.

This list could be continued, but let's return to Rutherford's experiments and his paper. Some phrases in it indicated that he couldn't help understanding that his model was at odds not only with Thomson's but also with the classical electrodynamics of Maxwell, because a continuously accelerated charge must continuously radiate electromagnetic energy. Therefore, in the planetary model, electrons must very quickly spiral in to the nucleus, and their "lifetime" (duration of fall) should be as short as 10^{-8} s. Could such an atom be a stable construction?

Moreover, according to classical views, the emission spectrum of atoms should consist not of lines, but of continuous frequency bands, because the frequency of an electron's revolution is not constant. Thus, the nuclear-planetary model dramatically highlighted the antagonism of the current theoretical views and evident atomic stability. Rutherford himself could not untangle this puzzle. Nevertheless, a solution was found.

Bohr's hypothesis

Niels Bohr showed early signs of becoming an outstanding scientist. In 1905, as a student at Cambridge University, he studied the oscillation of liquid jets in order to measure surface tension. His work earned him a gold medal. His masters dissertation was devoted to the electron theory of metals (1909), and thereafter he worked on his doctoral degree. In 1911 he proved the impossibility of creating a theory of the magnetic properties of matter entirely on the basis of classical views.

After defending his thesis, Bohr went to Cambridge for one year's work in Thomson's laboratory, where in October 1911 he participated in the traditional Cavendish Laboratory party together with Rutherford, who invited him to work in his lab in Manchester. This was the period (spring through autumn in 1912) when Bohr came to the conclusion that the contradictions of the nuclear-planetary model and classical electrodynamics could be solved only with the help of the exotic quantum theory of Max Planck.

Upon returning to Copenhagen, Bohr worked intensely and in March 1913 finished three papers describing the principles of his theory. In September 1913 Bohr reported his new results in Birmingham at the meeting of the British Association for the Advancement of Science. The audience was most authoritative, rigorous, and exacting. It included the cadre of classical physics: Rayleigh, Jeans, Lorentz, and

Thomson. The patriarchs of science met the report of the novice with a rather chilly response. However, Rayleigh wryly remarked that it makes no sense to have sexagenarians commenting on modern ideas. The situation changed for the better only after a number of papers were published by Bohr in scientific journals. Sir James Jeans was the first to support the ideas of Bohr: "Doctor Bohr gave the most witty, fruitful, and I suppose, the most convincing explanation of the relationships observed in the spectral lines."

As we have noted, Planck's idea on the discrete (quantum) nature of atomic energy was accepted in the mid 1920s together with Einstein's concept of the quantum structure of atoms. What was contributed by Bohr? First of all, he advanced the notion that the principal inference of classical electrodynamics on the continuous character of electromagnetic radiation emitted by electrons revolving around the atomic nucleus must be rejected. In place of this old concept, Bohr proposed the existence of stationary states of an atom, in which it doesn't emit energy. In addition, Bohr postulated the possibility of transitions between the stationary states accompanied by either the emission or absorption of energy. Clearly, one needs great scientific courage to make such steps—and Niels Bohr was equal to the task.

Bohr's hypotheses subsequently assumed the form of three famous postulates giving the rules for quantizing the electron's parameters in atoms. According to them, the quantum nature of angular momentum is described by the formula

$$mvr = \frac{h}{2\pi}n,$$

where $\hbar$ is Planck's constant. Combining it with the formula for the energy of an orbital electron

$$E = \frac{mv^2}{2} - \frac{ze^2}{4\pi\varepsilon_0 r}$$

and Newton's second law

$$\frac{mv^2}{r} = \frac{ze^2}{4\pi\varepsilon_0 r^2},$$

we obtain Bohr's famous formula for the energy of an electron in an atom:

$$E_n = -\frac{mz^2e^4}{4\varepsilon_0^2h^2n^2}.$$

This energy is quantized and assumes a set of discrete values corresponding to the integers $n = 1, 2, \ldots$. Bohr wrote that the "different numbers n correspond to series of values E_n, relating to various configurations of the system, in which there is no radiation, so they will remain unchanged until the system is disturbed from the outside."

It is interesting that in 1913, Bohr took the advice of a colleague and compared his formula with that of Balmer-Rydberg (he was unaware of this achievement in spectral physics). Bohr supposed that the spectral terms R/n^2 and R/m^2 were proportional to the energy of an electron in various stationary states. The next step was to assume that the transition of an atom from one state to another was accompanied by the radiation of a single quantum of energy, from which the famous rule of spectral frequencies immediately follows (Bohr's third postulate):

$$h\nu_{n_1 \to n_2} = E_{n_1} - E_{n_2},$$

$$\nu = \frac{mz^2e^4}{4\varepsilon_0^2h^3}\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right).$$

This coincidence with the Balmer-Rydberg formula was ideal, as it attested to the agreement of his theory with experimental data. According to George Hevesy, when Einstein was informed of the striking confirmation of Bohr's theory, he was astonished by the fact that the frequency of radiation really didn't depend on the frequency of an electron's revolution in the atom: "The large eyes of Einstein opened even wider and he said: 'In this case it is one of the greatest discoveries in history!'"

In his introduction to the paper "Binding of electrons by a positively charged nucleus," Bohr wrote on the principal role played by Planck's constant in his theory: "Only the existence of the quantum of action h prevents the fusion of electrons with nuclei and the creation of a neutral particle of virtually infinitely small size.... This fact alone provides a comprehensive explanation of the remarkable relationships between the physical and chemical properties of the elements, which are manifested in Mendeleev's Periodic Table."

Correspondence principle

Rejecting classical electrodynamics, Bohr nevertheless always tried to find a bridge between the new and old theories, and in 1912 he formulated the famous correspondence principle. According to this principle, in a number of limiting cases a physical theory based on the generalization and development of some classical theory should yield the same results as that produced by this old theory.

For Bohr's atomic theory the correspondence principle should be interpreted in the following way: For large quantum numbers n, the results of quantum theory should coincide with those given by the classical approach. For example, at large quantum numbers, the "distances" in the hydrogen atom between adjacent energy levels are very small (fig. 2), so these levels become almost continuous, which is the same as the concept of continuous energy in classical physics. In the paper "On the spectrum of hydrogen," Bohr calculated the Rydberg constant us-

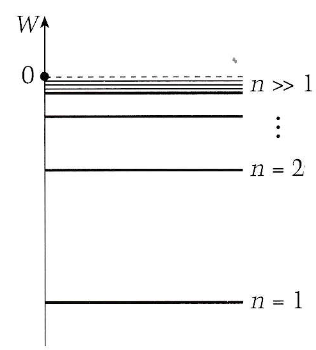


Figure 2

ing his correspondence principle. Try to repeat the course of his thoughts on your own.

Theory and experiment

Can any experimental data corroborate a theoretical conclusion? As Einstein once said: "Never will an experiment say 'yes' to a theory. In the best case it says 'maybe,' but mostly we hear only the flat 'no.'" Thus, the agreement of an experiment with a theory means only "it is possible," although disagreement is a negative verdict. Therefore, a reasoning on how many and what kind of experiments should be carried out to prove a certain theory cannot be absolutely blameless—the more experiments, the greater the confidence in a theory. However, there is no such thing as absolute confidence in any theory. At any moment a new phenomenon may be discovered that contradicts the current theory. If this phenomenon doesn't result from some flaw in the experiment, one should think seriously about whether the theory is always correct and even whether it is true at all.

In spite of the great number of experiments that attested to the validity of Bohr's theory, the theory was not perfect. Indeed, this theory could not explain differences in the intensities of the spectral lines, some parameters of the helium atom, the doubling of spectral lines, and many other phenomena. It became increasingly clear that there was an inherent contradiction in the attempts to combine two incompatible things—classical physics and quantum postulates—not only in the limiting cases but in the entire range of physical phenomena.

In the years 1926 and 1927, Erwin Schrödinger and Werner Heisenberg, backed by Bohr's theory and a large number of experimental and theoretical prerequisites, laid the foundation for a consistent theory of atomic structure—quantum mechanics. What fate was prepared for Bohr's theory? A number of its consequences such as the quantization rules of Bohr-Sommerfeld became the limiting cases, where modern quantum mechanics met Bohr's theory. The idea of discreteness at the atomic level was the starting point in further studies of many scientists. Other aspects of this theory are of great historical interest. According to Einstein, Bohr's theory is "music of the highest quality in the mental world." The works of Niels Bohr were also highly appreciated by Rutherford: "I consider the papers of Bohr as the greatest triumph of human endeavor."

In 1922 Niels Bohr won a Nobel Prize for "merits in the study of atomic structure."

Quantum articles about atomic theory:

S. L. Glashow, "The elementary particles," September/October 1990, pp. 49–51

M. Digilov, "A Strange Box and a Stubborn Brit," March/April 1991, pp. 26–27.

S. R. Filonovich, "The Power of Likeness," September/October 1991, pp. 23–27.

Bruck, M. Zelnikov, and A. Stasenko, "Wobbling Nuclear Drops," January/February 1997, pp. 12–17.

ANSWERS, HINTS & SOLUTIONS

Math

M251

The form of the expression in x and y reminds us of the formula for the tangent of the sum of two angles. In light of this, let's consider six angles $\alpha_1, \alpha_2, \ldots, \alpha_6$ whose tangents are equal to the given numbers. We can assume that $-\pi/2 < \alpha_1 < \alpha_2 < \ldots < \alpha_6 < \pi/2 < \pi + \alpha_1$. Points $\alpha_2, \ldots, \alpha_6$ partition the segment $[\alpha_1; \pi + \alpha_1]$ into six segments. At least one of them does not exceed $\pi/6$ in length. If, for example, $\alpha_2 - \alpha_1 \le \pi/6$, the desired inequality can be obtained by setting $x = \tan \alpha_2$ and $y = \tan \alpha_1$, because

$$\frac{x-y}{1+xy} = \tan(\alpha_2 - \alpha_1).$$

If the last segment, $[\alpha_6; \pi + \alpha_1]$, is the only one that is less than $\pi/6$, we can set $x = \tan \alpha_1$ and $y = \tan \alpha_2$ and use the identity $\tan(\pi + \alpha_1) = \tan \alpha_1$.

M252

Suppose ABC is the required triangle, h is the altitude to its base, and 2α is the measure of one of its base angles. Then we have (fig. 1):

$$\begin{aligned} KM &= KA + AB + BM \\ &= r\tan\alpha + 2h\cot 2\alpha + R\tan\alpha, \\ PQ &= PC + CQ = CE + CF \end{aligned}$$

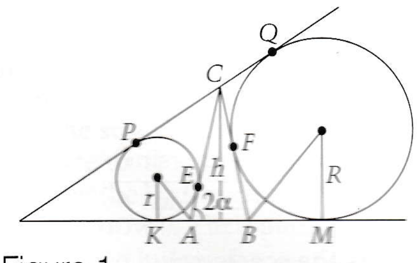


Figure 1

(since tangents to a circle from a point outside are equal)

$$\begin{aligned} &= CA - AE + CB - BF \\ &= 2CA - AE - BF \\ &= \frac{2h}{\sin 2\alpha} - r\tan\alpha - R\tan\alpha. \end{aligned}$$

Since $KM = PQ$, we obtain

$$r\tan\alpha + 2h\cot 2\alpha + R\tan\alpha = \frac{2h}{\sin 2\alpha} - r\tan\alpha - R\tan\alpha,$$

from which we get

$$h\left(\frac{1}{\sin 2\alpha} - \cot 2\alpha\right) = (R+r)\tan\alpha.$$

Since

$$\frac{1}{\sin 2\alpha} - \cot 2\alpha = \frac{1-\cos 2\alpha}{\sin 2\alpha} = \frac{2\sin^2\alpha}{2\sin\alpha\cos\alpha} = \tan\alpha$$

we find that $h = R + r$.

M253

We prove that the perpendiculars drawn to the sides of the triangle at points K, L, and M meet at a point. Suppose that this is not the case (fig. 2a). Take an arbitrary point P inside the triangle with the vertices at the points of intersection. We can see that in this case the given triangle admits a small rotation around point P.

Consider an equilateral triangle ABC with points K, L, and M at its sides such that $AK:KB = 2:1$ and $BL:LC = 3:2$. We must determine into what proportion point M divides side AC if the perpendiculars drawn to the sides of the triangle at points K, L, and M meet at a point. Let's call this point O (fig. 2b). If we let each side of triangle ABC be $15a$, then $AK = 10a$, $KB = 5a$, $BL = 9a$, and $LC = 6a$. We then set $AM = x$ and $CM = 15a - x$. By the Pythagorean theorem,

$$AO^2 - AK^2 = OK^2 = BO^2 - BK^2.$$

Therefore,

$$AO^2 - BO^2 = AK^2 - BK^2 = 75a^2.$$

Similarly,

$$BO^2 - CO^2 = BL^2 - CL^2 = 45a^2$$

and

$$\begin{aligned} CO^2 - AO^2 &= CM^2 - AM^2 \\ &= 225a^2 - 30ax. \end{aligned}$$

Adding these equations, we obtain

$$75a^2 + 45a^2 + 225a^2 - 30ax = 0,$$

from which we get

$$x = 23a/2.$$

We see that point M divides the side AC in the proportion of 23:7.

M254

Let the given plane intersect edge AB of the unit cube $ABCDA_1B_1C_1D_1$

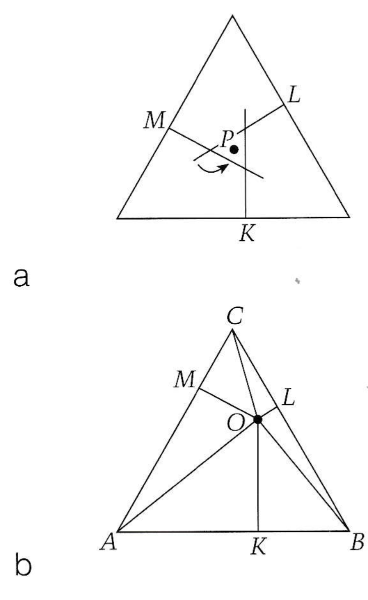


Figure 2

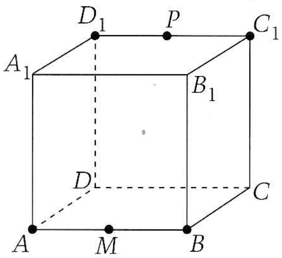
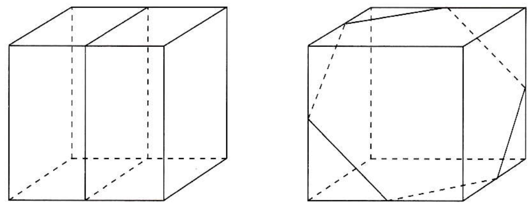


Figure 3



a b

Figure 4

(fig. 3) and let A belong to the first polyhedron and B to the second. It is not hard to see that the plane must intersect some edge of the cube at a point other than its endpoint. This plane must also intersect the opposite edge C_1D_1. In other words, point D_1 belongs to the first polyhedron and C_1 to the second. Let's label the midpoints of edges AB and C_1D_1 as M and P, respectively. Now the Pythagorean theorem shows that

$$\begin{aligned} AP^2 = BP^2 &= BC_1{}^2 + C_1P^2 \\ &= BB_1{}^2 + B_1C_1{}^2 + C_1P^2 \\ &= 1 + 1 + 1/4 = 9/4, \end{aligned}$$

so $AP = BP = 3/2$, and that for all points of segment PC_1 except for point P, the distance to A is greater than 3/2. Therefore, this segment belongs to the second polyhedron.

Similarly, we can prove that segment PD_1 belongs to the first polyhedron. Therefore, the given plane passes through point P. Similarly, we can prove that the plane passes through point M. This will also be true of any edge of the cube that passes through the plane. Thus the plane must pass through the cube's center as well. Now, consider a face of the cube that is intersected by the given plane. If the plane intersects two opposite sides of this face, it does so at the midpoints, and the section forms a square congruent to the cube's face (fig. 4a). If the plane intersects two adjacent edges of the face, it is not hard to see that the section forms a regular hexagon (fig. 4b), and the computation of the required area is now straightforward. **Answer:** 1 or

$$3\frac{\sqrt{3}}{4}.$$

M255

No participant of the conference makes more than 198 handshakes. Since the President of Illyria did not ask himself about the number of handshakes, the answers he received were the numbers 0, 1, 2, ..., 197, 198. The delegate who made 198 handshakes—delegate 198, whom we can assume to be a president—shook hands with everybody except his prime minister. Therefore, everybody except his own prime minister made at least one handshake, and thus delegate 0 must be the prime minister of the same country as delegate 198.

Eliminating this delegation together with all handshakes made by its members, we face the same situation as at the beginning of our reasoning, but for 99 countries. Reasoning as before, we see that the delegates who made 0 and 196 handshakes (they are delegates 1 and 197 of the initial situation) are from the same country. Continuing our reasoning in the same way, we find that delegate 99 is left without a pair. Since the list of those polled includes everybody except the President of Illyria, this delegate 99 is the Prime Minister of Illyria. Thus, the Prime Minister of Illyria shook hands with 99 persons.

Physics

P251

It is convenient to use a frame of reference that moves with the robber to the right with speed $v_0 = 5$ m/s. In this system the car's motion is similar to the oscillation of a load fastened to a spring attached to a fixed point (the robber). After the cord is stretched to the maximum length, it will start to contract. By the time its extension becomes zero, the car strikes the robber with speed v_0 directed to the right. In the rest frame of reference, the platform's speed is $2v_0$. As the robber lies on the car after the collision, the speeds of both bodies are identical. In physics such collisions are called completely inelastic. Conservation of momentum yields the mass ratio:

$$M \cdot 2v_0 + m \cdot v_0 = (M + m) \cdot 1.8v_0,$$

from which we obtain

$$M = 4m.$$

P252

The simplest way to solve this problem is based on conservation of energy. The kinetic energy of a wheel that rolls with speed v in a non-slipping manner is

$$E = \frac{mv^2}{2} + \frac{Mv^2}{2} + \frac{Mv^2}{2}.$$

The last two fractions are the kinetic energy of the translational motion of the rim with mass M and the energy of its rotation around the axis.

When the wheel has traveled the distance L down the incline, its center descended by $h = L \sin \alpha$. The corresponding decrease in the gravitational potential energy is equal to the increase in the kinetic energy:

$$(M + m)gL \sin \alpha = \frac{mv^2}{2} + Mv^2,$$

from which we get

$$v = \sqrt{\frac{2(M + m)gl \sin \alpha}{2M + m}}.$$

The uniformly accelerating motion of a wheel is described by the formulas $L = at^2/2$ and $v = at$, from which we obtain

$$a = \frac{(M + m)g \sin \alpha}{2M + m}.$$

This is the acceleration of the center of mass, which is determined by the total force applied to the wheel. Two forces act on the wheel along the inclined plane: friction and the corresponding projection of gravity. For the minimal coefficient of friction, the force of friction is simply expressed via the normal force and the coefficient of friction. Therefore,

$$(M+m)g\sin\alpha - \mu(M+m)g\cos\alpha = (M+m)a,$$

and

$$\mu = \frac{M\tan\alpha}{2M+m}.$$

P253

We consider only the case when the thickness H of the residual water layer is much smaller than the diameter d of the hole. Let's draw two vertical planes passing through the axis of symmetry of the hole and making some angle ϕ between the planes. Consider the water that is confined between these planes and an arbitrary vertical cylindrical surface (fig. 5, top view). The radius of

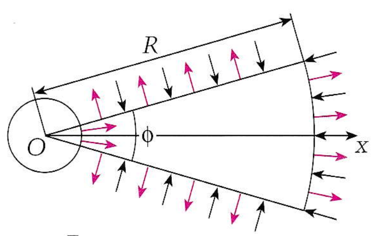


Figure 5

this surface is chosen such that the distance $R - d/2$ is much larger than H (although it is less than $d/2$). To a good approximation, at such distances the water surface can be considered horizontal.

The equilibrium condition for this water means the sum of all forces affecting it is zero, and the sum of the projection of these forces on the x-axis is zero. What are these forces?

The straight sides are affected by surface tension (red arrows in fig. 5) and by the hydrostatic pressure (the black arrows). Their projections are, respectively,

$$2\sigma R\sin\left(\frac{\phi}{2}\right)$$

and

$$-\rho g H^2 R\sin\left(\frac{\phi}{2}\right).$$

Similar forces act on the curved side, which have projections

$$2\sigma\left(R-\frac{d}{2}\right)\sin\frac{\phi}{2}$$

and

$$\rho g H^2\left(R-\frac{d}{2}\right)\sin\frac{\phi}{2}.$$

Here we have neglected the change in the thickness of the water layer near the hole (the length of this part of water is on the order of H, and in our case $H << R - d/2$). Now we must consider only the forces of surface tension affecting the water near the hole where it contacts the bottom. Figure 6 shows the case where the bottom of the vessel is somewhat water resistant, when angle θ between the tangent to the water surface and the horizontal plane of the bottom (wetting angle) is larger than $\pi/2$. The resulting projection of the surface tension acting at the point of contact (red inclined arrow in fig. 6) is

$$-\sigma d\sin\frac{\phi}{2}\cos\theta.$$

Thus, the equilibrium condition is

$$-2\sigma R\sin\frac{\phi}{2} - \rho g H^2 R\sin\frac{\phi}{2}$$
$$-2\sigma\left(R-\frac{d}{2}\right)\sin\frac{\phi}{2} + \rho g H^2\left(R-\frac{d}{2}\right)\sin\frac{\phi}{2}$$
$$-\sigma d\sin\frac{\phi}{2}\cos\theta = 0.$$

This equation yields the thick-

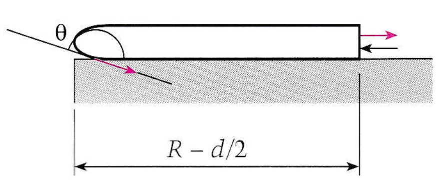


Figure 6

ness of the residual water layer:

$$H = \sqrt{\frac{2\sigma(1-\cos\theta)}{\rho g}}.$$

Now let's estimate the maximum mass of residual water. A bottom that does not wet at all ($\theta = 180°$) corresponds to the maximum depth

$$H_{max} = 2\sqrt{\frac{\sigma}{\rho g}} \cong 5.3 \text{ mm}.$$

In this case, the maximum water mass is

$$m_{max} = \rho\frac{\pi\left(D^2-d^2\right)}{4}H_{max} \cong 1 \text{ kg}.$$

In this case, the thickness of the residual water can be estimated by another method as well. It is based upon the equality of extra pressure over the convex surface (formed under the influence of surface tension) and the hydrostatic pressure at the depth of $H_{max}/2$:

$$\frac{2\sigma}{H_{max}} = \rho g\frac{H_{max}}{2}.$$

This yields $H_{max} = 2\sqrt{\sigma/(\rho g)}$, which coincides with the previous value. By the way, this coincidence attests to the equivalent character of the assumptions that underlie both methods of solving the problem.

P254

From the given conditions, the power of the source is 4 times that of the lamp. Thus, the voltage drop across the resistor is 3/4 of the source voltage, and the voltage across the lamp is only 1/4 of this value. Therefore, at the same current, the voltage drop across the lamp is one third that across the resistor.

If on figure 2 we plot the current-voltage relationship for the resistor, we get a straight line that crosses the curve for the lamp at 2.5 A and 6 A. These numbers are approximate, and their accuracy is determined by the scale of the graph. You may obtain slightly different values. Now we can find the solution. The conditions of the problem are met either for 2.5 A:

$$V_1 = 2.5\text{ A} \cdot 10\,\Omega + 2.5\text{ A} \cdot 10/3\,\Omega \cong 33.5\text{ V},$$

or for 6 A:

$$V_2 = 6\text{ A} \cdot 10\,\Omega + 6\text{ A} \cdot 10/3\,\Omega = 80\text{ V}.$$

P255

Let's construct the image S_1 of the source S in the flat mirror (fig. 7). Source S illuminates most of the screen's surface. On the contrary, rays from image S_1 (physically, the reflected light) cross the rays of source S only in the region AB. Let's denote $AB = h$ and $AC = z$. Similar triangles yield

$$\frac{b+d}{a} = \frac{L-b-d}{z-h},$$

and

$$\frac{b}{a} = \frac{L-b}{z},$$

from which we get

$$h = \frac{\frac{La}{b}}{1+\frac{b}{d}} \cong 16.7\text{ cm}.$$

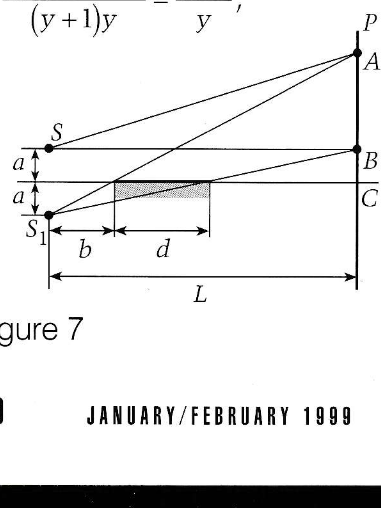


Figure 7

Brainteasers

B251

Let $a_1 = x$ and $a_2 = y$. Then,

$$a_3 = \frac{y+1}{x}, a_4 = \frac{\frac{y+1}{x}+1}{y} = \frac{y+x+1}{xy},$$

$$a_5 = \frac{\frac{y+x+1}{xy}+1}{\frac{y+1}{x}} = \frac{y+x+1+xy}{(y+1)y}$$

$$= \frac{(y+1)(x+1)}{(y+1)y} = \frac{x+1}{y},$$

Thus, $a_6 = a_1$. Similarly, we prove that $a_7 = a_2$, and our sequence is periodic with a period of 5. Therefore, $a_{2002} = a_2 = 1999$.

B252

It is sufficient to prove that the cubes can be superimposed so that two white vertices of the first coincide with white vertices of the second (even if the third white vertex of the first cube is paired with a black one from the second, the other four black vertices of the second cube must coincide with black vertices of the first). Figure 8 shows that the eight vertices can

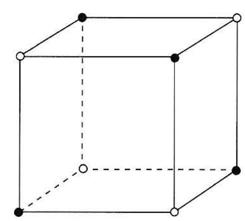


Figure 8

be partitioned into two quadruplets such that the distance between any two vertices of the same quadruplet equals a diagonal of the cube's face. Therefore, if we color any three vertices of the cube black, there will be at least one pair of vertices separated by a distance equal to a face diagonal. Thus, among three white vertices, at least two are at the distance of the diagonal of the cube face. Let us superimpose these two white vertices with the similar two white vertices of the other cube. Then four pairs of black vertices will coincide.

B253

This is impossible. Let's begin by coloring the towns two different

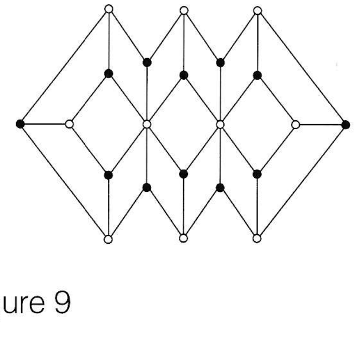


Figure 9

colors (black and white) as shown in figure 9. We have 12 black towns and 10 white towns, and as we travel the roads, the colors alternate. Therefore, if we visit all 12 black towns, the number of white towns we pass cannot be less than 11. However, there are only 10 white towns.

B254

The minimum value of n is 15. First, we prove that for 15 cards, the desired pair can be found. Suppose the contrary. Then the cards numbered 1 and 15 must be in different stacks, as must cards 1 and 3. Thus, cards 3 and 15 are in the same stack. Therefore, cards $6 = 9 - 3$ and $10 = 25 - 15$ are in the other stack, which contradicts the assumption, since $6 + 10 = 16$.

Now we show that 14 cards can be distributed between two stacks such that the sum of the numbers of any two cards of the same stack is not an exact square. Here is an example: 1, 2, 4, 6, 9, 11, 13 (the first stack) and 3, 5, 7, 8, 10, 12, 14 (the second stack). For any number of cards less than 14, the cards can be distributed between the two stacks in a similar way (with the desired condition holding true).

B255

The water starts to boil when its temperature (and that of the box) reaches 100°C. Then the energy of the candle is spent vaporizing the water, and this process goes on at the constant temperature of 100°C, which is too low to ignite paper.

Kaleidoscope

1. The collision between steel and marble is almost elastic, yet the collision between steel and asphalt is almost completely inelastic.

2. In the second case, the momentum of the fragile object decreases over a longer period, so a smaller force acts on the object.

3. Although the force exerted by

an anti-tank missile is very large, it acts for a short time and the impulse on the bus is much smaller than the impulse produced by several people over a longer period.

4. No, it is not a dangerous trick. The acceleration acquired by a massive anvil in the course of an elastic collision with a hammer is virtually zero. Therefore, the force affecting the athlete is also very small.

5. The lead.

6. After a shot, the suspended rifle is given more energy than a firmly fixed rifle. Therefore, in the first case the bullet is given less kinetic energy, so the range is shorter for a suspended rifle.

7. The impulse of the recoil produced by the bazooka is transferred not to the gun (and then to the soldier), but to the exhaust fumes that move in the direction opposite to the missile.

8. After the impact of a single ball on the right, the extreme left-hand ball will move through the same angle as the right-hand ball had initially. In the second case, the impact of two balls on the right results in the recoil of two balls on the left. Likewise, the impact of three balls on the right results in the recoil of three balls on the right.

9. The ball will bounce horizontally from the wedge and then move along a parabolic trajectory.

10. The more inflated ball is more elastic. A kick imparts less momentum in an inelastic collision than in an elastic collision. Therefore, an underinflated ball cannot be kicked as far.

11. By relaxing the hands and moving slightly backward, a player increases the duration of slowing down the ball, thereby diminishing the forces.

12. An important feature is the independence of the force and the velocity.

13. The center of mass of the projectile moves with the same velocity before and after the explosion (point C in fig. 10). Therefore it will proceed along the same

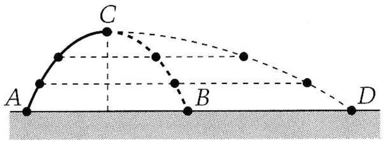


Figure 10

parabolic trajectory as the unexploded projectile. As both fragments have equal masses, they will move symmetrically relative to the parabolic trajectory of the center of mass. Thus, the second fragment will fall at point D, where $BD = AB$.

14. The motion of the molecules is impeded by their mutual collisions.

15. The smaller the surface of a particle, the more unbalanced impulse it receives from the molecules that collide with it.

16. At lower densities of a gas, the mean free path of electrons that ionize the atoms increases, so the electrons can acquire more kinetic energy at smaller voltages. This energy is spent ionizing atoms during collisions between an atom and an accelerated electron.

17. The atoms can be excited by the extra energy received in high-speed collisions.

18. The mass of a neutron is nearly the same as that of a proton. Therefore, a neutron loses more energy in a collision with a hydrogen atom than with a lead atom.

Microexperiment

When a large ball bounces off the floor, it hits the small ball and gives it a fraction of its momentum and kinetic energy. As a result, the small ball will bounce higher than the height from which it was dropped.

In the Lab

The experiment clearly demonstrates the physics of the boiling process. Boiling starts when the pressure of the saturated vapor at the temperature of the liquid becomes equal to the pressure of this liquid. When this happens, the bubbles appearing in the liquid do not collapse. On the contrary, they grow, rise to the surface, and burst. To support marked boiling, one needs a continuous supply of energy, which compensates for the energy expended by the intensive evaporation.

Near the open surface of the water in the test tube, the boiling process starts at approximately 100°C, when the pressure of the saturated water vapor equals that of the surrounding air (about 100 kPa). The rapid boiling drives out almost all the air that was in the tube, so the contents of the plugged tube consist almost entirely of water and the saturated vapor above it.

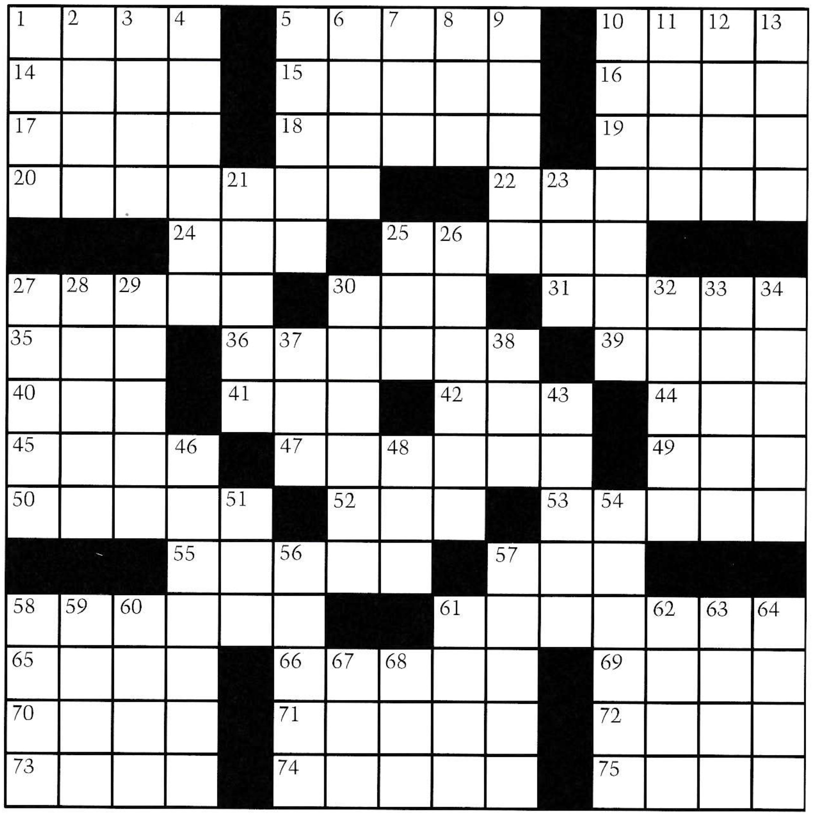
When cold water is poured over the tube, both the temperature and pressure of the vapor decrease. Therefore, the pressure of the water also drops, and its temperature has no time to change. Thus, the pressure in the water becomes lower than that of the saturated vapor corresponding to its temperature, so boiling recurs. The energy for boiling is taken from the cooling water. After a while, the temperature of the water decreases (the pressure in the tube increases), and boiling stops.

Successively pouring first cold then ice-cold water results in a gradual decrease in the temperature of the water and saturated vapor. Finally, the pressure in the tube is very small (at 0°C the pressure of saturated vapor is almost 17 times less than atmospheric pressure). To demonstrate this, it is enough to open the test tube—you will hear the air rushing into the tube.

There is another and very attractive way to demonstrate that there is almost a "vacuum" in the space above the water. Hold the tube vertically and shake it vigorously up and down. You will hear an unusual sound, as if there were some solid substance instead of water in the tube. The reason for this is the absence of air above the water, which would slow its otherwise free motion.

Criss × cross science

by David R. Martin



Across

1 God of war
5 Leaving
10 Clutched
14 Glossy
15 Love: comb. form
16 60,138 (in base 16)
17 Long narrative poem
18 Of the olfactory sense
19 Long wavelength EM wave
20 Ten-sided polygon
22 Drawings of functions
24 Gun club: abbr.
25 "___ and Nothingness," by Sartre
27 Moisten with drippings
30 cgs heat unit: abbr.
31 Major blood vessel
35 Abscisic acid: abbr.
36 Orbital extremum
39 Common or finishing follower
40 Uncle ___
41 Point
42 Reactive power unit
44 1054.5 joules
45 Book's ID
47 3-D circle
49 10^5 pascals
50 Unpredictable behavior
52 Billion years
53 Vertebral column
55 Saltpetre
57 Ten decibels
58 Four-sided polygon
61 Eight-sided polygon
65 Respiratory organ
66 Mud
69 Facial feature
70 Astronomer ___ Struve (1897–1963)
71 Tied one's shoes
72 Neurologist Antonio ___ Moniz (1874–1955)
73 Black
74 Polycyclic compound
75 Arrange in order

Down

1 Scored perfectly
2 Mellow
3 ___ the Red
4 Trigonometric function
5 Italian city
6 Algerian seaport
7 Greek island
8 Nitrilotriacetic acid
9 ___ apparatus (cellular structure)
10 Six-sided polygon
11 Wyatt ___
12 Jacob's wife
13 Units of time
21 ___ circle (equator, e.g.)
23 Nucleic acid
25 Sack
26 Prime number
27 Alkaline
28 Embarrass
29 Brazilian dance
30 Element 29
32 Jewish priest
33 Moon of Saturn
34 Cloister's passageway
37 Greek letters
38 Sense organ
43 Arrange again
46 Nine-sided polygon
48 Garden tool
51 Isaac Newton's title
54 Flat surfaces
56 Unit of magnetic flux
57 773,854 (in base 16)
58 Type of gin
59 A man empowered by Allah
60 Toward
61 Prediction
62 Type of dancer
63 Esker
64 Den
67 Resinous insect secretion
68 Dry ___

SOLUTION IN THE NEXT ISSUE

SOLUTION TO THE NOVEMBER/DECEMBER PUZZLE

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

Bulletin Board

Program in Mathematics for Young Scientists (PROMYS)

To be held at Boston University, July 3 to August 14, 1999, PROMYS offers a lively mathematical environment in which ambitious high school students explore the creative world of mathematics. Through their intensive efforts to solve a large assortment of unusually challenging problems in number theory, the participants practice the art of mathematical discovery—numerical exploration, formulation and critique of conjectures, and techniques of proof and generalization. More experienced participants may also study algorithms, geometry and topology, and combinatorics.

Problem sets are accompanied by daily lectures given by research mathematicians with extensive experience in professor Arnold Ross's longstanding Summer Mathematics Program at Ohio State University.

In addition, a highly competent staff of 15 college-aged counselors lives in the dormitories and is always available to discuss mathematics with students. Each participant belongs to a problem-solving group that meets with a professional mathematician three times per week. Special lectures by outside speakers offer a broad view of mathematics and its role in the sciences.

PROMYS is a residential program designed for 60 ambitious high school students entering grades 10 through 12. Admission decisions will be based on the following criteria: applicants' solutions to a set of challenging problems included with the application packet, teacher recommendations, high school transcripts, and student essays explaining their interest in the program.

The approximate cost of room and board is $1400. Books may cost an additional $100. Tuition is to be determined pending a proposal to the National Science Foundation. Financial aid is available. PROMYS is dedicated to the principle that no student will be unable to attend because of financial need.

PROMYS is directed by professor Glenn Stevens. Application materials can be obtained by writing to PROMYS, Department of Mathematics, Boston University, 111 Cummington Street, Boston, MA 02215, by e-mailing PROMYS at promys@math.bu.edu, or by calling (617) 353-2563. Applications will be accepted from March 1 until June 15, 1999.

Not-so-short stack

This month's CyberTeaser (B254 in this issue) proved controversial. Many contestants were quick to point out the ambiguity of the problem statement. Some contestants were led astray by the nebulous formulation, while some were able to think like a *Quantum* editor and make the assumptions necessary to reach the answer we expected.

As Nick Baxter pointed out in his solution, the ambiguity in the problem could have been avoided by simply adding the word *always*, as in, What is the minimum value of n such that at least one stack will *always* include a pair of cards whose numbers add up to an exact square? The answer $n = 4$ was quite popular with those who did not make this assumption.

Because we have often heard that life is unfair and that justice is hard to come by, we decided to consider $n = 15$ the only correct response. Therefore, the following 10 winners exhibited not only speedy reasoning but also the ability to read between the lines and divine exactly what we were looking for despite the inexactness of the question.

Bob Cordwell (Ellicott City, Maryland)
Max Bachmutsky (Kfar-Sava, Israel)
Matthew Wong (Edmonton, Alberta, Canada)
Leonid Borovskiy (Brooklyn, New York)
Elio Abbondanzieri (Houston, Texas)
Nick Baxter (Hillsborough, California)
John E. Beam (Bellaire, Texas)
Anastasia Nikitina (Pasadena, California)
Andrei Cipu (Bucharest, Romania)
Helio Waldman (Campinas, Brazil)

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

Hula Hoop

by Dr. Mu

WELCOME BACK TO COWCULATIONS, THE column devoted to problems best solved with a computer algorithm. When the snow begins to fly in Wisconsin and the temperature takes a dive, our bovine thoughts stray from the barn to places far away. One of the fondest destinations that warms my milk on a winter night is Waki-Kowa Beach in the Hawaiian islands. It was there, on a sunny afternoon, where I first learned how to move my hips like the Memphis Rock-and-Roller, Elvis Presley. It was back in 1958 when a curious cow first discovered a piece of Marlex in the barnyard and began flipping it around her hips. Soon people caught on and the hula hoop craze swept the country, creating the first big market for the newly discovered product—crystalline polypropylene, better known as plastic.

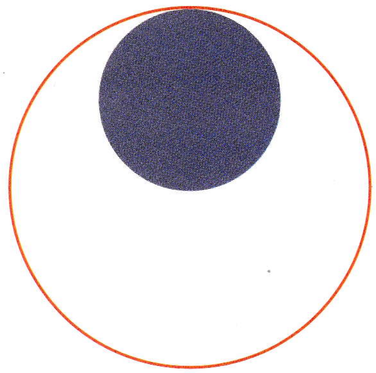
Today, the Hula Hoop is a great way to keep fit. And, believe me, if you're looking for a low-impact aerobic workout that protects your ankles and knees while stressing the flanks, obliques, quads, hams, calves, and deltoids and supplying toning tension to the udder ... the Hula Hoop is almost a total-body workout.

A Mathematica model

We can create a simple model of the hula hoop by drawing a red circle just touching a blue disk.

```
red = RGBColor[1, 0, 0];
blue = RGBColor[0, 0, 1];
Show[Graphics[{blue, Disk[{0, 0}, 1]}],
  Graphics[{red, Thickness[0.01],
Circle[{0, (-1)}, 2]}],
AspectRatio -> Automatic]
```

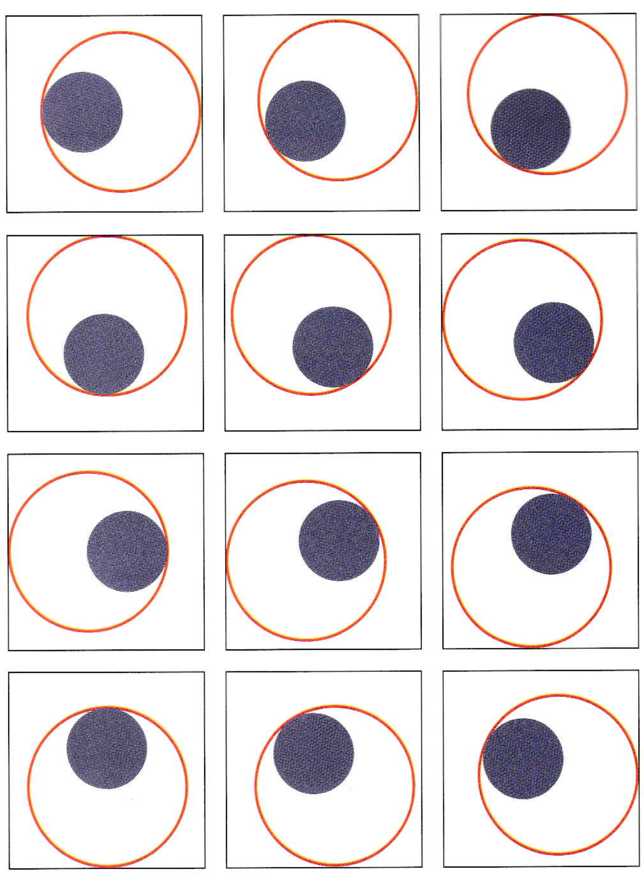
Art by Mark Brenneman



Now we add the motion of hips to the disk, and—presto—we have a hula hoop!

```
Animate[{x, y} = {Cos[t + Pi], Sin[t + Pi]};
  {u, v} = {Cos[t + 2 Pi], Sin[t + 2 Pi]};
    Show[Graphics[{blue, Disk[{x, y}, 2]}],
    Graphics[{red, Thickness[0.01],
      Circle[{u, v}, 4]}],
  AspectRatio -> Automatic, Frame ->
    True, FrameTicks -> None,
PlotRange -> {{-5, 5}, {-5, 5}}], {t,
  0, 2 Pi, Pi/6}]
```

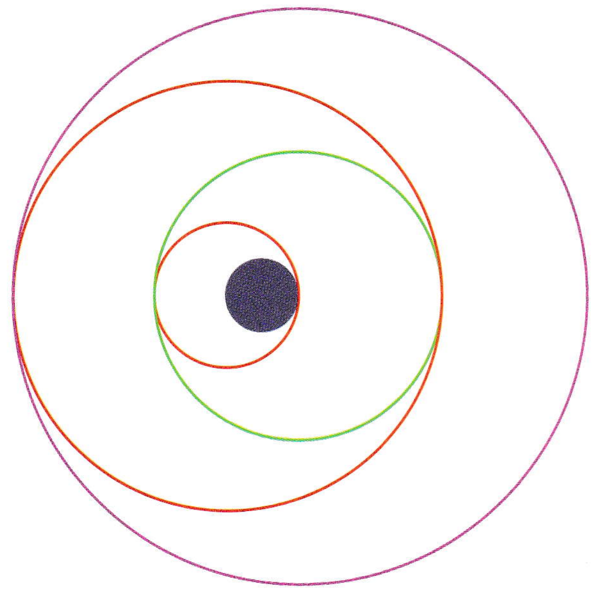
To see the animation, use *Mathematica* and enter the code above. Now double click on the first graphic and sway to the rhythm of the Hula Hula. The best I can do on a static page is to show the output in a graphics array. Notice that the blue disk sways as does the red circle.



What would happen if this hoop could support a larger hoop, and that one a larger hoop, and so on, and a cow was skilled enough to keep them all going? This suggests a problem, which, you guessed it, is your Challenge Outta Wisconsin.

COW 14

Write a program that will animate four Hula Hoops that contact each other as shown below, and run the animation.



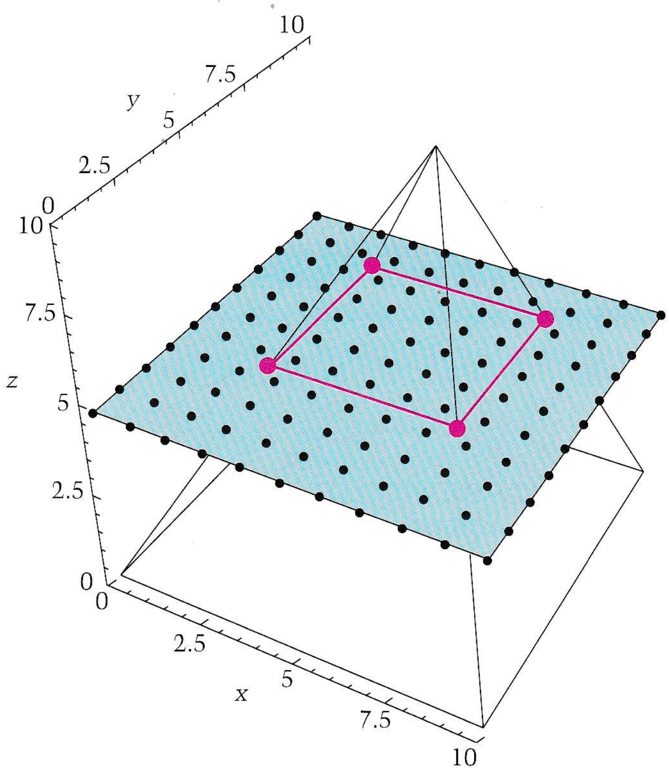
A circle around a disk forms a loop,
That's the model for a Hula Hoop.
Add more hoops and loop them all,
And now be careful so they don't fall.
When that's done, do the animations.
You've solved this COW, congratulations!
—Dr. Mu

COW12

In COW12 you were asked to write a program that would accept any zap point $\{x, y, z\}$ with positive integers and any base with positive integer length and have it cowculate the number of lattice points inside the polyhedron formed by the square base and the zap point. How many flies can be zapped with a single jolt from a Fly Zapper at the zap point {6000, 9000, 10000}, assuming a square base of length = 12000.

Solution

The key is to look at intersecting planes $z = \zeta$. Say that the square base of the polyhedron has opposite corners {0, 0, 0} and {base, base, 0}, and that the top point has coordinates {xtop, *y*top, *z*top}. Notice that for any value ζ between 0 and *z*top, the plane $z = \zeta$ will intersect the polyhedron in four points that form a square with opposite corners {xmin[ζ], *y*min[ζ], ζ} and {xmax[ζ], *y*max[ζ], ζ}, where xmin, xmax, *y*min, and *y*max are determined below.



A point on the line connecting the origin to the top point has the parametric form {xtop t, ytop t, ztop t}. So for $z = \zeta$, we have $t = \zeta$/ztop, and then the x and y coordinates of the intersection of this line with the plane $z = \zeta$ are easily determined:

```
xmin[ζ_] = ζ xtop / ztop;
ymin[ζ_] = ζ ytop / ztop;
```

Similarly, the line connecting {base, base, 0} to the top point can be parametrized as {(xtop – base) t + base, (ytop – base) t + base, ztop t}. For $z = \zeta$, we again have $t = \zeta$/ztop, and then xmax and ymax can be defined.

```
xmax[ζ_] = ζ (xtop - base) / ztop + base;
ymax[ζ_] = ζ (ytop - base) / ztop + base;
```

Armed with these coordinates of the square intersection of the plane $z = \zeta$ with the polyhedron, we next count the lattice points contained in the interior of each such square. The **zcount** function does this in a straightforward way.

```
zcount[ζ_] :=
  (Ceiling[xmax[ζ]] - Floor[xmin[ζ]] - 1) *
  (Ceiling[ymax[ζ]] - Floor[ymin[ζ]] - 1)
```

Summing over all z with 0 < z < ztop finishes the procedure. Here is the assembled function:

```
interiorLatticePoints3D[base_, {xtop_,
  ytop_, ztop_}] :=
Module[{xmin, ymin, xmax, ymax, zcount},
  xmin[ζ_] = ζ xtop / ztop;
  ymin[ζ_] = ζ ytop / ztop;
  xmax[ζ_] = ζ (xtop - base) / ztop + base;
  ymax[ζ_] = ζ (ytop - base) / ztop + base;
  zcount[ζ_] :=
    (Ceiling[xmax[ζ]] - Floor[xmin[ζ]] - 1) *
    (Ceiling[ymax[ζ]] - Floor[ymin[ζ]] - 1)
Sum[zcount[ζ],{ζ,0+1,ztop - 1}]]
```

This function counts the suggested case correctly:

```
interiorLatticePoints3D[12000,{6000,9000,10000}]
479910022999
```

This solution was submitted by Louis J. D'Andria. Correct solutions were also submitted by Eric Rimbey and Robert Dickau.

And finally...

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

If you like to hula hoop around the competition while programming a computer, stop by the USA Computing Olympiad web site at http://usaco.uwp.edu. The 1999 USA Computing Olympiad is in progress, but there is still time to enter the remaining Internet competitions and the National Championship in April.