SCIENTIFIC AMERICAN

METALLIDING

SEVENTY-FIVE CENTS

August 1969

If you want to know the kind of man he is, take a good look at the car he drives.

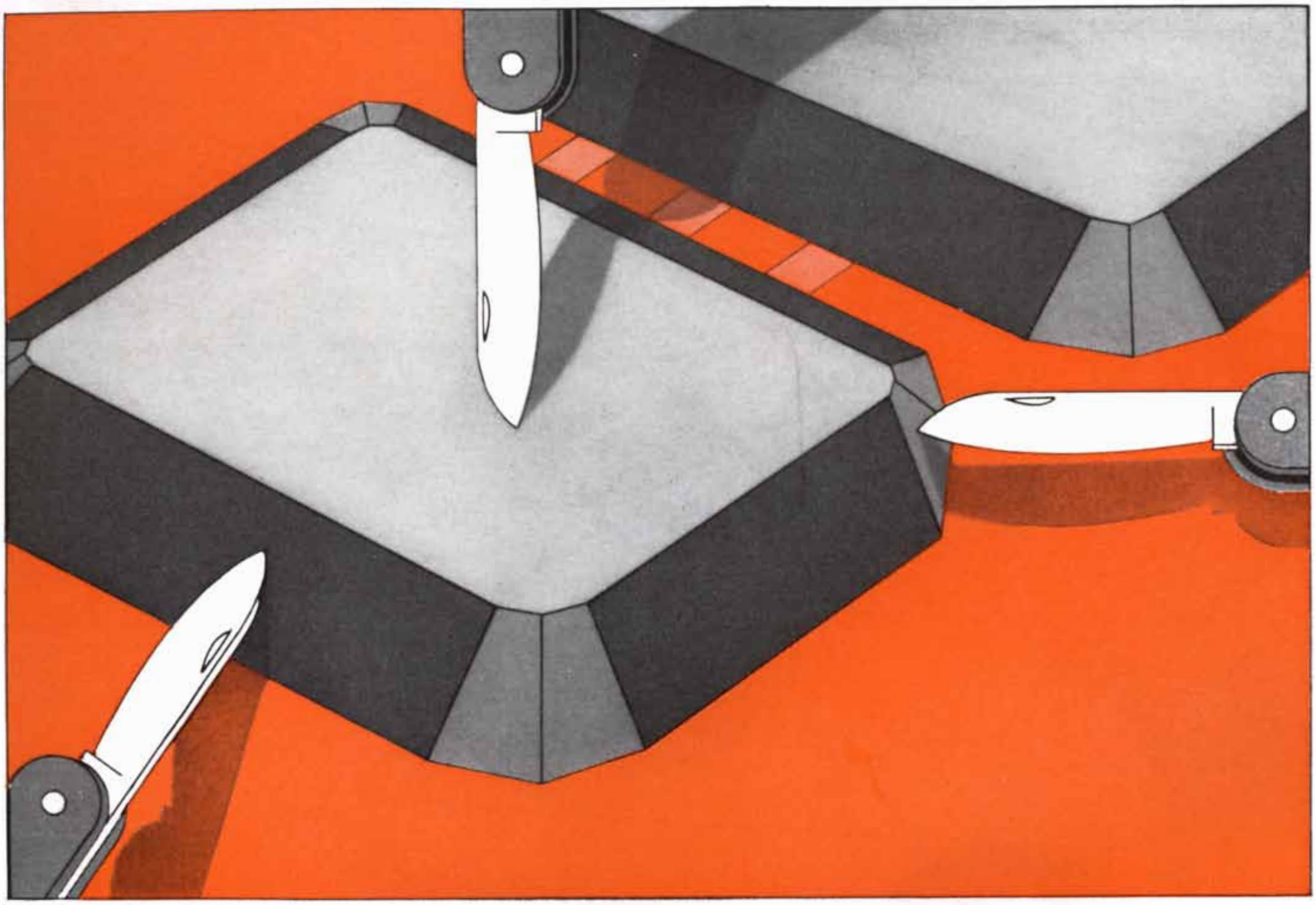
MGB

The MGB man likes to shift for himself. So he insists on a 4-speed synchromesh gearbox. He's a no-nonsense guy. So he wants heavy-duty suspension and front disc brakes. And he's got an eye for beauty and comfort. So we give him leather bucket seats, wire wheels and full carpeting. Point for point, the MGB is in a class by itself. But then again, so is the man who drives it.

BRITISH LEYLAND

For overseas delivery information, write: Dept. SA-1, British Leyland Motors Inc., Leonia, N.J. 07605

MG AUSTIN AT AUSTIN-MG DEALERS.



We sharpen a chemical knife.

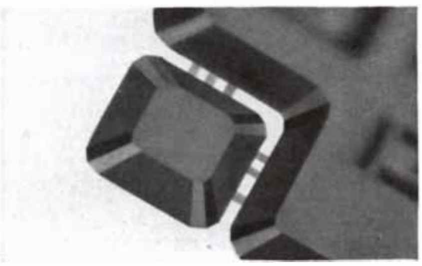
We make many integrated circuits at a time . . . on a single wafer of silicon crystal. Then, we divide the wafer into individual circuits (and, sometimes, the circuits into electrically isolated components) by etching narrow slots through the silicon.

But conventional etchants don't cut downward only. They cut sideways, too, making a slot that's wide as well as deep. So, for safety, we have had to leave plenty of space between circuits. Then, Bell Laboratories scientists Herbert A. Waggener, Roger C. Kragness, and A. Lamont Tyler discovered a means of "one-directional" etching.

The new technique makes wedge-shaped slots, separating the circuit elements along precise lines. It depends on a "preferential" etchant, which most strongly attacks the semiconductor perpendicular to a particular crystal-lattice plane. The slot is wedge-shaped because its walls are other planes toward which the etchant is almost inert.

The process is self-limiting; once a slot goes through a wafer, there is very little further etching. So, we can leave the wafer in the etchant long enough for complete separation of the parts without the careful thickness control formerly required. And, because of the fixed slope of the slot walls, minimum slot width can be much less than wafer thickness.

In the drawing, the light, medium, and dark surfaces represent three crystallographic planes in silicon. The solution etches perpendicular to each of these at different rates. The surface shown light, for example, etches away most rapidly. To cut out an area, a mask is applied onto the fastest—"light"—plane, with the mask edges parallel to the slowest—"dark"—planes of the crystal.



In one example of the new technique, hundreds of 1-mm-square circuits, each with 10 transistors, 14 diodes, and 12 resistors on nine air-isolated areas, were etched at once on a single wafer. Slots were about 17 microns wide with strong "beam leads" for structural and electrical connections. The slot in the photo is only 6 microns wide.

This technique is another step toward making ever better and smaller integrated circuits.

From the Research and Development Unit of the Bell System—

Bell Labs

Established 1845 **SCIENTIFIC AMERICAN** August 1969 Volume 221 Number 2

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1969 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $8 PER YEAR.

THE COVER

The photograph on the cover shows the effects of the technique called metalliding, which involves the diffusion of atoms of one metal into the surface of another by means of a high-temperature electrolytic process (see "Metalliding," page 38). The coating thus formed has properties not possessed by the original material. In this case the wire on the left, consisting of molybdenum alone, and the one on the right, which has had atoms of silicon diffused into the surface of the molybdenum to form an outer layer of silicided molybdenum, differ in their resistance to oxidation. The wires have been heated to a temperature of about 1,000 degrees Celsius; the wire on the left is oxidizing and vaporizing but the silicided wire remains unaffected.

THE ILLUSTRATIONS

Cover photograph by Paul Weller

Page	Source
18–22	Jerome Kühl
23	U.S. Army
24–26	Jerome Kühl
27	U.S. Air Force
28–29	Jerome Kühl
30–37	Eric Mose
39–40	Newell C. Cook, General Electric Research and Development Center
41–44	Allen Beechel
46	Newell C. Cook, General Electric Research and Development Center (*top*); Paul Weller (*bottom*)
58–73	Jim Egleson
74	A. P. Okladnikov
76–77	Enid Kotschnig
79	A. P. Okladnikov
80–82	Enid Kotschnig
86–87	Rod Johnson, Commonwealth Scientific and Industrial Research Organization
88–90	Bunji Tagawa
91	R. D. B. Fraser, Commonwealth Scientific and Industrial Research Organization (*top*); Bunji Tagawa (*middle*); Ruth M. Bonès, University of Leeds, courtesy Jan Sikorski (*bottom left*); George E. Rogers and B. K. Filshie, Commonwealth Scientific and Industrial Research Organization (*bottom middle and right*)
92	R. D. B. Fraser, Commonwealth Scientific and Industrial Research Organization
93	Bunji Tagawa (*top*); R. D. B. Fraser, Commonwealth Scientific and Industrial Research Organization (*bottom*)
94–96	Bunji Tagawa
100	Gerald L. Kooyman
102–103	Thomas Prentiss
104	Thomas Prentiss (*top*), Gerald L. Kooyman (*bottom*)
105	Thomas Prentiss
106	Thomas Prentiss (*top*), Gerald L. Kooyman (*bottom*)
109	Maschinenfabrik Archiv, Augsburg
110–111	Dan Todd
112	Dan Todd (*top*); Maschinenfabrik Archiv, Augsburg (*bottom*)
113–117	Dan Todd
118–121	Alan D. Iselin
124–128	Roger Hayward

Pollution that assaults the lungs, the digestive tract, the ear ...and how effective instruments can lead to abatement

The Lungs

Until very recently, Molecular Rotational Resonance (MRR) Spectroscopy often seemed like a brilliant scientific breakthrough destined to remain an ivory tower curiosity for lack of a practical application in the real world of quantitative analysis.

In its pristine form, MRR allowed the scientist to look into molecular structure by measuring changes in the absorption of microwave energy which result from transitions between rotational energy levels in a polar molecule. Because differences exist in the composition or geometry of individual molecular species, there is a characteristic MRR spectrum for each molecule. Absorption peaks are unique for each molecule and MRR readily differentiates between them, even in a complex mixture, because of its inherent specificity. In the usual case, measuring the frequency of a single absorption line completely identifies the molecule.

MRR has recently been shown to be a practical quantitative tool too. In a paper published in the *Journal of Chemical Physics* (46, 3698, 1967) the response of the HP 8400B MRR Spectrometer was shown to be linear with concentration from the lowest detectable limit to 100%. More recent work with common air pollutants (SO_2, NO_2, hydrocarbons) has demonstrated that MRR gives a quantitative response for each gas, even in the complex mixtures that are commonly associated with air pollution samples. The actual sensitivity limit for SO_2 has been determined at 3.5 nanograms without using concentration techniques (... this corresponds to a concentration of 11.6 ppb in a one liter sample). To further enhance its usefulness in the quantitative analysis of air pollutants, most MRR experiments are carried out at low pressures—typically 10-15 μ Hg—a condition that greatly reduces the rate at which the pollutants react with each other.

Precisely where the MRR Spectrometer fits into the pattern of analytical chemistry is still being studied. Based on the work reported above, it certainly should be considered for air pollution analysis, especially for calibrating on-site air pollution monitors. Results of experimental work in air pollution and other significant analyses with the MRR Spectrometer are published regularly in *Molecules and Microwaves*, a copy of which awaits your request.

The Digestive Tract

In the days before Rachel Carson's *Silent Spring*, the only popular connection between pesticides and the human digestive tract was benign: one was reassured that large parts of the world would be hungry, even suffer famine, except for the beneficial effect of pesticides on agricultural production. Nowadays, it's more common to hear warnings from respected scientific sources that pesticides constitute a real and present danger to life on this planet because they are ingested as residues in the food we eat and the liquids we drink.

These are not mutually contradictory arguments so much as they are accurate descriptions of both sides of the split personality of pesticides. The only conceivable solution to this very human dilemma is better control of the use of pesticides, and more careful analysis of pesticide residues in foodstuffs.

Enter the gas chromatograph (GC). While the men engaged in pesticide detection are many and far-flung, instrumentation for this sensitive work falls almost solely on the GC. On this basis, Hewlett-Packard has directed much research effort towards perfecting both instrumentation and technique. Although pesticide detection is still most often recorded in the nanogram range, an HP GC—more than four years ago—separated a laboratory pesticide sample at the picogram level. Most of this chemical detective work is being performed on the HP Model 402 High-Efficiency GC—an instrument perfected especially for this and other biochemical research. HP's pesticide analysts prefer to use this instrument equipped with an electron capture type of detector. The latter employs a radioactive tritium source to produce electrons whose capture by the pesticide molecules is a direct measure of their presence. Recently, HP chemist-designers have perfected a new electron capture detector that employs a radioactive Ni^{63} source that is more stable at higher temperatures thereby holding out a promise of more searching pesticide detection than the older tritium type can accomplish.

Sometimes the inherent difficulty of pesticide analysis is resolved by improvements in technique rather than hardware. HP chemists have developed special techniques for the analysis of pesticide residues in many foodstuffs, and sample extraction techniques for the analysis of bovine and human milk.

If you'd care to pursue this subject in more depth, write for Applications Lab Report 1003, yours on request.

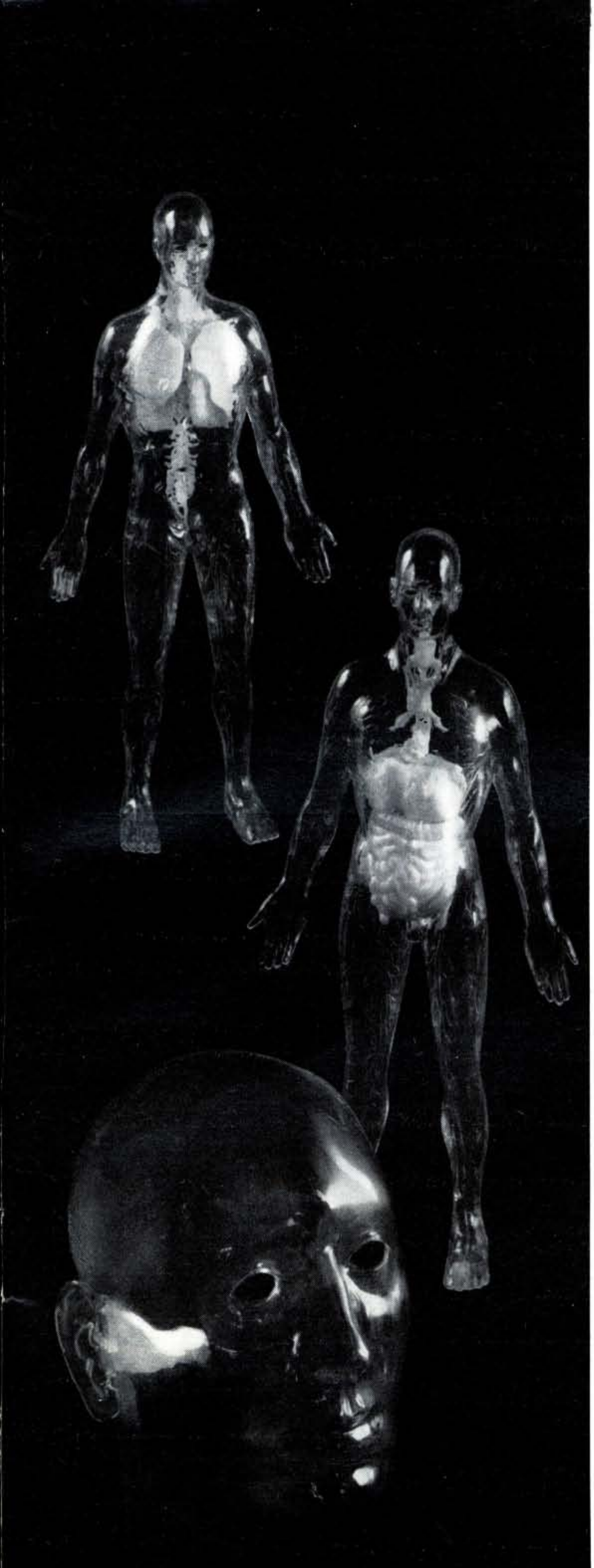
The Ear

Well played by a fine orchestra, Brahms can only be described as beautiful. But reproduced too loud on a cheap phonograph, it's noise. An increasingly widespread and serious form of pollution, noise can make us uncomfortable; prolonged loud noise damages hearing; very loud noises can cause pain, psychosis and even death.

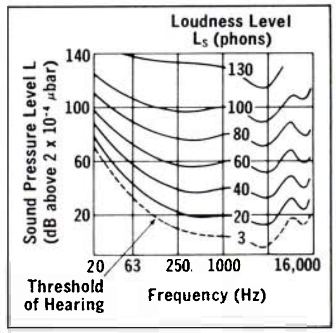
Obviously the time has come to control this form of 20th century environmental pollution. When HP scientists turned their talents to noise measurement, they ran into a very unusual problem. Objectively sound is simply a matter of rapidly changing air pressure, easy to measure with traditional sound level meters. But noise is really not an objective phenomenon: what the ear hears is a subjective sensation of loudness involving complicated physiological and psychological mechanisms.

For an instrument to measure sound as the ear hears it, it must imitate the unique properties of the ear. Take loudness level which is traditionally measured in *phons*. Although the logarithmic phon scale covers the large dynamic range of the ear—120 dB—it does not fit a subjective loudness scale. The trouble is that a noise that sounds twice as loud as another does not measure double the number of phons. So a subjective measure of loudness was developed by international agreement in which the unit is a *sone* and whose scale corresponds closely to the subjective sensation of loudness. For example, the comparison between a jet takeoff and a quiet conversation is 3:1 in phons (120 vs. 40) ... and a much more realistic 60:1 in sones (256 vs. 4).

Neither is the frequency response of the human ear a straightforward thing: the ear responds differently to sounds of different



frequencies and loudness levels. Although there is a small variation from person to person, normal ears agree within a few dB with the plot reproduced here (ISO Recommendation 226).



An even more significant peculiarity of the ear is its response to the pitch and bandwidth of a noise. Broadband sounds, like those of jet aircraft, seem much louder than narrow-band noise of the same sound pressure level. Thus accurate loudness measurements can be made only by taking into account the spectral distribution of the sound and relating it to empirically determined critical bandwidths. This phenomenon has given rise to the *Bark* scale: the audio range comprises 24 Bark, each of which equals the ear's critical bandwidth at a given center frequency.

Probably the most significant difference between objective and subjective measure of loudness occurs when two sounds are presented to the ear simultaneously. If the two sounds are widely separated in frequency, their partial loudnesses simply add to form the total loudness. But if they are not separated by a critical bandwidth, one sound masks the other: the closer together, the greater the influence. The noise analyst expresses this characteristic quantitatively in terms of *loudness density*, in sones/Bark.

The HP 8051A Loudness Analyzer is, in effect, a calibrated electronic ear that takes all of these subjective reactions of the human ear into consideration in measuring loudness based on ISO Recommendation 532 (Zwicker's Method). It listens to sound through a calibrated microphone or tape recorder, automatically produces a continuous spectral analysis and displays it as a plot of loudness density vs. subjective pitch. The instrument also computes and displays the total loudness of the sound, that is the integral of the Zwicker diagram.

The instrument is a great help in noise abatement studies because it shows how noise reduction techniques can be applied most effectively. Its spectral analysis points the finger at the most obvious sound-producing component, suggests what kind of sound-absorbing material may be needed, offers quick *before* and *after* comparisons of noise abatement programs.

A much more complex and versatile instrument for audio spectrum analysis, the recently announced HP 80501A Audio Data Processor combines the equivalent of a Loudness Analyzer with a powerful HP 2115A Digital Computer. The 80501A measures loudness with Kryter, Stevens, TALARM, SAE or dB weightings depending on the choice of standard computer programs. Results are available immediately: for example, the 80501A yields a complete analysis of aircraft noise while the plane is still overhead.

Our new 116-page Acoustics Handbook does justice to this rather complex subject. For your copy, write to Hewlett-Packard, 1502 Page Mill Road, Palo Alto, California 94304. In Europe: 1217 Meyrin-Geneva, Switzerland.

HEWLETT hp PACKARD

ANALYTICAL INSTRUMENTS

00970

LETTERS

Sirs:

The article "The First Electron Tube," by George Shiers, in the March issue of *Scientific American,* is attractive looking, with its many illustrations. But the title, as applied to Fleming's valve, and in fact the article as a whole are quite erroneous historically.

In the first place, electron tubes existed, and were being widely experimented with in Europe, well before Fleming, or even Edison. For example, the following book, published in London in 1870, has some 40 pictures of such tubes and their discharges, with 63 pages of text: the second volume of *A Physical Treatise on Electricity and Magnetism,* by J. E. H. Gordon.

The experimentation had started in the 1850's with the first sealed-in tubes made by Geissler, using his improved vacuum pump and using for excitation largely the newly devised Ruhmkorff induction coil. The properties of the discharges, reaction to magnetic fields and so on were studied by careful scientists throughout the latter half of the 19th century as higher vacuums were attained. The flowering began in the 1890's with the discovery of X rays from cathode rays, the identification of the electron and the devising of the cathode ray oscilloscope (the television picture tube). "The Development of the Electron Idea" is the subject of an illuminating article by one of the discoverers of the electron, W. Kaufmann, and of an accompanying editorial in *The Electrician* of London, November 9, 1901. Hence even the title was not left to Fleming. The accumulation of knowledge was collected in the first comprehensive textbooks on electronic science and technology, under the title *Electrical Conduction through Gases,* one by Johannes Stark (1901) and one by J. J. Thomson (1902). In them are to be found hundreds of references to the original scientific literature, mostly German, with precious little of Edison or Fleming.

These pioneering researches were conducted mostly on cold-electrode diodes. But the heated cathode, the thermionic vacuum tube, had also been discovered and studied, by Hittorf, Goldstein, and Elster and Geitel in the 1870's and 1880's, as Mr. Shiers mentions. But he goes on to say: "The knowledge gathered by the Germans does not appear to have been utilized by them for further research or for practical ends." That further research was done is amply registered in the scientific literature, ahead of Fleming, and we find a very practical end in the electric valve of Artur Wehnelt. It was the subject of German Patent No. 157843, filed January 15, 1904, issued January 13, 1905, titled *Electric Valve.* The one simple broad claim reads: "An electric valve characterized by a discharge tube with a hot metal-oxide cathode and a cold anode of any metal."

The term "metal oxide," in translation, refers to the barium and calcium oxides Wehnelt had found to greatly enhance the electron emission. He had first published on this effect in 1903 (*Sitzungsberichte der Physikalische-Medizinischen Sozietät zu Erlangen,* pages 150–158), and then on July 12, 1904 (*Annalen der Physik,* Vol. 14, pages 425–468), and in the *Physikalische Zeitschrift* of October 20, 1904, Vol. 5, pages 680–681. These last two papers went into the thermionic valve in considerable detail and exactness. A final paper, "An Electric Valve-Tube," in 1906 (*Annalen der Physik,* Vol. 19, pages 138–156) carries a footnote indicating that Fleming's use of the valve as a wireless detector would seem obvious!

Thus Fleming was not a "first," nor did his patent disclose anything that was very useful (the device was too insensitive), other than something on which Lee De Forest could build by adding an anode electromotive force and a telephone receiver, in his 1906 paper, and then the triode. That was to be matched by the corresponding von Lieben device of Europe, which proved to be the first electron amplifier.

LLOYD ESPENSCHIED

Kew Gardens, N.Y.

Sirs:

I thank Dr. Espenschied for his interest in my article. However, his comments are irrelevant to the theme—the evolution of elementary hot-cathode (thermionic) devices, generally known as "electron tubes." Cold-cathode (discharge) tubes were excluded because historically these are two distinct lines of development.

Serious studies of electric conduction in gases (cold-cathode tubes) were undertaken from the early 1850's by W. R. Grove, J. Plücker, J. P. Gassiot and M. Faraday. The induction coil, developed primarily by C. G. Page and N. J. Callan (1835–1838), improved by A. Fizeau, J. Foucault and H. D. Ruhmkorff (1851–1856), along with improved vacuum tubes and the mercury vacuum pump, both contributed by H. Geissler at this time, were vital components in these studies. The apparatus and the experiments quickly became prominent features in both professional and popular literature from 1860 onward.

In contrast, thermionic phenomena were almost completely ignored for 20 years. The valuable work of A. Wehnelt (1903–1904), concerning heated oxide emitters, was later incorporated in F. Braun's (1897) cathode ray indicator tube and in R. von Lieben's cathode ray telephone amplifier tube (1906). Thus two different paths merged at this time, but without any effect on the period (1880–1904) or results under discussion.

The somewhat irresponsible statement that the article is "quite erroneous" is not supported in Dr. Espenschied's letter. Therefore, since none of the items mentioned have any bearing on the matter or intent of the article, further comments on my part seem unnecessary.

GEORGE SHIERS

Santa Barbara, Calif.

Scientific American, August, 1969; Vol. 221, No. 2. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and treasurer.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Manuscripts are submitted at the author's risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Allan Wittman, Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 20 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 357B.

Please enter change of address ☐
Please enter subscription ☐
☐ 1 year, $8 ☐ 2 years, $15 ☐ 3 years, $21

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1919: "It goes without saying that such a radical departure as the theory of evolution could not become generally accepted without vast protest and a soul-satisfying fight. The question might appropriately be asked why Darwin was able to get his revolutionary ideas accepted, whereas other ideas, no more revolutionary and no more absurd from the standpoint of entrenched orthodoxy, never graduate from the 'nut' class. We would urge that the real reason Darwinism swept the world like wild-fire was the character of the first converts. Darwin was unbelievably fortunate in finding among scientists of standing, right at the outset, three disciples in as many countries who may fairly be characterized as zealots in the new science. Asa Gray in America, Thomas Huxley in Britain, Ernst Heinrich Haeckel on the Continent—their enthusiastic acceptance and championing of the new teachings was the chief instrumentality in overturning the established foundations of biology in a few months and starting all over again on the new basis. The first thought of every biologist reading that Haeckel had died in Jena on August 9th in his 86th year must have been that the last member of the quadrumvirate that had revolutionized the science had passed."

"Evidence sustaining the temperature hypothesis of a cooling globe has been obtained by the U.S. Geological Survey in the large number of thermometer readings made during the sinking of a well near Valley Falls, W.Va., to a depth of 7,579 feet. The temperature near the surface was 52°, whereas at a depth of 7,500 feet it was approximately 170°. A slight reduction in the steepness of the gradient is noticed below 6,500 feet. The well is the deepest in the world."

"The International Health Board (Rockefeller Foundation) continues to wage relentless war against yellow fever in the parts of the world that are regarded as the seed-beds of the disease. It is hoped that by this radical plan yellow fever may eventually be altogether eliminated from the world. Since last November the yellow fever work of the International Health Board has been directed by General Gorgas, late surgeon-general of the Army."

"Three times within the past few weeks accidents due to the inflammable hydrogen of the airship have shown with tragic emphasis how profound a peril rests in the highly combustible gas on which the lighter-than-air machine depends for its flotation. The outstanding lesson of this series of disasters to hydrogen-supported airships is that the elimination of hydrogen must be made the constant aim of aeronautical engineers and of the laboratories which are devoting time to aeronautical research. The wholesale substitution of helium would be the obvious solution, if only that ideal gas could be obtained at a reasonable cost in sufficient quantities to meet the demand. If travel through the air, whether by airplane or airship, is to engage the interest and secure the confidence of the civilian world, it must be made reasonably safe. It is certain that a continued recurrence of these horrible conflagrations will go far to kill the public confidence in travel by lighter-than-air machines."

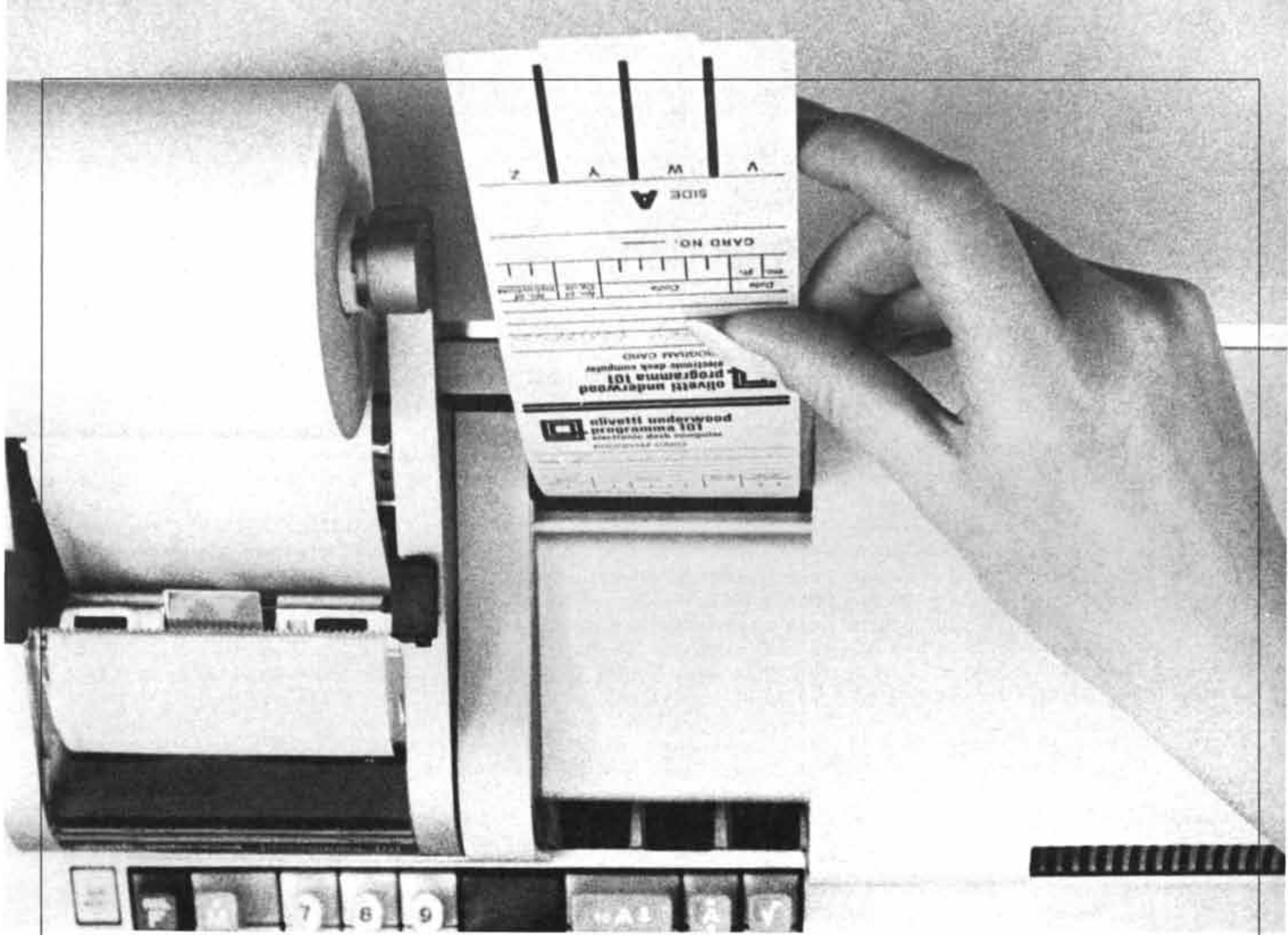
SCIENTIFIC AMERICAN

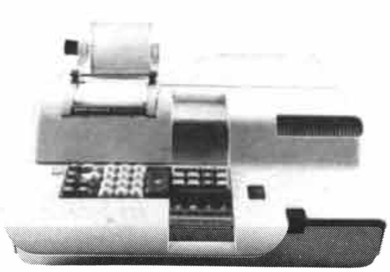
AUGUST, 1869: "It is officially announced by M. Lesseps that the ceremonies of the opening of the Suez canal will take place on the 17th of next November. The two great enterprises by which the year 1869 will be distinguished in history are the Pacific railroad and the Suez canal."

"The United States may as well look the subject of Chinese labor squarely in the face, and make timely provision to absorb and utilize this new accession to our population. Some are bitterly opposed to the coming of the Chinese. This opposition is based on groundless prejudice. The policy of the Government has hitherto opened the doors of immigration to people of every race and clime. Shall we now close it on the Mongolian, and if so, why? We have heretofore spoken of the intelligence, industry, frugality and order-loving disposition of the Chinese. Since we assert that the Chinese character possesses in an eminent degree the qualities we have ever been taught to regard as the elements of citizenship, we do not see how it is possible, with any show of consistency, to attempt, either by persecution or legislation, to shut our doors against them. The Chinaman wants to work for us and we want him. Then let an end be speedily put to the disgraceful treatment he has hitherto received, a blot upon the history of the 'Golden State' which makes humanity blush. Let us welcome him, with all the rest of the oppressed and suffering who now find refuge here, confident that by the process of assimilation we can absorb, and render homogeneous, the mixed races which are destined to people this continent."

"John A. Roebling, C.E., whose fame as an engineer has made his name familiar throughout the civilized world, died at the residence of his son in Hicks Street, Brooklyn, on the 22nd of July. His death resulted from lockjaw, caused from an injury to his foot, which rendered amputation necessary. The bruise was received while he was in company with his son engaged in surveying the approaches to the projected East River Suspension Bridge, about to be erected between New York and Brooklyn. That the elder Roebling has been cut off thus on the threshold of his greatest undertaking adds to our sincere regret, but that he could not live to see the bridge's completion will not detract from the well-won renown of its gifted and accomplished designer."

"The returns from the different scientific expeditions sent out to various points in the Middle West to observe the solar eclipse have been only such as have been made in the daily papers, and the details are extremely meager. The principal points to be determined in the observation of this eclipse were, first and foremost, the nature so far as could be ascertained of the rose-colored prominences, second, the true nature of the solar corona, and third, the existence or nonexistence of planets between the orbit of Mercury and the sun. The belief that the corona is concentric with the sun will be open to question, if the reports that reach us are correct in regard to the appearance presented by it in this eclipse. The form is stated to have been rhomboidal rather than circular, as hitherto observed. The rose-colored protuberances appeared to the number of five or six. The results of the observations made on them are not, however, yet sufficiently collated to justify any positive conclusion. So far as we can gather, no planets within the orbit of Mercury were discovered."





THE AUTHORS

HERBERT F. YORK ("Military Technology and National Security") is professor of physics at the University of California at San Diego and a member of the general advisory committee of the U.S. Arms Control and Disarmament Agency. York returned to teaching in 1964 after a long career as an administrator and Government official; he was director of the Livermore Laboratory from 1952 to 1958; chief scientist with the Advanced Research Projects Agency of the Department of Defense in 1958; director of defense research and engineering in the office of the Secretary of Defense from 1958 to 1961; chancellor of the University of California at San Diego from 1961 to 1964, and twice a member of the President's Science Advisory Committee, serving as its vice-chairman in the latter years of the Johnson Administration. He was graduated from the University of Rochester in 1942 and received a master's degree there in 1943; he obtained his Ph.D. from the University of California at Berkeley in 1949.

LUIGI LUCA CAVALLI-SFORZA ("'Genetic Drift' in an Italian Population") is visiting professor in the department of genetics at the Stanford University School of Medicine; his permanent position is professor of genetics and chairman of the department of genetics at the University of Pavia. "I have worked in bacterial genetics since 1942," he writes, "and have also worked on antibiotic resistance and other bacteriological problems. My interest in human population genetics started when I was lecturing in genetics at Parma in 1952. I have also engaged in an overall study of human evolution and racial differentiation in man and, in the past four winters, in expeditions to the Central African Republic and other places in Africa for an investigation of African pygmies." Cavalli-Sforza obtained a degree in medicine at the University of Pavia in 1944 and a master's degree at the University of Cambridge in 1950.

NEWELL C. COOK ("Metalliding") is a research associate in the organic chemistry branch of the chemical laboratory of the General Electric Research and Development Center. He was graduated from Florida Southern College in 1939 and obtained his Ph.D. from Pennsylvania State University in 1943. From 1945 to 1950, when he joined General Electric, he was assistant professor of chemistry at Penn State. His research interests include hydrocarbon chemistry, high-temperature reactions and electrolysis in fused salts.

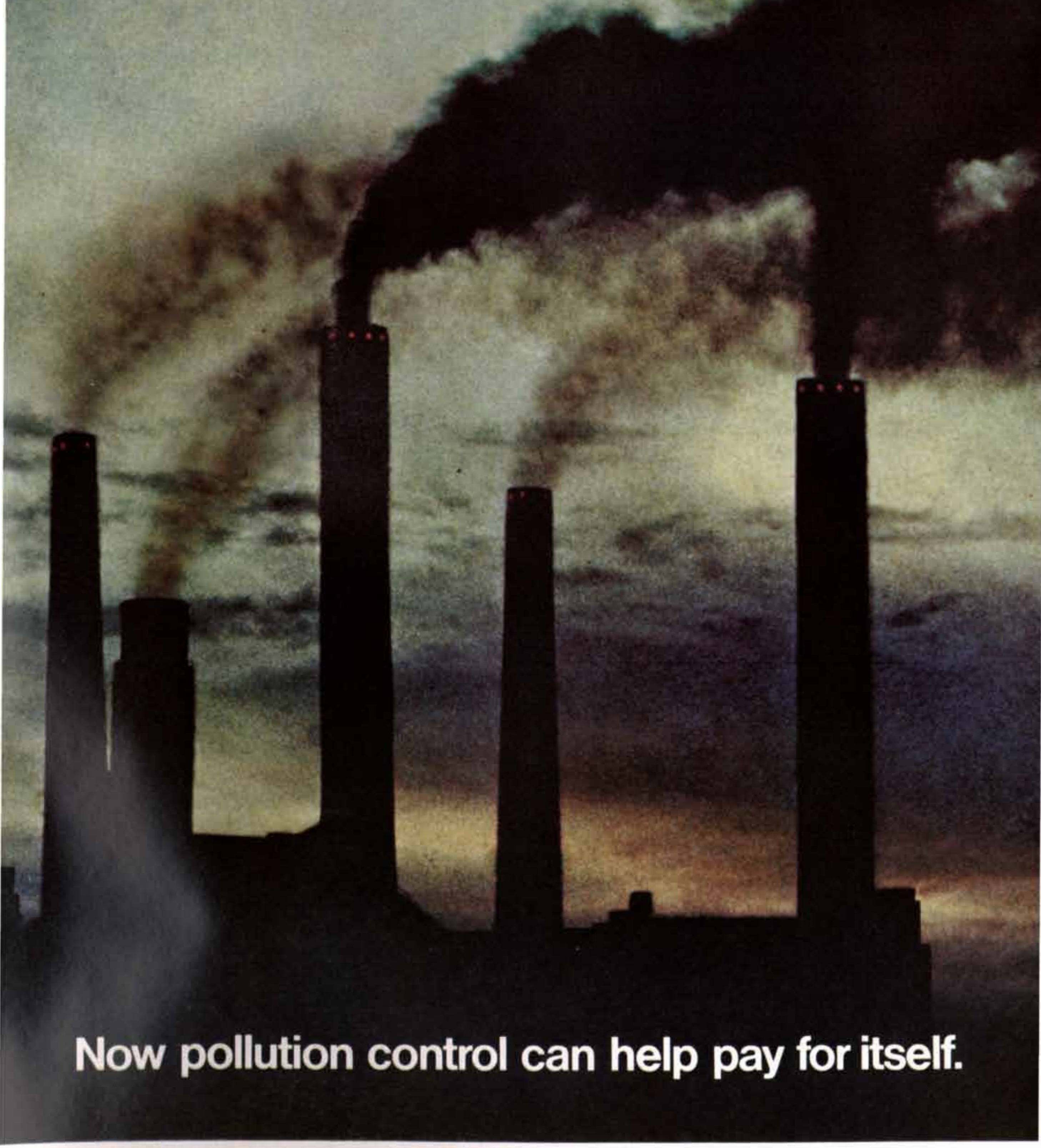
MICHEL BARANGER and RAYMOND A. SORENSEN ("The Size and Shape of Atomic Nuclei") were together until recently as professors of physics at Carnegie-Mellon University; in July, Baranger took up an appointment at the Massachusetts Institute of Technology. Baranger writes: "I came to the U.S. in 1949 after receiving most of my graduate education in France, where I was born. The graduate school I attended in Paris is the École Normale Supérieure, from 1945 to 1949. In those days, however, physics education in France was terribly old-fashioned; for instance, I had to learn quantum mechanics entirely by myself out of books." Baranger obtained his Ph.D. from Cornell University in 1951 and went to the Carnegie Institute of Technology in 1955. Sorensen took his bachelor's, master's and doctor's degrees at Carnegie Tech between 1953 and 1958. He became a member of the faculty there in 1961 after a year as a postdoctoral fellow at the Niels Bohr Institute in Copenhagen and two years at Columbia University.

A. P. OKLADNIKOV ("The Petroglyphs of Siberia") is rector of the Institute of History, Philology and Philosophy of the Siberian Branch of the Academy of Sciences of the U.S.S.R. He is also an academician of the Academy of Sciences of the U.S.S.R. Born in a Siberian village where his father was a country teacher and his mother a peasant, he studied in Irkutsk and then at the Academy of History of Material Culture in Leningrad, which is now known as the Leningrad Branch of the Institute of Archaeology of the Academy of Sciences. He received his doctor's degree from the University of Leningrad in 1947. From 1928 to 1935 he worked in the Irkutsk Museum and from 1938 to 1961 in the Leningrad Branch of the Institute of Archaeology. He has been in the Siberian Branch of the Academy of Sciences since 1961. Okladnikov has led a number of archaeological expeditions.

R. D. B. FRASER ("Keratins") is chief research scientist in the Division of Protein Chemistry of the Commonwealth Scientific and Industrial Research Organization of Australia. "I graduated in physics in 1948 at King's College in the University of London," he writes, "and then studied the application of infrared spectrometry to biological problems in the Medical Research Council Biophysics Unit at King's College. For this work I was awarded a Ph.D. in biophysics in 1951. In 1952 I took up an appointment with C.S.I.R.O. in Melbourne to study keratin structure by infrared spectrometry. I was awarded the degree of D.Sc. by the University of London in 1960 for contributions to the study of the structure of biological molecules and currently work with a small group who share an interest in unraveling the intricacies of fibrous protein structure. Outside the laboratory my main recreations are flying and trout fishing."

GERALD L. KOOYMAN ("The Weddell Seal") is assistant research physiologist in the Physiological Research Laboratory of the Scripps Institution of Oceanography at the University of California at San Diego. He was graduated from the University of California at Los Angeles in 1957 and received his Ph.D. from the University of Arizona in 1966, joining Scripps after a year as a postdoctoral fellow at London Hospital Medical College. Kooyman writes: "My broad scientific interest is environmental physiology, which I find particularly appealing because it requires sound knowledge of the physical aspects of the environment as well as understanding of the behavior and physiology of the species in question. At present my specific interests relate to the adaptation of marine birds and mammals to aquatic environments, especially with regard to deep diving and the concomitant effects of pressure."

LYNWOOD BRYANT ("Rudolf Diesel and His Rational Engine") is professor of history at the Massachusetts Institute of Technology, where he teaches American intellectual history and American economic history. He was graduated from Harvard College in 1929 with a degree in English literature and did graduate work in American civilization at Harvard University. Bryant writes: "I am not an engineer, only an observer of engineering—a kind of sidewalk superintendent. My chief professional interest is in the history, philosophy and sociology of technology. In recent years I have concentrated on the history of the automobile and the internal-combustion engine." He has also been interested in technical writing and publication; he was director of the M.I.T. Press from 1957 to 1962. He was the author of "The Origin of the Automobile Engine" in the March 1967 issue of Scientific American.



With our Cat-Ox process it's a real possibility. The Monsanto catalytic oxidation system cleans all of the fly ash from the stacks of coal-burning plants. Then converts 90% of the sulfur dioxide from stack gases into salable sulfuric acid.

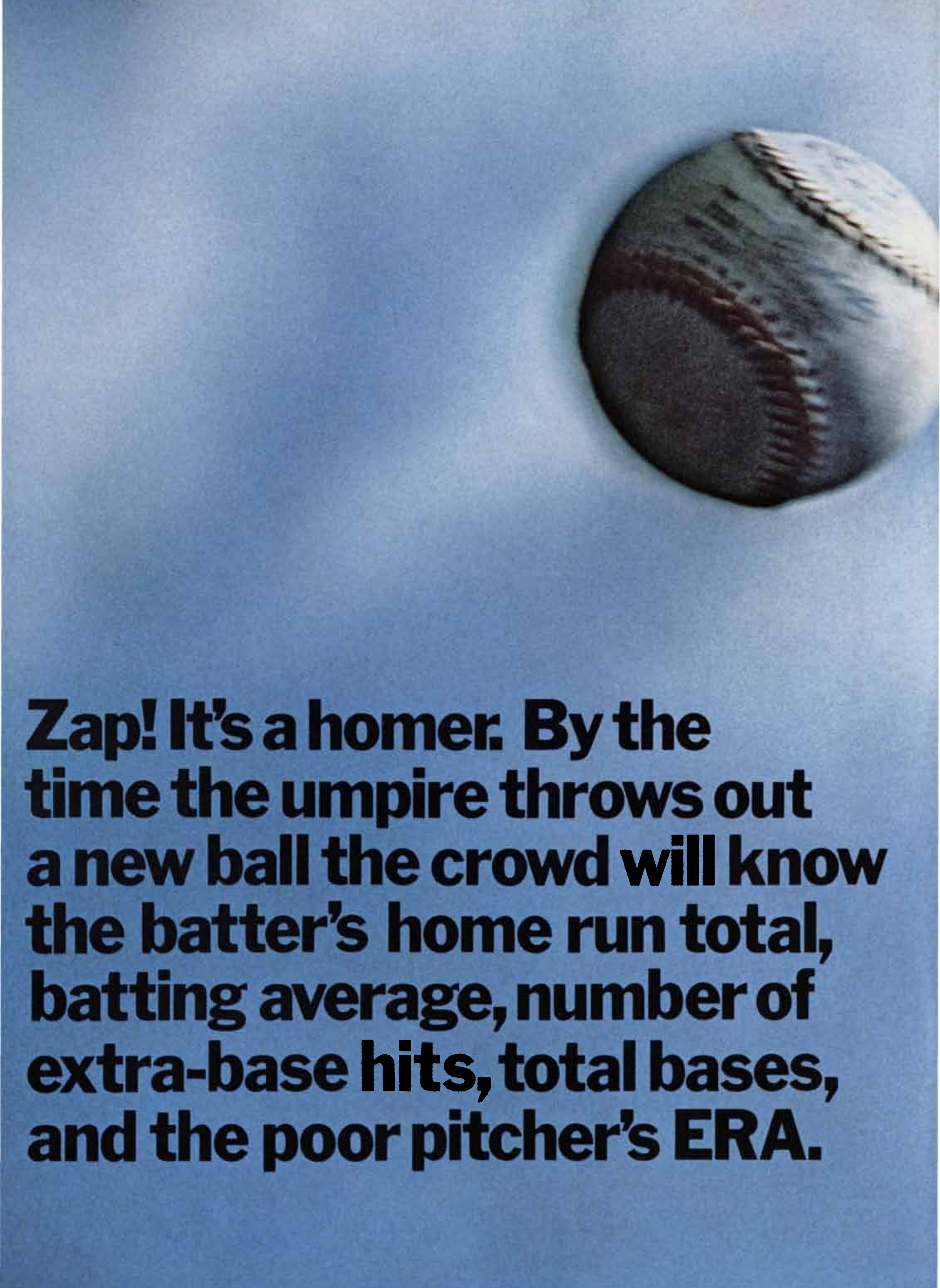
But that's only one way Monsanto's new development helps make pollution control self-supporting.

Heat extracted from the gas helps generate more electricity. Tall, expensive stacks aren't needed. And low-cost, high-sulfur coal can be used without washing.

At Monsanto, we're constantly solving problems by exploring the total science spectrum. Monsanto—where science and research are working for people.

Monsanto Company; 800 N. Lindbergh Blvd.; St. Louis, Missouri 63166

Monsanto


Zap! It's a homer. By the time the umpire throws out a new ball the crowd will know the batter's home run total, batting average, number of extra-base hits, total bases, and the poor pitcher's ERA.

Remember Glendon, the kid who spent all summer reciting batting averages? There's a little bit of Glendon in all of us. "Look, Harry, this guy's batting .295 and he's hit safely in 11 straight games. We walk him, right?" Zap.

The fans get all the facts they need to manage their team from the scoreboard at the Oakland-Alameda County Coliseum. The Athletics put in a computer to collect, store and calculate the statistics. We engineered and installed the world's largest computer output terminal—a scoreboard with twin 126-foot units—one to get the dope out where 50,000 fans could read it, fast... another to display instant animated cartoons and comment.

The hunger for instant baseball information is just one small symptom of a national mania... call it Glendonitis. We have to know what's going on. Now. Right. Clear.

"What's the last quote on General Tel 5s?

"Have you a first-class seat to Tulsa next Tuesday?"

"Miss, would you please tell me the balance in my checking account? Number 536-377-9842."

"Wasn't that the turn for the turnpike west, dear?"

One of our special areas is providing all sorts of gear to get the answers to people who need them. A giant electronically controlled bond board for a major New York bank. Systems that let bank tellers ask questions of a computer and get the answers on a TV screen. A computerized election-returns display for a TV network. Militarized TV monitors that show a pilot what's on the ground when his eyes tell him it's as black as a hat.

We're even dreaming of new types of highway signs that react to conditions. Imagine a sign that senses that there's a traffic slowdown 10 miles ahead and adds,

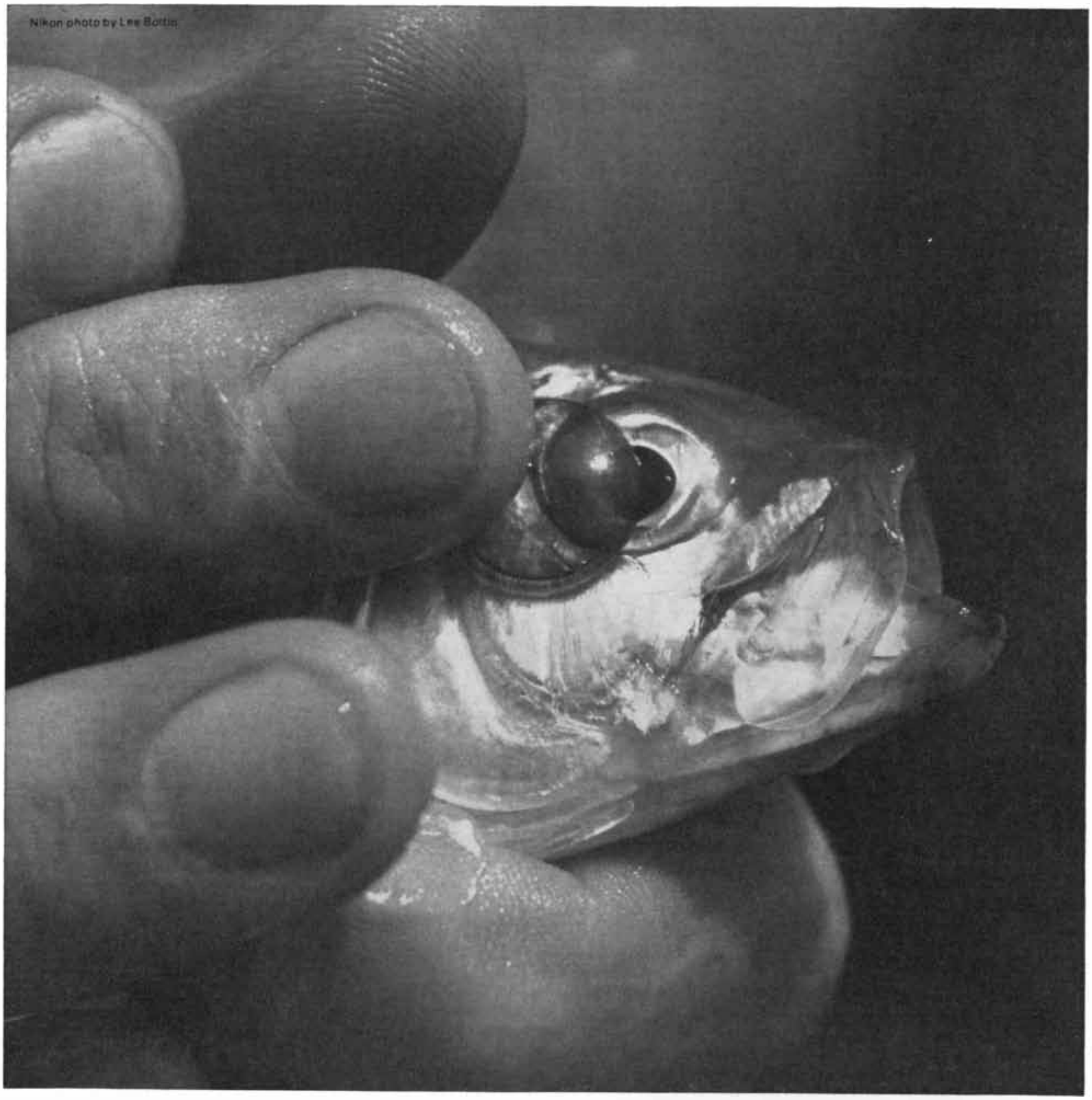
"JAM AHEAD—LEFT FOR BEST ROUTE TO SHORE."

Nobody knows what we'll do next. Not even us. At the moment we're in sensing systems, aerospace flight controls, numerically controlled machine tools, computer peripheral equipment, information display, automatic assembly.

We're not really selling products, though. We're selling answers. Any questions?

Conrac Corporation,
330 Madison Avenue,
New York 10017.

CONRAC
CORPORATION


Nikon photo by Lee Boltin


Nikon

Established 1845 SCIENTIFIC AMERICAN August 1969 Volume 221 Number 2

Military Technology and National Security

The ABM debate is analyzed in the context of a larger dilemma: the futility of searching for technological solutions to what is essentially a political problem

by Herbert F. York

The recent public hearings in the Senate and the House of Representatives on anti-ballistic-missile (ABM) systems have provided an unprecedented opportunity to expose to the people of this country and the world the inner workings of one of the dominant features of our time: the strategic arms race. Testimony has been given by a wide range of witnesses concerning the development and deployment of all kinds of offensive and defensive nuclear weapons; particular attention has been paid to the interaction between decisions in these matters and the dynamics of the arms race as a whole.

In my view the ABM issue is only a detail in a much larger problem: the feasibility of a purely technological approach to national security. What makes the ABM debate so important is that for the first time it has been possible to discuss a major aspect of this larger problem entirely in public. The reason for this is that nearly all the relevant facts about the proposed ABM systems either are already declassified or can easily be deduced from logical concepts that have never been classified. Thus it has been possible to consider in a particular case such questions as the following:

1. To what extent is the increasing complexity of modern weapons systems and the need for instant response causing strategic decision-making authority to pass from high political levels to low military-command levels, and from human beings to machines?

2. To what extent is the factor of secrecy combined with complexity leading to a steadily increasing dominance of military-oriented technicians in some vital areas of decision-making?

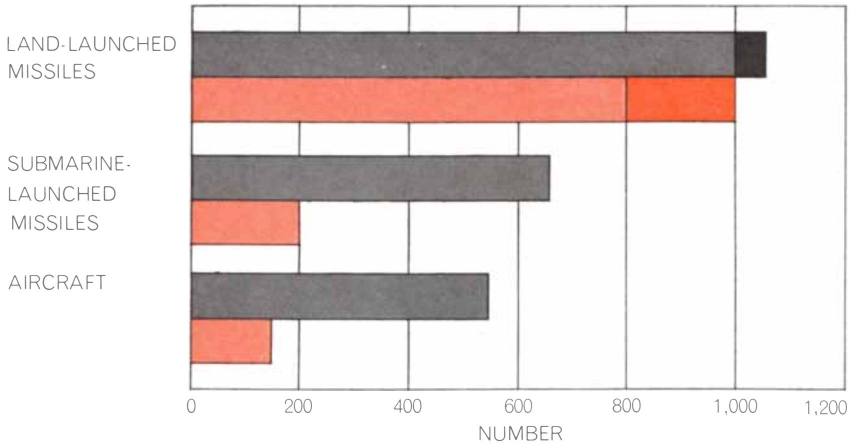
3. To what extent do increasing numbers of weapons and increasing complexity—in and of themselves—complicate and accelerate the arms race?

My own conclusion is that the ABM issue constitutes a particularly clear example of the futility of searching for technical solutions to what is essentially a political problem, namely the problem of national security. In support of this conclusion I propose in this article to review the recent history of the strategic arms race, to evaluate what the recent hearings and other public discussions have revealed about its present status and future prospects, and then to suggest what might be done now to deal with the problem of national security in a more rational manner.

The strategic arms race in its present form is a comparatively recent phenomenon. It began in the early 1950's, when it became evident that the state of the art in nuclear weaponry, rocket propulsion and missile guidance and control had reached the point in the U.S. where a strategically useful intercontinental ballistic missile (ICBM) could be built. At about the same time the fact that a major long-range-missile development program was in progress in the U.S.S.R. was confirmed. As a result of the confluence of these two events the tremendous U.S. long-range-missile program, which dominated the technological scene for more than a decade, was undertaken. The Air Force's Thor, Atlas and Titan programs and the Army's Jupiter program were started almost simultaneously; the Navy's Polaris program and the Air Force's Minuteman program were phased in just a few years later.

More or less at the same time the Army, which had had the responsibility for ground-based air defense (including the Nike Ajax and Nike Hercules surface-to-air missiles, or SAM's), began to study the problem of how to intercept ICBM's, and soon afterward initiated the Nike Zeus program. This program was a straightforward attempt to use existing technology in the design of a nuclear-armed rocket for the purpose of intercepting an uncomplicated incoming warhead. The Air Force proposed more exotic solutions to the missile-defense problem, but these were subsequently absorbed into the Defender Program of the Department of Defense's Advanced Research Projects Agency (ARPA). The Defender Program included the study of designs more advanced than Nike Zeus, and it also incorporated a program of down-range measurements designed to find out what did in fact go on during the terminal phases of missile flight.

By 1960 indications that the Russians were taking the ABM prospect seriously, in addition to progress in our own Nike Zeus program, stimulated our offensive-



PRESENT STATUS of the deployment of strategic offensive forces by the U.S. (*gray*) and the U.S.S.R. (*color*) shows that the two superpowers are about even in numbers of intercontinental ballistic missiles (ICBM's), and that the U.S. is ahead in both long-range aircraft and submarine-launched ballistic missiles (SLBM's) of the Polaris type. The U.S. ICBM's consist almost entirely of Minutemen (*light gray*), which carry a nuclear warhead with an explosive yield in the megaton range; there currently remain only 54 of the larger Titans (*dark gray*) in our strategic forces. The smaller Russian missiles (*light color*) are mostly SS-11's, which are roughly equivalent in size to our Minutemen. The larger Russian missiles (*dark color*) are SS-9's, which are comparable in size to our Titans. The figures used are from a speech given by Secretary of Defense Melvin R. Laird on March 25.

missile designers into seriously studying the problem of how to penetrate missile defenses. Very quickly a host of "penetration aid" concepts came to light: light and heavy decoys, including balloons, tank fragments and objects resembling children's jacks; electronic countermeasures, including radar-reflecting clouds of the small wires called chaff; radar blackout by means of high-altitude nuclear explosions; tactics such as barrage, local exhaustion and "rollback" of the defense, and, most important insofar as the then unforeseen consequences were concerned, the notion of putting more than one warhead on one launch vehicle. At first this notion simply involved a "shotgun" technique, good only against large-area targets (cities), but it soon developed into what we now call MIRV's (multiple independently targeted reentry vehicles), which can in principle (and soon in practice) be used against smaller, harder targets such as missile silos, radars and command centers.

This avalanche of concepts forced the ABM designers to go back to the drawing board, and as a result the Nike-X concept was born in 1962. The Nike-X designers attempted to make use of more sophisticated and up-to-date technology in the design of a system that they hoped might be able to cope with a large, sophisticated attack. All through the mid-1960's a vigorous battle of defensive concepts and designs versus offensive concepts and designs took place. This battle was waged partly on the Pacific Missile Range but mostly on paper and in committee meetings. It took place generally in secret, although parts of it have been discussed in earlier articles in this magazine [see "National Security and the Nuclear-Test Ban," by Jerome B. Wiesner and Herbert F. York, October, 1964; "Anti-Ballistic-Missile Systems," by Richard L. Garwin and Hans A. Bethe, March, 1968; "The Dynamics of the Arms Race," by George W. Rathjens, April, 1969].

This intellectual battle culminated in a meeting that took place in the White House in January, 1967. In addition to President Johnson, Secretary of Defense Robert S. McNamara and the Joint Chiefs of Staff there were present all past and current Special Assistants to the President for Science and Technology (James R. Killian, Jr., George B. Kistiakowsky, Jerome B. Wiesner and Donald F. Hornig) and all past and current Directors of Defense Research and Engineering (Harold Brown, John S. Foster, Jr., and myself). We were asked that simple kind of question which must be answered after all the complicated ifs, ands and buts have been discussed: "Will it work?" The answer was no, and there was no dissent from that answer. The context, of course, was the Russian threat as it was then interpreted and forecast, and the current and projected state of our own ABM technology.

Later that year Secretary McNamara gave his famous San Francisco speech in which he reiterated his belief that we could not build an ABM system capable of protecting us from destruction in the event of a Russian attack. For the first time, however, he stated that he did believe we could build an ABM system able to cope with a hypothetical Chinese missile attack, which by definition would be "light" and uncomplicated. In recommending that we go ahead with a program to build what came to be known as the Sentinel system, he said that "there are *marginal* grounds for concluding that a light deployment of U.S. ABM's against this possibility is prudent." A few sentences later, however, he warned: "The danger in deploying this relatively light and reliable Chinese-oriented ABM system is going to be that pressures will develop to expand it into a heavy Soviet-oriented ABM system." The record makes it clear that he was quite right in this prediction.

Meanwhile the U.S.S.R. was going ahead with its own ABM program. The Russian program proceeded by fits and starts, and our understanding of it was, as might be supposed in such a situation, even more erratic. It is now generally agreed that the only ABM system the Russians have deployed is an area defense around Moscow much like our old Nike Zeus system. It appears to have virtually no capability against our offense, and it has been, as we shall see below, extremely counterproductive insofar as its goal of defending Moscow is concerned.

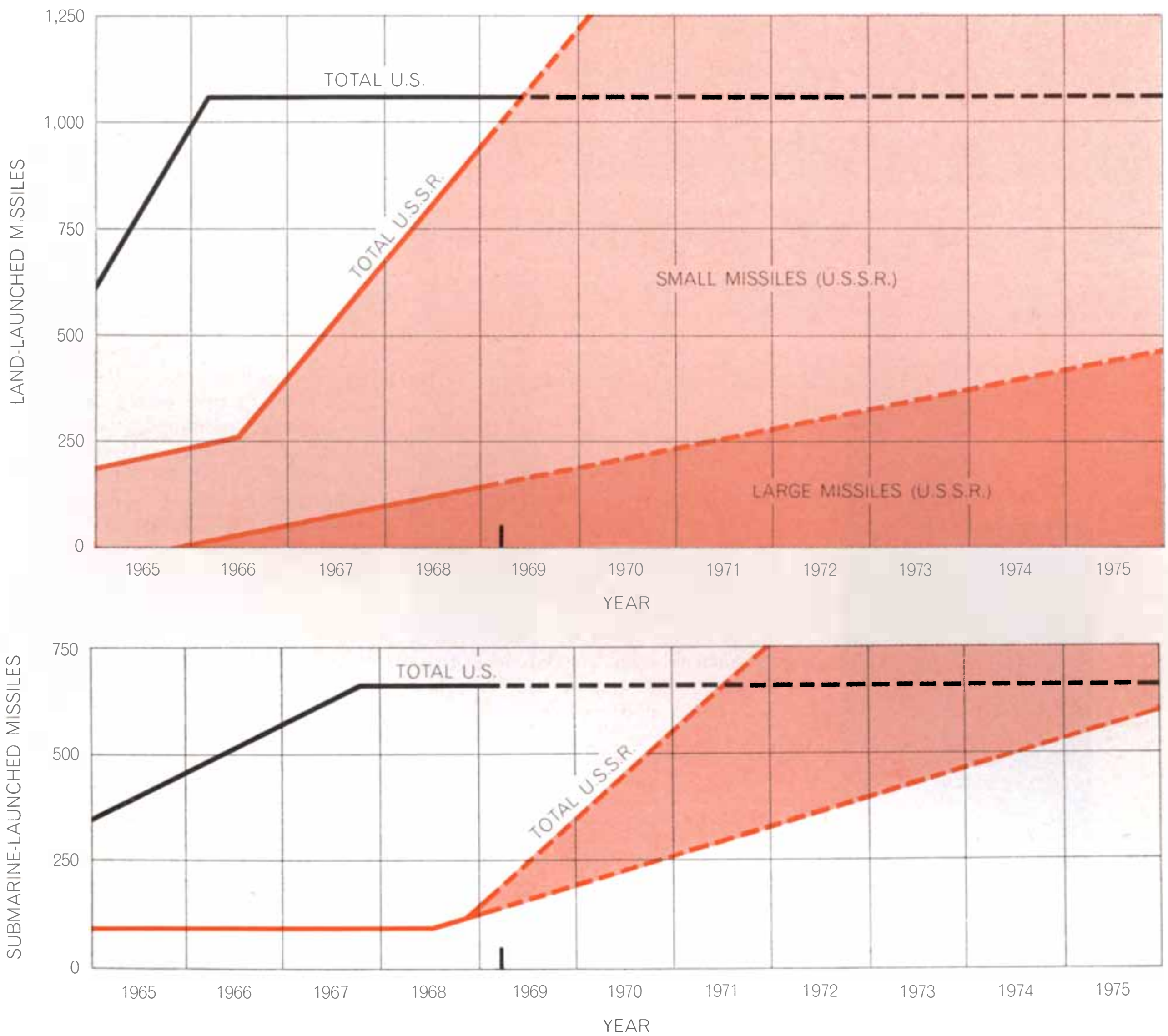
Development and deployment of offensive-weapons systems on both sides progressed rapidly during the 1960's, but rather than discuss these historically I shall go directly to the picture that the Administration has given of the present status and future projection of such forces.

Data recently presented by the Department of Defense show that the U.S. and the U.S.S.R. are about even in numbers of intercontinental missiles, and that the U.S. is ahead in both long-range aircraft and submarines of the Polaris type [*see illustration on this page*]. The small Russian missiles are mostly what we call SS-11's, which were described in the hearings as being roughly the equivalent of our Minutemen. The large Russian missile is what we call the SS-9. Deputy Secretary of Defense David Packard characterized its capability as one 20-megaton warhead or three five-megaton warheads. Our own missiles are almost entirely the smaller Minutemen. There currently remain only 54 of the larger Titans in our strategic forces. Not covered in the table are "extras" such as the U.S.S.R.'s FOBS (fractional orbital bombardment system) and IRBM's (intermediate-range ballistic missiles), nor the U.S.'s bombardment aircraft deployed on carriers and overseas bases in Europe and elsewhere. There are, of course, many important details that do not come out clearly in such a simple tabular presentation; these include payload capacity, warhead yield, number of warheads per missile and, often the most important, warhead accuracy.

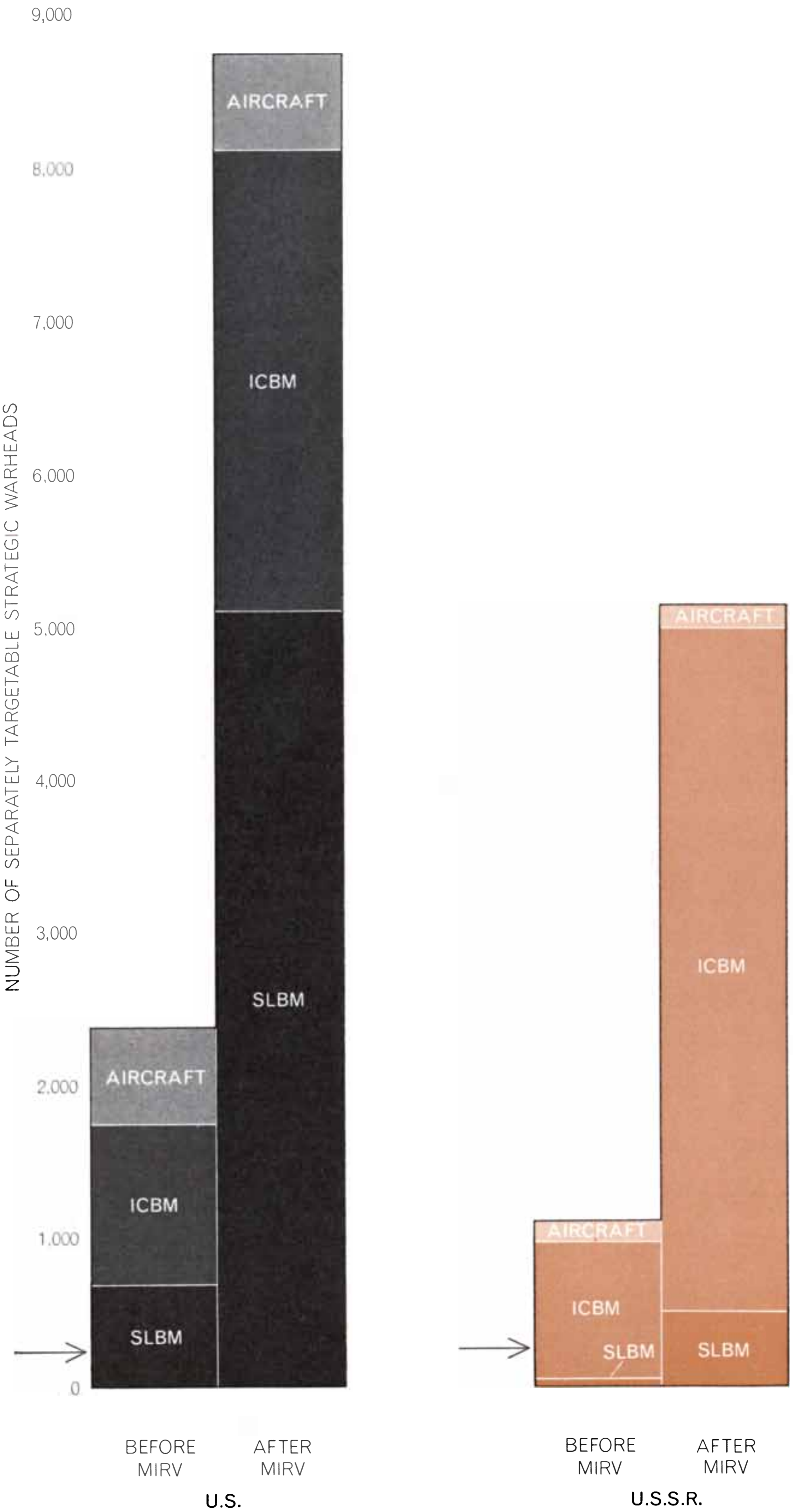
In the area of defensive systems designed to cope with the offensive systems outlined above, both the U.S. and the U.S.S.R. have defenses against bombers that would probably be adequate against a prolonged attack using chemical explosives (where 10 percent attrition is enough) and almost certainly inadequate against a nuclear attack (where 10 percent penetration is enough). In addition the U.S.S.R. has its ineffective ABM deployment around Moscow, usually estimated as consisting of fewer than 100 antimissile missiles.

What all these complicated details add up to can be expressed in a single word: parity. This is clearly not numerical equality in the number of warheads or in the number of megatons or in the total "throw weight"; in fact, given different design approaches on the two sides, simultaneous equality in these three figures is entirely impossible. It is, rather, parity with respect to strategic objectives; that is, in each case these forces are easily sufficient for deterrence and entirely insufficient for a successful preemptive strike. In the jargon of strategic studies either side would retain, after a massive "first strike" by the other, a sufficiently large "assured destruction capability" against the other in order to deter such a first strike from being made.

There is much argument about exactly what it takes in the way of "assured destruction capability" in order to deter, but even the most conservative strategic



EXTRAPOLATED TRENDS in the deployment of strategic offensive missiles by the U.S. and the U.S.S.R. are indicated by the broken lines in this pair of charts, which are based on a presentation by Deputy Secretary of Defense David Packard to the Senate Foreign Relations Committee on March 26. The chart at top shows the numbers of deployed ICBM's for both sides during the period 1965–1975. The Russian total is broken down into "small" missiles (the one-megaton SS-11's) and "large" missiles (the multimegaton SS-9's). The chart at bottom shows the deployed SLBM's during the same period. The extrapolations suggest that the Russians will be even with us in ICBM's quite soon and will catch up in SLBM's sometime between 1971 and 1977. One important factor omitted from the charts is the imminent deployment by both sides of MIRV's (multiple independently targetable reentry vehicles).



IMPACT OF MIRV on the strategic balance is emphasized in this chart, which is based on one prepared by the staff of the Senate Foreign Relations Committee and presented by Senator Albert Gore of Tennessee on March 26. The chart depicts the strategic balance in terms of separately targetable strategic warheads before and after MIRVing, which is expected to take place in the next five years. The two black arrows near the bottom indicate the number of warheads that could devastate the 50 largest cities on each side.

planners conclude that the threat of only a few hundred warheads exploding over population and industrial centers would be sufficient for the purpose. The large growing disparity between the number of warheads needed for the purpose and the number actually possessed by each side is what leads to the concept of "overkill." If present trends continue, in the future all or most missiles will be MIRVed, and so this overkill will be increased by perhaps another order of magnitude.

Here let me note that it is sometimes argued that there is a disparity in the present situation because Russian missile warheads are said to be bigger than U.S. warheads, both in weight and megatonnage; similarly, it is argued that MIRVing does not increase overkill because total yield is reduced in going from single to multiple warheads. This argument is based on the false notion that the individual MIRV warheads of the future will be "small" when measured against the purpose assigned to them. Against large, "soft" targets such as cities bombs *very much* smaller than those that could be used as components of MIRV's are (and in the case of Hiroshima were proved to be) entirely adequate for destroying the heart of a city and killing hundreds of thousands of people. Furthermore, in the case of small, "hard" targets such as missile silos, command posts and other military installations, having explosions bigger than those for which the "kill," or crater, radius slightly exceeds "circular error probable" (CEP) adds little to the probability of destroying such targets. Crater radius depends roughly on the cube root of the explosive power; consequently, if during the period when technology allows us to go from one to 10 warheads per missile it also allows us to improve accuracy by a little more than twofold, the "kill" per warhead will remain nearly the same in most cases, whereas the number of warheads increases tenfold.

In any case, it is fair to say that in spite of a number of such arguments about details, nearly everyone who testified at the ABM hearings agreed that the present situation is one in which each side possesses forces adequate to deter the other. In short, we now have parity in the only sense that ultimately counts.

Several forecasts have been made of what the strategic-weapons situation will be in the mid-1970's. In most respects here again there is quite general agreement. Part of the presentation by Deputy Secretary Packard to the Senate Foreign Relations Committee on March

26 were two graphs showing the trends in numbers of deployed offensive missiles beginning in 1965 and extending to 1975 [*see illustrations on page 19*]. There is no serious debate about the basic features of these graphs. It is agreed by all that in the recent past the U.S. has been far ahead of the U.S.S.R. in all areas, and that the Russians began a rapid deployment program a few years ago that will bring them even with us in ICBM's quite soon and that, if extended ahead without any slowdown, would bring them even in submarine-launched ballistic missiles (SLBM's) sometime between 1971 and 1977.

One important factor that the Department of Defense omitted from its graphs is MIRV. Deployment plans for MIRV's have not been released by either the U.S. or the U.S.S.R., although various rough projections were made at the hearings about numbers of warheads per vehicle (three to 10), about accuracies (figures around half a mile were often mentioned, and it was implied that U.S. accuracies were better than Russian ones) and about development status (the U.S. was said to be ahead in developments in this field). A pair of charts emphasizing the impact of MIRV was prepared by the staff of the Senate Foreign Relations Committee [*see illustration on opposite page*].

One could argue with both of these sets of charts. For example, one might wonder why the Senate charts show so few warheads on the Russian Polaris-type submarine and why they show only three MIRV's on U.S. Minutemen; on the other hand, one might wonder whether the Department of Defense's projected buildup of the Russian Polaris fleet could be that fast, or whether one should count the older Russian missile submarines. Nonetheless, the general picture presented cannot be far wrong. Moreover, the central arguments pursued throughout the ABM hearings (in both the Senate Foreign Relations Committee hearings in March and the Senate Armed Services Committee hearings in April) were not primarily concerned with these numerical matters. Rather, they were concerned with (1) Secretary of Defense Melvin R. Laird's interpretation of these numbers insofar as Russian *intentions* were concerned, (2) the validity of the Safeguard ABM system as a response to the purported strategic problems of the 1970's and (3) the arms-race implications of Safeguard.

As for the matter of intentions, those favoring the ABM concept generally held that the only "rational" explanation of the Russians' recent SS-9 buildup, coupled with their multiple-warhead development program and the Moscow ABM system, was that they were aiming for a first-strike capability. One must admit that almost anything is conceivable as far as intentions are concerned, but there certainly are simpler, and it seems to me much more likely, explanations. The simplest of all is contained in Deputy Secretary Packard's chart. The most surprising feature of this chart is the fact that the Russians were evidently satisfied with being such a poor second for such a long time. This is made more puzzling by the fact that all during this period U.S. defense officials found it necessary to boast about how far ahead we were in order to be able to resist internal pressures for still greater expansion of our offensive forces.

Another possible reason, and one that I believe added to the other in the minds of the Russian planners, was that their strategists concluded in the mid-1960's that, whatever the top officials here might say, certain elements would eventually succeed in getting a large-scale ABM system built, and that penetration-aid devices, including multiple warheads, would be needed to meet the challenge. Whether or not they were correct in this latter hypothetical analysis is still uncertain at this writing. Let us, however, pass on from this question of someone else's intentions and consider whether or not the proposed Safeguard ABM system is a valid, rational and necessary response to the Russian deployments and developments outlined above.

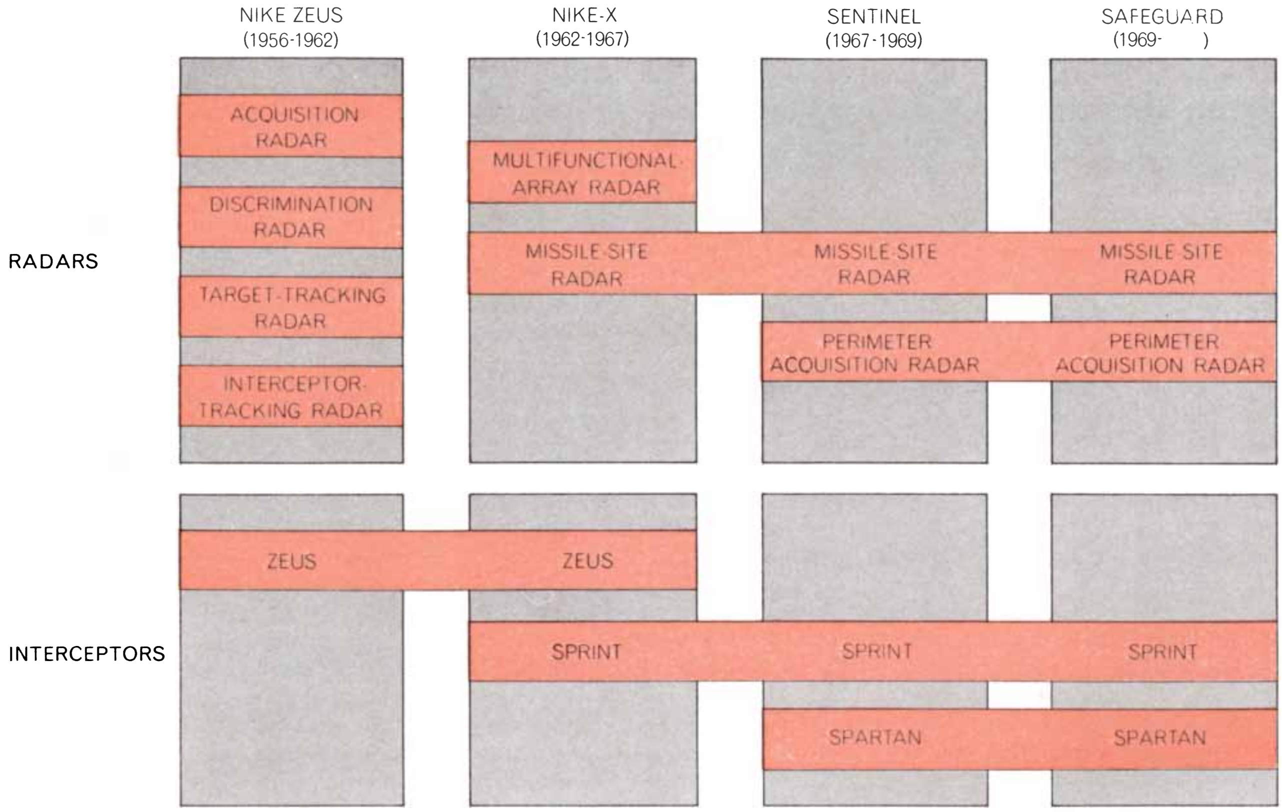
To many of those who have recently written favorably about ABM defenses or who have testified in their favor before the Congressional committees, Safeguard is supported mainly as a prototype of something else: a "thick" defense of the U.S. against a massive Russian missile attack. This is clearly not at all the rationale for the Safeguard decision as presented by President Nixon in his press conference of March 14, nor is it implied as more than a dividend in the defense secretaries' testimony. The President said that he wanted a system that would protect a *part* of our Minuteman force in order to increase the credibility of our deterrent, and that he had overruled moving in the direction of a massive city defense because "even starting with a thin system and then going to a heavy system tends to be more provocative in terms of making credible a first-strike capability against the Soviet Union. I want no provocation which might deter arms talks." The top civilian defense officials give this same rationale, although they put a little more emphasis on the "prototype" and "growth potential" aspects of the system. For simplicity and clarity I shall focus on the Administration's proposal, as stated in open session by responsible officials.

From a technical point of view and as far as components are concerned, President Nixon's Safeguard system of today is very little different from President Johnson's Sentinel system [*see illustrations on page 24*]. There are only minor changes in the location of certain components (away from cities), and elements have been added to some of the radars so that they can now observe submarine-launched missiles coming from directions other than directly from the U.S.S.R. and China. As before, the system consists of a long-range interceptor carrying a large nuclear weapon (Spartan), a fast short-range interceptor carrying a small nuclear weapon (Sprint), two types of radar (perimeter acquisition radar, or PAR, and missile-site radar, or MSR), a computer for directing the battle, and a command and control system for integrating Safeguard with the national command. I shall not describe the equipment in detail at this point but pass on directly to what I believe can be concluded from the hearings and other public sources about each of the following four major questions: (1) Assuming that Safeguard could protect Minuteman, is it needed to protect our deterrent? (2) Assuming that Safeguard "works," can it in fact safeguard Minuteman? (3) Will it work? (4) Anyway, what harm can it do?

First: Assuming that Safeguard could protect Minuteman, is it needed to protect our deterrent?

Perhaps the clearest explanation of why the answer to this first question is "no" was given by Wolfgang K. H. Panofsky before the Senate Armed Services Committee on April 22. He described how the deterrent consists of three main components: Polaris submarines, bombers and land-based ICBM's. Each of these components alone is capable of delivering far more warheads than is actually needed for deterrence, and each is currently defended against surprise destruction in a quite different way. ICBM's are in hard silos and are numerous. Polarises are hidden in the seas. Bombers can be placed on various levels of alert and can be dispersed.

Since the warning time in the case of an ICBM attack is generally taken as being about 30 minutes, the people who believe the deterrent may be in serious danger usually imagine that the bomb-



EVOLUTION OF U.S. ABM SYSTEMS is represented in this illustration, which is adapted from a chart introduced by Daniel J. Fink in his testimony before the Senate Foreign Relations Committee on March 6. In general the radar components of the successive designs have progressed from slow, mechanically steered, single-function radars to fast, electronically steered, multifunction radars. The slow Zeus ABM missile has been superseded by the short-range Sprint (for terminal defense) and the long-range Spartan (for area defense). The components of the Safeguard system are the same as those that were originally intended for the earlier Sentinel system.

ers are attacked by missile submarines, and therefore have only a 15-minute warning. This is important because a 30-minute warning gives the bombers ample time to get off the ground. In that case, however, an attack on all three components cannot be made simultaneously; that is, if the attacking weapons are launched simultaneously, they cannot arrive simultaneously, and vice versa.

Thus it is incredible that all three of our deterrent systems could become vulnerable in the same time period, and it is doubly incredible that we could not know that this would happen without sufficient notice so that we could do something about it. There is, therefore, no basis for a frantic reaction to the hypothetical Russian threat to Minuteman. Still, it is sensible and prudent to begin thinking about the problem, and so we turn to the other questions. We must consider these questions in the technological framework of the mid-1970's, and we shall do this now in the way defense officials currently seem to favor: by assuming that this is the best of all possible technological worlds, that everything works as intended and that direct extrapolations of current capabilities are valid.

Second: Assuming that Safeguard "works," can it in fact safeguard Minuteman?

One good approach to this problem is the one used by George W. Rathjens in his testimony before the Senate Armed Services Committee on April 23. His analysis took as a basis of calculation the implication in Secretary Laird's testimony that the Minuteman force may become seriously imperiled in the mid-1970's. Rathjens then estimated how many SS-9's would have to be deployed at that time in order to achieve this result. From this number, and the estimate of the current number of SS-9's deployed, he got a rate of deployment. He also had to make an assumption about how many Sprints and Spartans would be deployed at that time, and his estimates were based on the first phase of Safeguard deployment. These last numbers have not been released, but a range of reasonable values can be guessed from the cost estimates given. Assuming that the SS-9's would have four or five MIRV warheads each by that time, Rathjens found that by prolonging the SS-9 production program by a few months the Russians would be able to cope with Safeguard by simply exhausting it and would still have enough warheads left to imperil Minuteman, if that is indeed their intention [*see illustration on page 29*].

The length of this short safe period does depend on the numbers used in the calculations, and they of course can be disputed to a degree. Thus if one assumes that it takes fewer Russian warheads to imperil Minuteman (it can't be less than one for one!), then the assumed deployment rate is lower and the safe period is lengthened; on the other hand, if one notes that the missile-site radars in our system are much softer than even today's silos, then the first attacking warheads, fired directly at the radars, can be smaller and less accurate, so that a higher degree of MIRVing can be used for attacking these radars and a shorter safe period results. To go further, it was suggested that the accuracy/yield combination of the more numerous SS-11's might be sufficient for attacking the missile-site radars, and therefore, if the Russians were to elect such an option, there would be no safe period at all. In short, the most that Safeguard can do is either delay somewhat the date when Minuteman would be imperiled or cause the attacker to build up his forces at a somewhat higher rate if indeed imperiling Minuteman by a fixed date is his purpose.

In the more general case this problem is often discussed in budgetary terms, and the "cost-exchange ratio" between offense and defense is computed for a wide variety of specific types of weapon. Such calculations give a wide variety of results, and there is much argument about them. However, even using current offense designs (that is, without MIRV), such calculations usually strongly favor the offense. This exchange ratio varies almost linearly with the degree of MIRVing of the offensive missiles, and therefore it seems to me that in the ideal technological future we have taken as our context this exchange ratio will still more strongly favor the offense.

Third: Will it work? By this question I mean: Will operational units be able to intercept enemy warheads accompanied by enemy penetration aids in an atmosphere of total astonishment and uncertainty? I do not mean: Will test equipment and test crews intercept U.S. warheads accompanied by U.S. penetration aids in a contrived atmosphere? A positive answer to the latter question is a necessary condition for obtaining a positive answer to the former, but it is by no stretch of the imagination a sufficient condition.

This basic question has been attacked from two quite different angles: by examining historical analogies and by examining the technical elements of the problem in detail. I shall touch on both here. Design-oriented people who consider this a purely technical question emphasize the second approach. I believe the question is by no means a purely technical question, and I suggest that the historical-analogy approach is more promising, albeit much more difficult to use correctly.

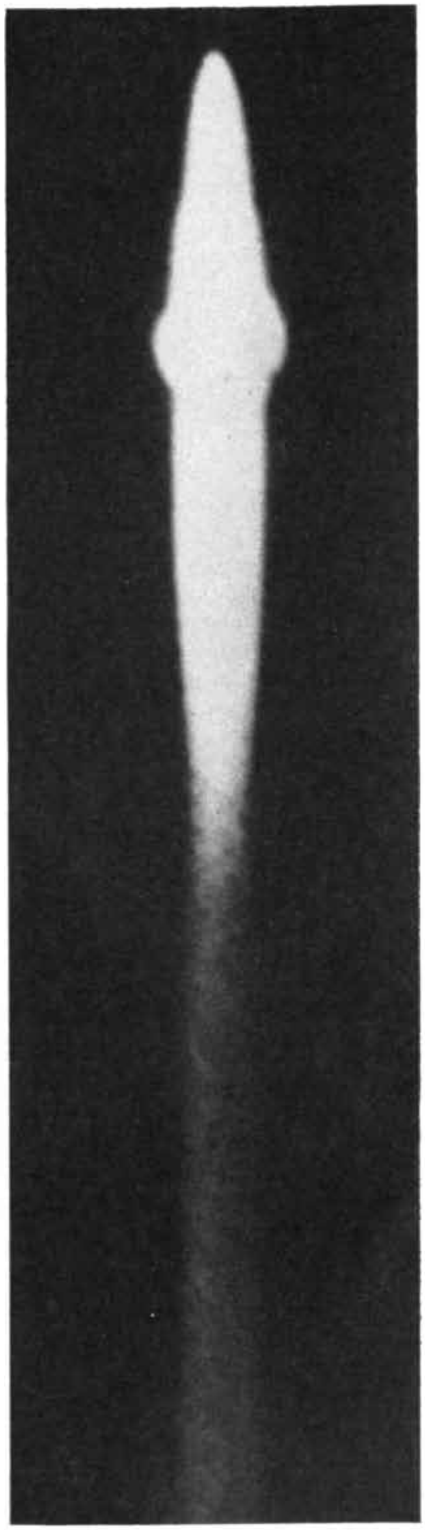
False analogies are common in this argument. We find that some say: "You can't tell me that if we can put a man on the moon we can't build an ABM." Others say: "That's what Oppenheimer told us about the hydrogen bomb." These two statements contain the same basic error. They are examples of successes in a contest between technology and nature, whereas the ABM issue involves a contest between two technologies: offensive weapons and penetration aids versus defensive weapons and discrimination techniques. These analogies would be more pertinent if, in the first case, someone were to jerk the moon away just before the astronauts landed, or if, in the second case, nature were to keep changing the nuclear-reaction probabilities all during the development of the hydrogen bomb and once again after it was deployed.

Proper historical analogies should involve modern high-technology defense systems that have actually been installed and used in combat. If one examines the record of such systems, one finds that they do often produce some attrition of the offense, but not nearly enough to be of use against a nuclear attack. The most up-to-date example is provided by the Russian SAM's and other air-defense equipment deployed in North Vietnam. This system "works" after a fashion because both the equipment designers and the operating crews have had plenty of opportunities to practice against real U.S. targets equipped with real U.S. countermeasures and employing real U.S. tactics.

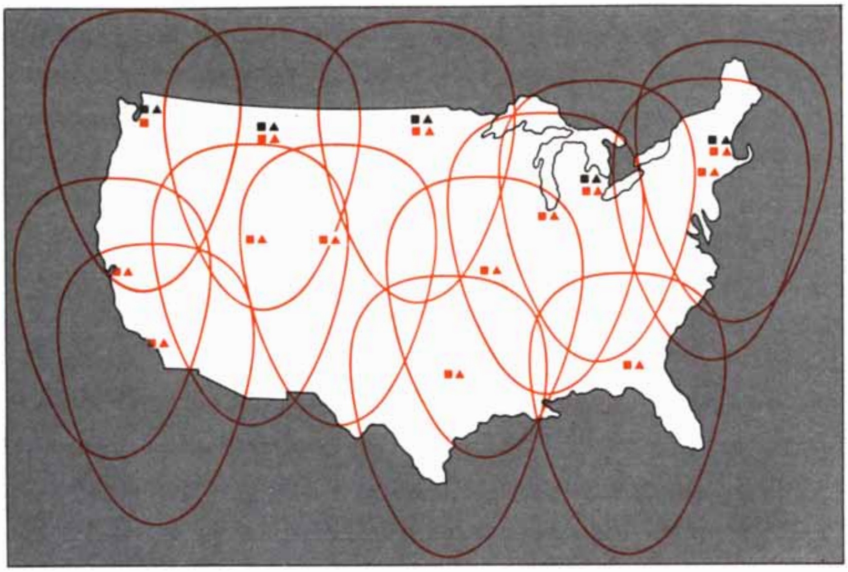
The best example of a U.S. system is somewhat older, but I believe it is still relevant. It is the SAGE system, a complex air-defense system designed in the early 1950's. All the components worked on the test range, but by 1960 we came to realize, even without combat testing, that SAGE could not really cope with the offense that was then coming into being. We thereupon greatly curtailed and modified our plans, although we did continue with some parts of the system. To quote from the recent report on the ABM decision prepared by Wiesner, Abram Chayes and others: "Still, after fifteen years, and the expenditure of more than $20 billion, it is generally conceded that we do not have a significant capability to defend ourselves against a well-planned air attack. The Soviet Union, after even greater effort, has probably not done much better."

So much for analogies; let us turn to the Safeguard system itself. Doubts about its being able to work were raised during the public hearings on a variety of grounds, some of which are as follows:

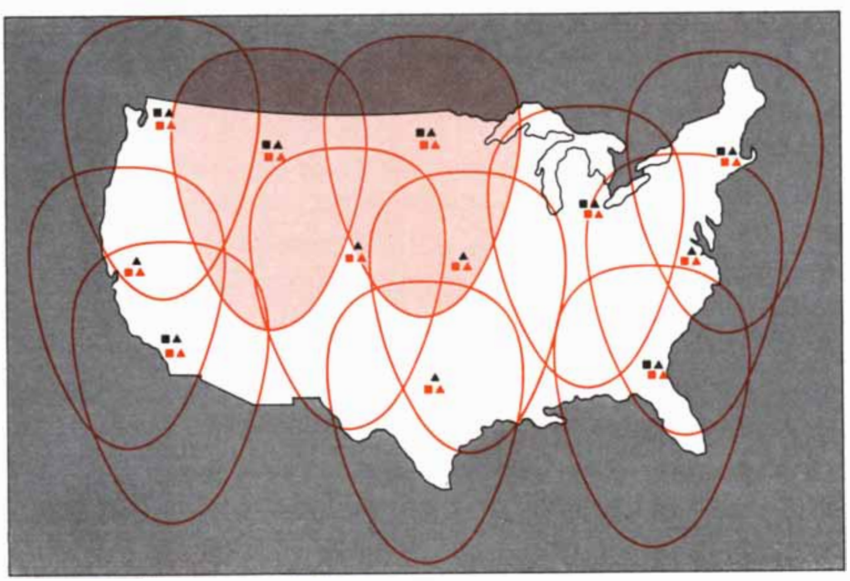
First, and perhaps foremost, there is the remarkable fact that the new Safeguard system and the old Sentinel system use virtually the same hardware deployed in a very similar manner, and yet they have entirely different primary purposes. Sentinel had as its purpose defending large soft targets against the so-called Chinese threat. The Chinese threat by definition involved virtually no sophisticated penetration aids and no



SPRINT ABM MISSILE was photographed during a test flight at the White Sands Missile Range in New Mexico. The photograph, which was released by the U.S. Army's Nike-X Project Office in March, 1966, shows the second stage of the missile heated to incandescence by friction with the atmosphere. Because of the extremely high speed of the Sprint, the missile's skin in places reaches temperatures hotter than inside its rocket motor. The bulges are created by the guidance fins at the rear of the second stage.



SENTINEL SYSTEM was described by the Johnson Administration as a "thin" ABM system designed to defend the U.S. against a hypothetical Chinese missile attack in the 1970's. The main defense was to be provided by long-range Spartan missiles. The Spartans would be deployed at about 14 locations in order to provide an area defense of the whole country. The range of each "farm" of Spartans is indicated by the egg-shaped area around it; for missiles attacking over the northern horizon the intercept range of the Spartan is elongated somewhat to the south. The Sentinel system would also include some short-range Sprint missiles, which were originally to be deployed to defend the five or six perimeter acquisition radars, or PAR's, which were to be deployed at five sites located across the northern part of the country. Missile-site radars, or MSR's, were to be deployed at every ABM site.



SAFEGUARD SYSTEM, President Nixon's proposed modification of the Sentinel scheme, uses essentially the same components in a slightly different array to accomplish an entirely different primary purpose: the defense of a part of our Minuteman force against a hypothetical surprise attack by the Russians. Phase I of Safeguard covers the construction of ABM sites at two Minuteman "fields": one near Malmstrom Air Force Base in Montana and the other near Grand Forks Air Force Base in North Dakota (*colored areas*). The completed system would have a total of 12 sites, each with Sprint and Spartan coverage, located somewhat farther away from the cities. In addition two new PAR sites would be included in order to observe submarine-launched missiles coming from directions other than due north.

possibilities of exhausting the defense; thus were "solved" two of the most difficult problems that had eliminated Nike Zeus and Nike-X.

Safeguard has as its primary purpose defending a part of the Minuteman force against a Russian attack. It is not credible that a Russian attack against the part of the Minuteman force so defended would be other than massive and sophisticated, so that we are virtually right back to trying to do what in 1967 we said we could not do, and we are trying to do it with no real change in the missiles or the radars. It is true that defending hard points is to a degree easier than defending cities because interception can be accomplished later and at lower altitudes, thus giving discrimination techniques more time to work. Moreover, only those objects headed for specific small areas must be intercepted. These factors do make the problem somewhat easier, but they do not ensure its solution, and plenty of room for doubt remains.

Second, there is the contest between penetration aids and discrimination techniques. This was discussed at length by Garwin and Bethe in their March 1968 article in *Scientific American* and mentioned also in varying degrees of detail by many of those who testified recently concerning the ABM issue. The Russian physicist Andrei D. Sakharov, in his essay "Thoughts on Progress, Coexistence and Intellectual Freedom," put the issue this way: "Improvements in the resistance of warheads to shock waves and the radiation effects of neutron and X-ray exposure, the possibility of mass use of relatively light and inexpensive decoys that are virtually indistinguishable from warheads and exhaust the capabilities of an antimissile defense system, a perfection of tactics of massed and concentrated attacks, in time and space, that overstrain the defense detection centers, the use of orbital and fractional-orbital attacks, the use of active and passive jamming and other methods not disclosed in the press—all of this has created technical and economic obstacles to an effective missile defense that, at the present time, are virtually insurmountable."

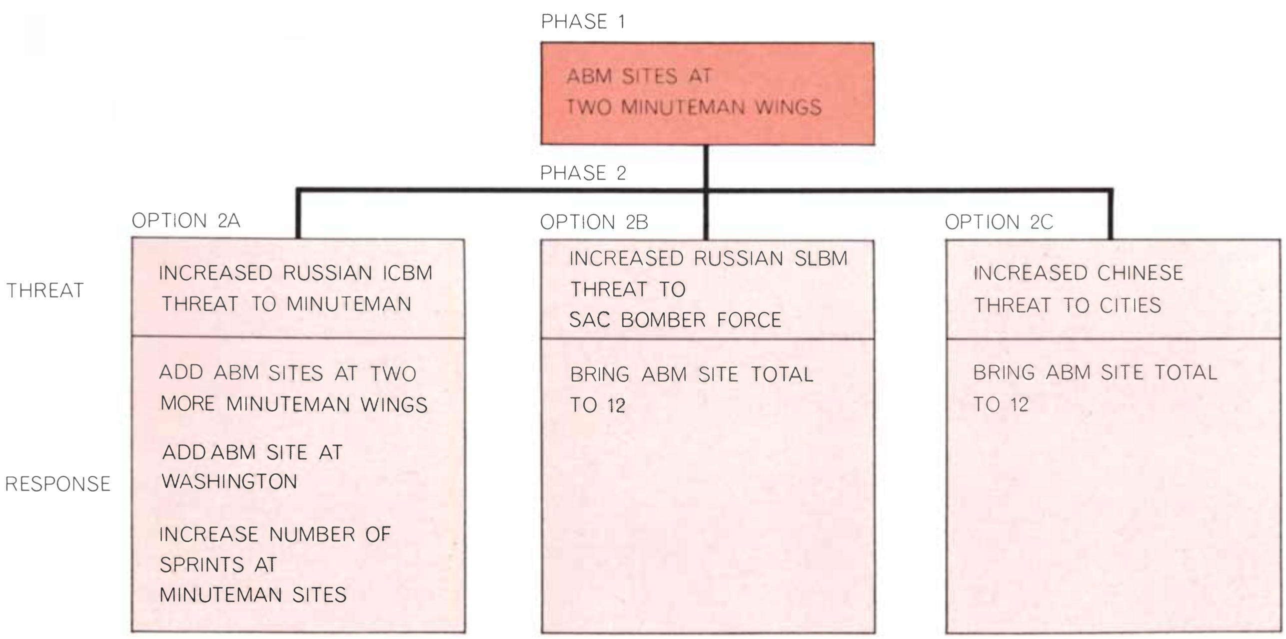
I would add only MIRV to Sakharov's list. Pitted against this plethora of penetration aids are various observational methods designed to discriminate the real warheads. Some of the penetration devices obviously work only at high altitudes, but even these make it necessary for the final "sorting" to be delayed, and thus they still contribute to making the defense problem harder. Other devices can continue to confuse the defense even down to low altitudes. Some of the problems the offense presents to the defense can no doubt be solved (and have been solved) when considered separately and in isolation. That is, they can be solved for a time, until the offense designers react. One must have serious reservations, however, whether these problems can ever be solved for any long period in the complex combinations that even a modestly sophisticated attacker can present. Further, such a contest *could* result in a catastrophic failure of the system in which all or nearly all interceptions fail.

Third, there is the unquantifiable difference between the test range and the real world. The extraordinary efforts of the Air Force to test operationally deployed Minutemen show that it too regards this as an important problem. Moreover, the tests to date do seem to have revealed important weaknesses in the deployed forces. The problem has many aspects: the possible differences between test equipment and deployed equipment; the certain differences between the offensive warheads and penetration aids supplied by us as test targets and the corresponding equipment and tactics the defense must ultimately be prepared to face; the differences between the installation crews at a test site and at a deployment site; the differences in attitudes and motivation between a test crew and an operational crew (even if it is composed of the same men); the differences between men and equipment that have recently been made ready and whom everyone is watching and men and equipment that have been standing ready for years during which nothing happened; the differences between the emotional atmosphere where everyone knows it is not "for real" and the emotional atmosphere where no one can believe what he has just been told. It may be that all that enormously complex equipment will be ready to work the very first time it must "for real," and it may be that all those thousands of human beings have performed all their interlocking assignments correctly, but I have very substantial doubts about it.

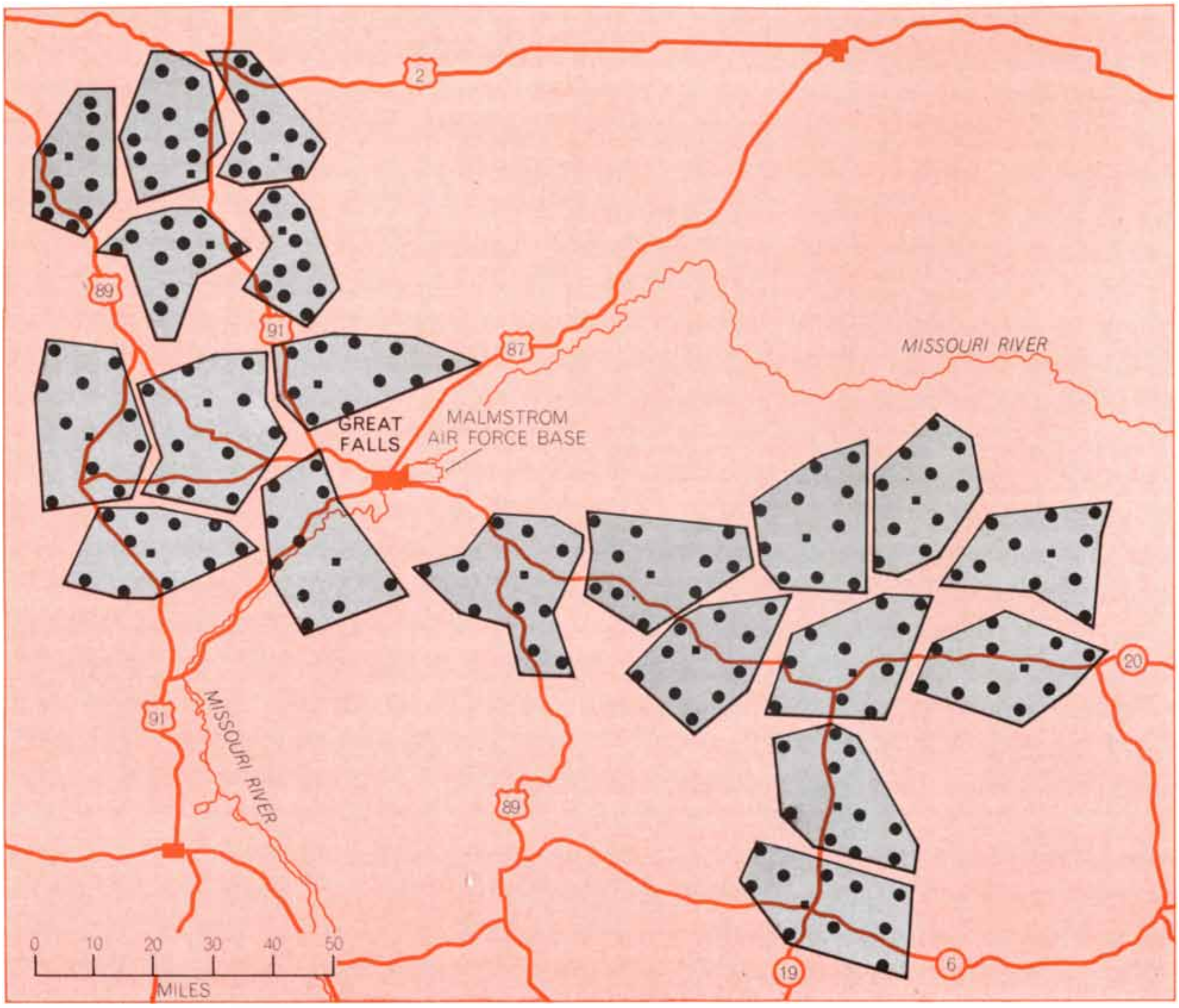
Fourth, there is the closely related "hair-trigger/stiff-trigger" contradiction. Any active defense system such as Safeguard must sit in readiness for two or four or eight years and then fire at precisely the correct second following a warning time of only minutes. Furthermore, the precision needed for the firing time is so fine that machines must be used to choose the exact instant of firing no matter how the decision to fire is made. In the case of offensive missiles the situation is different in an essential way: Although maintaining readiness throughout a long, indefinite period is necessary, the moment of firing is not so precisely controlled in general and hence human decision-makers, including even those at high levels, may readily be permitted to play a part in the decision-making process. Thus if we wish to be certain that the defense will respond under conditions of surprise, the trigger of the ABM system, unlike the triggers of the ICBM's and Polarises, must be continuously sensitive and ready—in short, a hair trigger—for indefinitely long periods of time.

On the other hand, it is obvious that we cannot afford to have an ABM missile fire by mistake or in response to a false alarm. Indeed, the Army went to some pains to assure residents of areas near proposed Sentinel sites that it was imposing requirements to ensure against the accidental launching of the missile and the subsequent detonation of the nuclear warhead it carries. Moreover, Army officials have assured the public that no ABM missiles would ever be launched without the specific approval of "very high authorities."

These two requirements—a hair trigger so that the system can cope with a surprise attack and a stiff trigger so that it will never go off accidentally or without proper authorization—are, I believe, contradictory requirements. In saying this I am not expressing doubt about the stated intentions of the present Army leaders, and I strongly endorse the restrictions implied in their statements. I am saying, however, that if the system cannot be fired without approval of "the highest authorities," then the probability of its being fired under conditions of surprise is less than it would be otherwise. This probability depends to a degree on the highly classified technical details of the Command and Control System, but in the last analysis it depends more on the fact that "the highest authority" is a human being and therefore subject



SAFEGUARD PHASE II provides for three different optional responses to various potential threats in the 1970's. A possible further addition would be sites in Alaska and Hawaii. This chart is also adapted from Deputy Secretary Packard's testimony on March 26.



MINUTEMAN MISSILE BASE in the vicinity of Malmstrom Air Force Base is shown on this map, which is based on information released by the Department of the Air Force. The Minuteman missiles are grouped in 20 flights of 10 missiles each for a total of 200 missiles. Every flight has its own control center, each of which is capable of launching an entire squadron of 50 missiles.

to all the failures and foibles pertaining thereto.

This brings us to our fourth principal question: Anyway, what harm can it do?

We have just found that the total deterrent is very probably not in peril, that the Safeguard system probably cannot safeguard Minuteman even if it "works," that there is, to say the least, considerable uncertainty whether or not it will "work." Nonetheless, if there were no harm in it, we might be prudent and follow the basic motto of the arms race: "Let us err on the side of military safety." There seem to be many answers to the question of what harm building an ABM system would do. First of all, such a system would cost large sums of money needed for nondefense purposes. Second, it would divert money and attention from what may be better military solutions to the strategic problems posed by the Administration. Third, it would intensify the arms race. All these considerations were discussed at the hearings; I shall comment here only on the third, the arms-race implications of the ABM decision.

It is often said that an ABM system is not an accelerating element in the arms race because it is intrinsically defensive. For example, during the hearings Senator Henry M. Jackson of Washington, surely one of the best-informed senators in this field, said essentially that, and he quoted Premier Kosygin as having said the same thing. I believe such a notion is in error and is based on what we may call "the fallacy of the last move." I believe that in the real world of constant change in both the technology and the deployed numbers of all kinds of strategic-weapons systems, ABM systems are accelerating elements in the arms race. In support of this view let us recall one of the features of the history recited at the start of this article.

At the beginning of this decade we began to hear about a possible Russian ABM system, and we became concerned about its potential effects on our ICBM and Polaris systems. In response the MIRV concept was invented. Today there are additional justifications for MIRV besides penetration, but that is how it started. Now, the possibility of a Russian MIRV is used as one of the main arguments in support of the Safeguard system. Thus we have come one full turn around the arms-race spiral. No one in 1960 and 1961 thought through the potential destabilizing effects of multiple warheads, and certainly no one predicted, or even could have predicted, that the inexorable logic of the arms race would carry us directly from Russian talk in 1960 about defending Moscow against missiles to a requirement for hard-point defense of offensive-missile sites in the U.S. in 1969.

By the same token I am sure the Russians did not foresee the large increase in deployed U.S. warheads that will ultimately result from their ABM deployment and that made it so counterproductive. Similarly, no one today can describe in detail the chain reaction the Safeguard deployment would lead to, but it is easy to see the seeds of a future acceleration of the arms race in the Nixon Administration's Safeguard proposal. Soon after Safeguard is started (let us assume for now that it will be) Russian offense planners are going to look at it and say something such as: "It may not work, but we must be prudent and assume it will." They may then plan further deployments, or more complex penetration systems, or maybe they will go to more dangerous systems such as bombs in orbit. A little later, when some of our optimistic statements about how "it will do the job it is supposed to do" have become part of history, our strategic planners are going to look at Safeguard and say something such as: "Maybe it will work as they said, but we must be prudent and assume it will not and besides, *now* look at what the Russians are doing."

This approach to strategic thinking, known in the trade as "worst-case analysis," leads to a completely hopeless situation in which there is no possibility of achieving a state of affairs that both sides would consider as constituting parity. Unless the arms race is stopped by political action outside the two defense establishments, I feel reasonably sure there will be another "crash program" response analogous to what we had in the days of the "missile gap"—a situation some would like to see repeated.

I also mentioned in my own testimony at the ABM hearings that "we may further expect deployment of these ABM systems to lead to the persistent query 'But how do you know it *really* works?' and thus to increase the pressures against the current limited nuclear-test ban as well as to work against amplifying it." I mentioned this then, and I mention it again now, in the hope that it will become a self-defeating prediction. It is also important to note that the response of our own defense establishment to the Russian ABM deployment, which I have outlined above, was not the result of our being "provoked," and I emphasize this because we hear so much discussion about what is a "provocative" move and what is not. Rather, our response was motivated by a deep-seated belief that the only appropriate response to any new technical development on the other side is further technical complexity of our own. The arms race is not so much a series of political provocations followed by hot emotional reactions as it is a series of technical challenges followed by cool, calculated responses in the form of ever more costly, more complex and more fully automatic devices. I believe this endless, seemingly uncontrollable process was one of the principal factors President Eisenhower had in mind when he made his other (usually forgotten) warning: "We must be alert to the... danger that public policy could itself become the captive of a scientific-technological elite." He placed this other warning, also from his farewell address, on the same level as the much more familiar comment about the military-industrial complex.

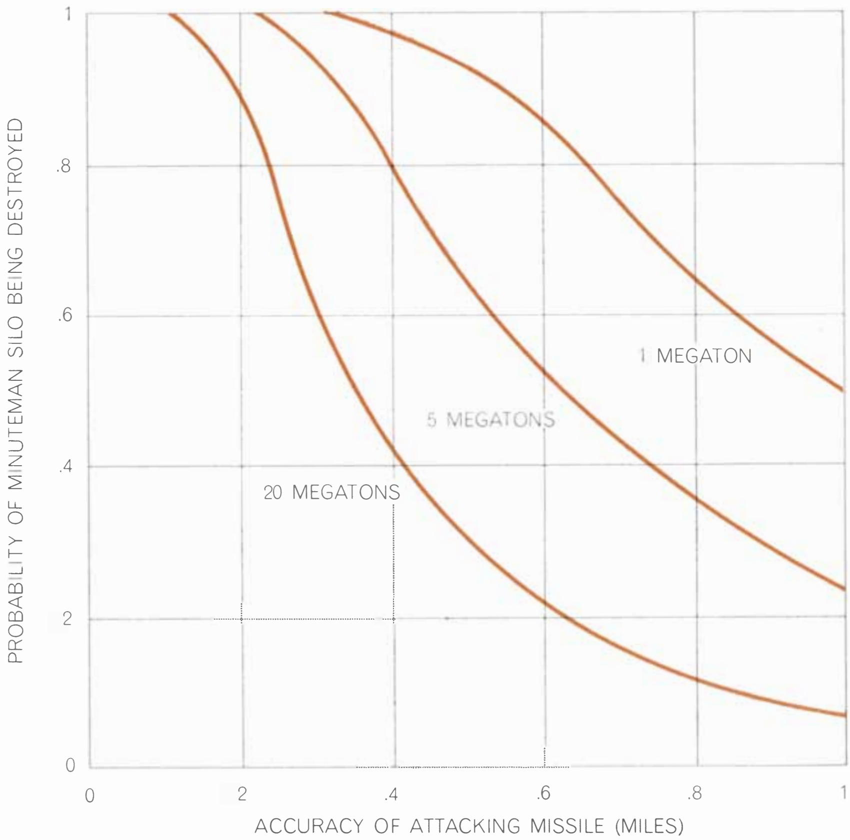
Several alternative approaches to Safeguard for protecting Minuteman have been discussed recently. These include superhardening, proliferation, a "shell game" in which there are more silos than missiles, and land-mobile missiles. Although I was personally hopeful before the hearings that at least one of these approaches would maintain its invulnerability, a review of the recent debates leaves me now with the pessimistic view that none of them holds much promise beyond the next 10 years.

Silo-hardening most probably does work now, in the sense that the combination of SS-11 accuracy and yield and Minuteman silo-hardening works out in such a way that one incoming warhead (and hence one SS-11 missile) has less than a 50-50 chance of destroying a Minuteman. If one considers the technological trends in hardening, yield per unit weight, MIRVing and accuracy, however, it does seem convincing that this is a game in which the offense eventually will win. Albert Wohlstetter, testifying in favor of the Safeguard system before the Senate Armed Services Committee, quoted a paper he wrote with Fred Hoffman in 1954 (long before any ICBM's were actually in place anywhere) predicting that the ability of silo-hardening to protect offensive missiles would run out by the end of the 1960's. That was a remarkably prescient study and is wrong only in numerical detail.

If we take the same rosy view of technology that was taken in almost all the pro-ABM arguments, then hardening will not work for more than another five years. My own view of the technological future is clearly much less rosy, but I do believe that the situation in which hardening is no longer the answer could come by, say, 1980 or, more appropriately, 1984.

Proliferation of Minuteman would have worked in the absence of MIRV. Now, however, it would seem that the ability to MIRV, which no doubt can eventually be carried much further than the fewfold MIRV we see for the immediate future, clearly makes prolifera-

FIRST SALVO LAUNCH of Minuteman ICBM's was made at Vandenberg Air Force Base in California on February 24, 1966. Photograph was released by U.S. Air Force.



VULNERABILITY OF MINUTEMAN is revealed in this graph, which relates probability of destruction of a hardened Minuteman silo to accuracy for three different sizes of attacking warhead. This graph was interpreted by Deputy Secretary Packard as demonstrating the seriousness of the threat to Minuteman posed by the large Russian SS-9 missile, which he said is capable of carrying either one 20-megaton warhead or three five-megaton warheads.

tion a losing game as well as the dangerous one it always was.

The "shell game" has not in my view been analyzed in satisfactory detail, but it would appear to have a serious destabilizing effect on the arms race. Schemes have been suggested for verifying that a certain fraction of the missile holes are in fact empty, but one can foresee a growing and persistent belief on each side that the "other missiles" must be hidden somewhere.

Road-mobile and rail-mobile versions of Minuteman have been seriously studied for well over a decade. These ideas have always foundered on two basic difficulties: (1) Such systems are inherently soft and hence can be attacked by large warheads without precise knowledge of where they are, and (2) railroads and highways all pass through population centers, and large political and social problems seem unavoidable.

Where does all this leave us insofar as finding a technical solution for protecting Minuteman is concerned? One and only one technically viable solution seems to have emerged for the long run: Launch on warning. Such an idea has been considered seriously by some politicians, some technical men and some military officers. Launch on warning could either be managed entirely by automatic devices, or the command and control system could be such as to require authorization to launch by some very high human authority.

In the case of the first alternative, people who think about such things envision a system consisting of probably two types of detection device that could, in principle, determine that a massive launch had been made and then somewhat later determine that such a launch consisted of multiple warheads aimed at our missile-silo fields. This information would be processed by a computer, which would then launch the Minutemen so that the incoming missiles would find only empty holes; consequently the Minutemen would be able to carry out their mission of revenge. Thus the steady advance of arms technology may not be leading us to the ultimate weapon but rather to the ultimate absurdity: a completely automatic system for deciding whether or not doomsday has arrived.

To me such an approach to the problem is politically and morally unacceptable, and if it really is the only approach, then clearly we have been considering the wrong problem. Instead of asking how Minuteman can be protected, we should be asking what the alternatives to Minuteman are. Evidently most other people also find such an idea unacceptable. As I mentioned above, the Army has found it necessary to reassure people repeatedly that ABM missiles would not be launched without approval by "the highest authorities," even though this is clearly a far less serious matter in the case of the ABM missiles than in the case of Minuteman.

The alternative is to require that a human decision-maker, at the level of "the highest authorities," be introduced into the decision-making loop. But is this really satisfactory? We would be asking that a human being make, in just a few minutes, a decision to utterly destroy another country. (After all, there would be no point in firing at *their* empty silos.) If, for any reason whatever, he was responding to a false alarm, or to some kind of smaller, perhaps "accidental," attack, he would be ensuring that a massive deliberate attack on us would take place moments later. Considering the shortness of the time, the complexity of the information and the awesomeness of the moment, the President would himself have to be properly preprogrammed in order to make such a decision.

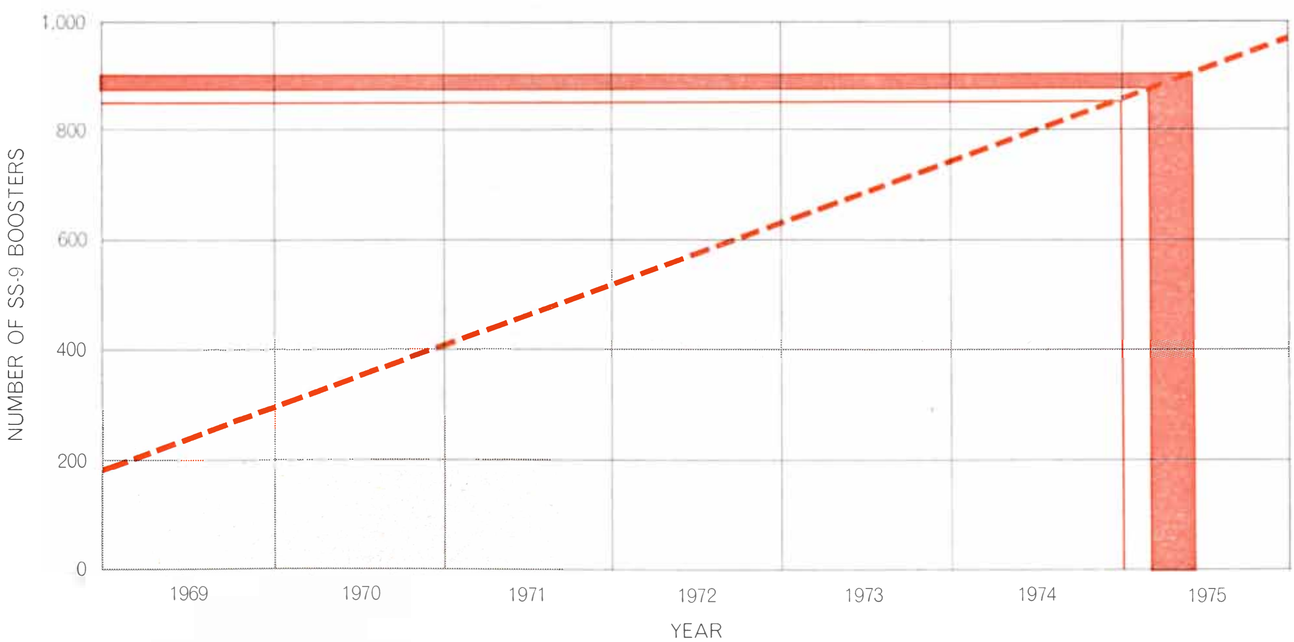
Those who argue that the Command and Control System is perfect or perfectable forget that human beings are not. If forced to choose, I would prefer a preprogrammed President to a computer when it came to deciding whether or not doomsday had arrived, but again I feel that this solution too is really unacceptable, and that once again, in attempting to defend Minuteman, we are simply dealing with the wrong problem. For the present it would seem the Polarises and the bombers are not, as systems, subject to the same objections, since there are now enough other approaches to the problem of ensuring their invulnerability to sudden massive destruction.

In my view, all the above once again confirms the utter futility of attempting to achieve national security through military technology alone. We must look elsewhere. Fortunately an opportunity does seem to be in the offing. There appears to be real promise that serious strategic-arms-limitation talks will begin soon. The time is propitious. There is in the land a fairly widespread doubt about the strictly military approach to security problems, and even military-minded politicians are genuinely interested in exploring other possibilities. The essay by Academician Sakharov, as well as the statements of Russian officials, indicate genuine interest on the other side. The time is propitious in another sense: both sides will be discussing the matter from a position of parity. Moreover, this parity seems reasonably stable and likely to endure for several years.

Later, however, major deployments of sophisticated ABM systems and, even more important, widespread conversion of present single-warhead systems to MIRV will be strongly destabilizing and will at least give the impression that parity is about to be upset. If so, the motto of the arms race, "Let us err on the side of military safety," will come to dominate the scene on both sides and the present opportunity will be lost. Therefore in the short run we must do everything possible to ensure that the talks not only start but also succeed. Although the ABM decision may not forestall the talks, it would seem that success will be more likely if we avoid starting things that history has shown are difficult to stop once they are started.

Such things surely include deployment of ABM missiles and MIRV's. There have been successes in stopping programs while they were in the development phase, but seldom has anything been stopped after deployment had started. The idea of a freeze on deployment of new weapons systems at this time and for these reasons is fairly widespread already, but achieving it will require concerted action by those believing strongly in the validity and necessity of arms limitations as a means of increasing national security. Thus the principal result of the recent national debate over the ABM issue has been to make it clear that Safeguard will safeguard nothing, and that the right step for the immediate future is doing whatever is necessary (such as freezing present deployments and developments) to ensure the success of the coming strategic-arms-limitation talks.

In addition, the ABM debate has served to highlight more serious issues (for example the implications of MIRV for the arms race) and to raise serious questions about other weapons systems. For instance, I suggest that we have also found that silo-based missiles will become obsolete. The only sure method for defense of Minuteman beyond, say, the mid-1970's seems to be the unacceptable launch on warning. As long as we must have a strategic deterrent, we must find one that does not force us to turn the final decision over to either a computer or a preprogrammed President. Minuteman was conceived in the 1950's and served its purpose as a deterrent through the 1960's, but it appears that in the 1970's its threat to us will exceed its value, and that it and other silo-based missiles will have to go. The deterrent must have alternatives other than "go/no-go," and for the 1970's at least it would now appear that other strategic weapons (Polaris/Poseidon and bombers) could provide them. I expect, however, that as the continuing national debate subjects the whole matter of strategic arms to further public scrutiny we shall learn that these other alternatives also have dangerous flaws, and we shall see confirmed the idea that there is no technical solution to the dilemma of the steady decrease in our national security that has for more than 20 years accompanied the steady increase in our military power.



SAFEGUARD COULD BE NULLIFIED within a few months after its Phase I deployment, according to this graph, which is based on calculations presented by George W. Rathjens in his testimony before the Senate Armed Services Committee on April 23. His analysis took as a basis of calculation the implication in Secretary Laird's testimony that the Minuteman force may become seriously imperiled in the mid-1970's. Assuming that the Russian SS-9's would have four or five MIRV warheads each by that time, Rathjens then estimated that approximately 850 SS-9's would have to be deployed in order to achieve this result. From this number, and the estimate of the current number of SS-9's deployed (about 200), he got a rate of deployment (about 100 per year). Making certain assumptions about the numbers and effectiveness of the Spartan and Sprint ABM missiles that would be deployed at that time, Rathjens found that by prolonging the SS-9 production program by two to five months the Russians would be able to cope with Safeguard by simply exhausting it and would still have enough warheads to imperil Minuteman. Recently different numerical assumptions have been made, but they do not change the general conclusion that the proposed Safeguard system is much too thin to safeguard Minuteman.

"Genetic Drift" in an Italian Population

Studies of blood-group frequencies and consanguineous marriages among the people of the Parma Valley indicate that this random change in hereditary type distinctly influences human evolution

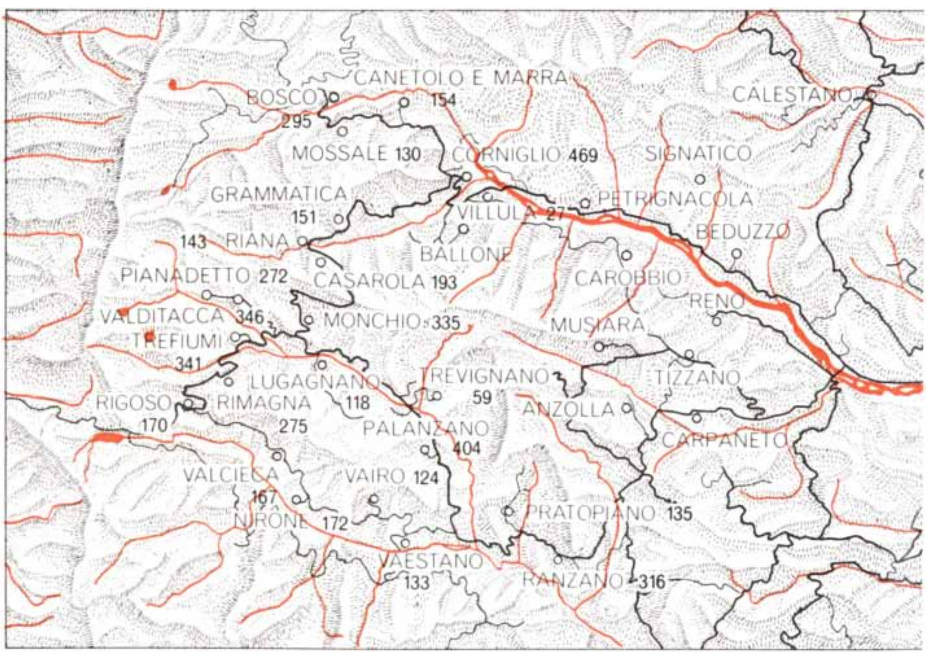
by Luigi Luca Cavalli-Sforza

The variety of hereditary types in a human population originates with mutations in the genetic material. The survival and preferential multiplication of types better adapted to the environment (natural selection) is the basis of evolution. Into this process, however, enters another kind of variation that is so completely independent of natural selection that it can even promote the predominance of genes that oppose adaptation rather than favoring it. Called genetic drift, this type of variation is a random, statistical fluctuation in the frequency of a gene as it appears in a population from one generation to the next. Sometimes genetic drift seems to exert only a moderate influence, causing the frequency of a gene to fluctuate by 5 or 10 percent. At other times it may result in one gene overwhelming other genes responsible for the same characteristic.

How strong is the influence of genetic drift in evolution, and what factors control it? Together with my colleagues Franco Conterio and Antonio Moroni of the University of Parma, Italo Barrai and Gianna Zei of the University of Pavia and our collaborators at other institutions, I have for the past 15 years been investigating genetic drift in the populations of the cities and villages in the Parma Valley in Italy. We have examined parish books, studied marriage records in the Vatican archives, made surveys of blood types, developed mathematical theories and finally simulated some of the region's populations on a computer. We have found that genetic drift can affect evolution significantly, and we have been successful in identifying factors that control it.

Hypothetically genetic drift can happen in the following way. Suppose a small group of Europeans, perhaps 10 people, colonized an island (as the mutinous sailors of the *Bounty* and their Tahitian women did). Among 10 such randomly chosen people there might well be no one with blood of Type B or Type AB, because the genes for these blood types are respectively carried by only 15 percent and 5 percent of Europeans. Forty percent of Europeans have blood of Type O, and the same percentage have blood of Type A. In this small group, then, the frequency of the Type B gene might be zero rather than 15 percent because, in the nature of statistical processes, it is absent and cannot reappear unless a rare mutation takes place. The gene may also be extinguished if only one or a few members of the group carry the Type B gene and they produce no descendants. Conversely, the frequency of a rare gene may sometimes increase until it becomes "fixed," or predominant. In remote valleys of the Alps, for instance, there is a relatively high frequency of such traits as albinism, mental deficiency and deaf-mutism, which are normally subject to negative selection.

This view of genetic drift suggests that two factors determine its strength: population size and migration. In the population of an alpine village, or among our hypothetical island colonists, a gene might vanish or become fixed in relatively few generations because the popula-



THE PARMA VALLEY stretches from the ridges behind the village of Rigoso at left to the plains of the Po River, 90 kilometers to the north. Because the settlement patterns of the Parma Valley include isolated villages in its steep-sided upper reaches, hill towns at lower tion is so small that even a slight change in the actual number of people carrying a gene causes a large change in the percentage of the population endowed with that trait. In a larger population a change in gene frequency would affect a smaller percentage of the people, and thus drift would be less pronounced. Migration can offset the movement of a gene toward predominance or extinction by increasing the frequency of rival genes. Like migration, natural selection may also restore equilibrium, by promoting adaptive combinations of genes, even after the situation has been profoundly disturbed by genetic drift.

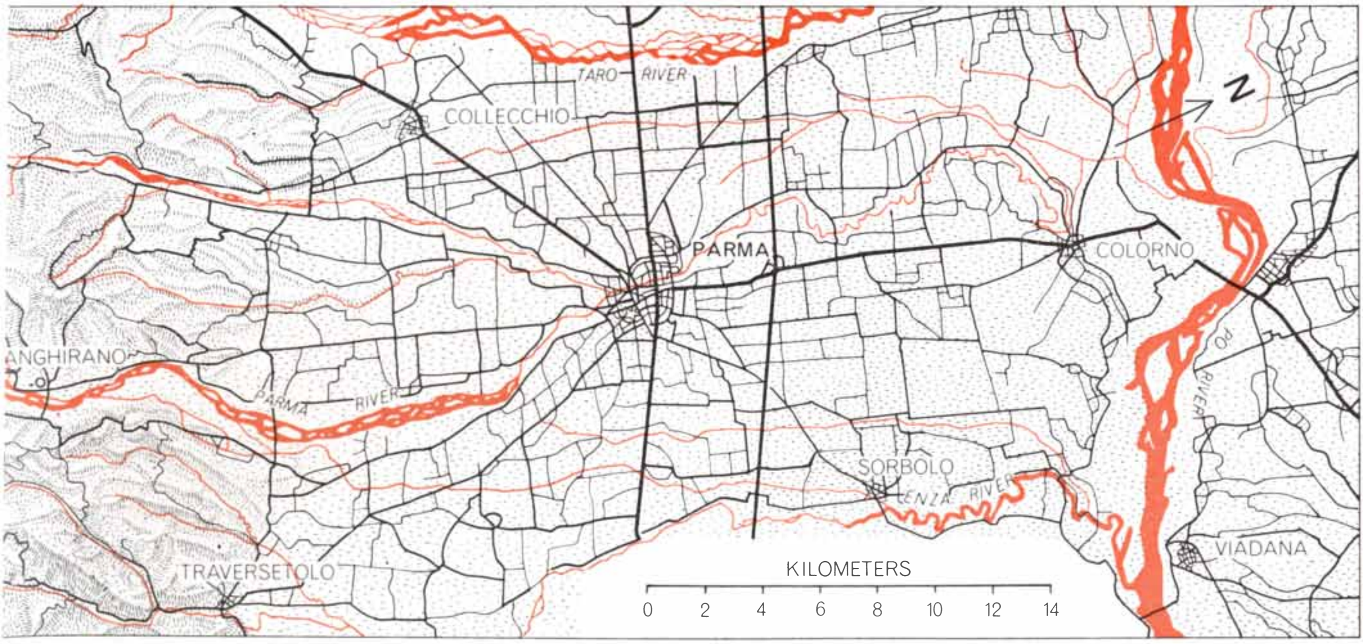
Isolated observations, however, offer only glimpses of the significance of genetic drift, and speculating on such observations cannot provide us with a basis for measuring the phenomenon or identifying the factors that control it. Accordingly my colleagues and I began to search for an experimental group in which we could conduct tests that would clarify these issues. We decided that the Parma Valley, which stretches for 90 kilometers to the south of the city of Parma, would provide an ideal population.

The Parma Valley, located in north-central Italy, is named after a stream that flows through it into the Po from the Apennines. The river has carved out what in its upper reaches is a steep-sided, inhospitable valley that gradually opens out into gentle hills and finally into a broad plain on which lies the city of Parma. The very geology of the valley creates an almost complete spectrum of the patterns of human habitation. In the highlands the steep countryside encourages people to gather in small villages of about 200 to 300 inhabitants. Farther downstream the rural villages become bigger as the hills give way to the plain, and where the stream flows into the Po stands the city of Parma [*see illustration below*].

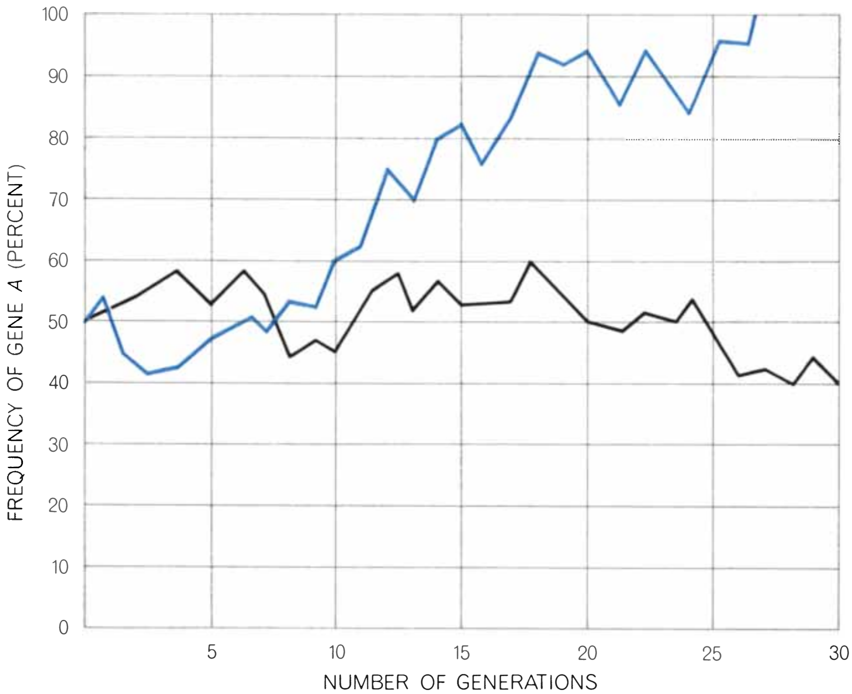
People have lived in the Parma Valley since prehistoric times. Because there have been no major immigrations since the seventh century B.C. a certain demographic and genetic equilibrium has been reached. The effect of natural factors such as migration, natural selection and genetic drift can therefore be studied under the simplest conditions: when they are, or can be reasonably believed to be, in equilibrium. On the plain and in the hills, however, immigration and migratory exchanges are more frequent than in the mountains. Important demographic information is supplied by the parish books of marriages, births and deaths in the area since the end of the 16th century. Accordingly the valley offers excellent opportunities for measuring the effects of population size and migration on drift.

One of the first hypotheses we decided to test was the one that drift should be more pronounced in a small, isolated population than in a large one, since a large population has a wide variety of gene types and is more susceptible to migration. If this proposition were true, we could expect to detect the strongest drift in small, isolated mountain villages in the uplands of the Parma Valley, less drift in the hill communities farther down the valley and the least drift on the plain.

The most convenient way to measure drift is suggested by the nature of the phenomenon itself. Under the influence of drift village populations will tend to become more and more different, even if at the beginning they were homogeneous in their composition of hereditary types. Taking a particular hereditary characteristic, say blood of Type A, individuals with blood of this type may become more frequent in one village and rare in other villages. What is needed is a measure of this kind of variation among villages. If we had only two villages, we could take the percentage of individuals with blood of Type A in one village and the percentage in the other village and compute the difference between the two percentages as a measure of the variation between the two villages. In examining many villages we might consider the differences between all possible pairs of villages and average them out. In actuality we use a somewhat different way



altitudes and the city of Parma on the plain south of the Po, the valley constitutes an excellent natural laboratory for studying how genetic drift affects human evolution. In order to study genetic drift, blood samples were taken throughout the cities and towns of the valley, and those upland villages whose population sizes are indicated by numbers were simulated in a computer experiment.



GENETIC DRIFT can cause the frequency of a gene to vary markedly from one population to another. In this calculation performed with random numbers Gene *A* appears in 50 percent of the members of two populations. Only 27 generations later Gene *A* has become "fixed" in one (*color*) and in the other its frequency fluctuates from 40 to 60 percent.

of measuring the variation, but the principle is much the same. Here let us simply call the result "measure of genetic variation between villages." Estimates of the variation were obtained after we had grouped villages in somewhat larger local areas, from the highlands down through the hill towns to the city of Parma. The measure we used excludes the effects due to sampling because we used a fraction of the total population.

As predicted, the variation between villages declined as population size increased, from .03 in the high valley, where the population density was well under 50 people per square kilometer, to less than .01 in the hill country, where there are about 100 people per square kilometer, to almost nothing on the plain, where the density reaches 200 people per square kilometer [*see bottom illustration on page 34*].

It is possible that the variation between villages could be caused by adaptation to different environments rather than by genetic drift. As unlikely as it may seem, it is possible that the environmental conditions differ from village to village so that different genes are favored in each place. It may also be that people of diverse origins and therefore of diverse blood groups have settled in the more populous regions, and that because of these historical accidents there has not yet been time to reach an equilibrium.

If natural selection or historical accidents were responsible, the percentage of individuals possessing a certain gene would vary from village to village. The percentage would not necessarily vary in the same way for all genes, since there is no reason why the selective factors or historical accidents should operate with equal force on all genes. It would, in fact, be a strange coincidence if they did. If the variations in genes were caused by genetic drift, however, they would be the same, on the average, for any gene. The reason is that genetic drift, being a property of the population rather than of the gene, should affect all genes in the same way. Our evidence shows that the variations between villages are indeed the same for any gene.

This first test of our ideas about genetic drift was convincing, but in order to test our analysis more severely we wanted to make exact forecasts of the amount of genetic variation caused by drift. Such an exercise would require a precise quantitative prediction rather than a simple qualitative statement. Unfortunately the classical mathematical theories of population genetics (put forward by Sewall Wright, Motoo Kimura and Gustave Malécot) require that villages be of equal size, and that migrations between them follow a highly homogeneous pattern, simplifications that are rather far from reality.

To avoid this difficulty we developed other methods of predicting variations in the frequency of a gene on the basis of population size and migration. With the help of Walter Bodmer of Stanford University it was possible to devise a new theory that takes account of the actual observed migration pattern from village to village, however complex it may be. The model removes many of the oversimplifications of the classical theories but not all of them. We have therefore also developed a more general method that makes it possible on the basis of simple demographic information to predict the expected amount of drift with unlimited adherence to reality. This method consists in the use of artificial populations generated in a computer. Before I describe it, however, I should mention an apparently independent but in fact closely related approach, using a substantially different body of data, that we have followed in parallel with the study of genes.

This alternative approach we have followed is the study of relationships between individuals, which can be obtained from pedigrees or similar sources. One intuitively understands that the relationship between people must be associated with the similarities (or the differences) between the genes they carry. Both depend on common ancestry. Greater isolation between villages implies a lesser degree of common ancestry between the people of the villages, and therefore both a lesser degree of relationship between them and more differences between their genes. Thus the study of pedigrees, making it possible to estimate common ancestry, or degrees of relationship, should yield almost the same information as an analysis of the frequency of genes in the various villages. It has been shown that even data as simple as the identity of surnames can, in indicating common ancestry, supply information similar to what can be obtained from the direct study of genes.

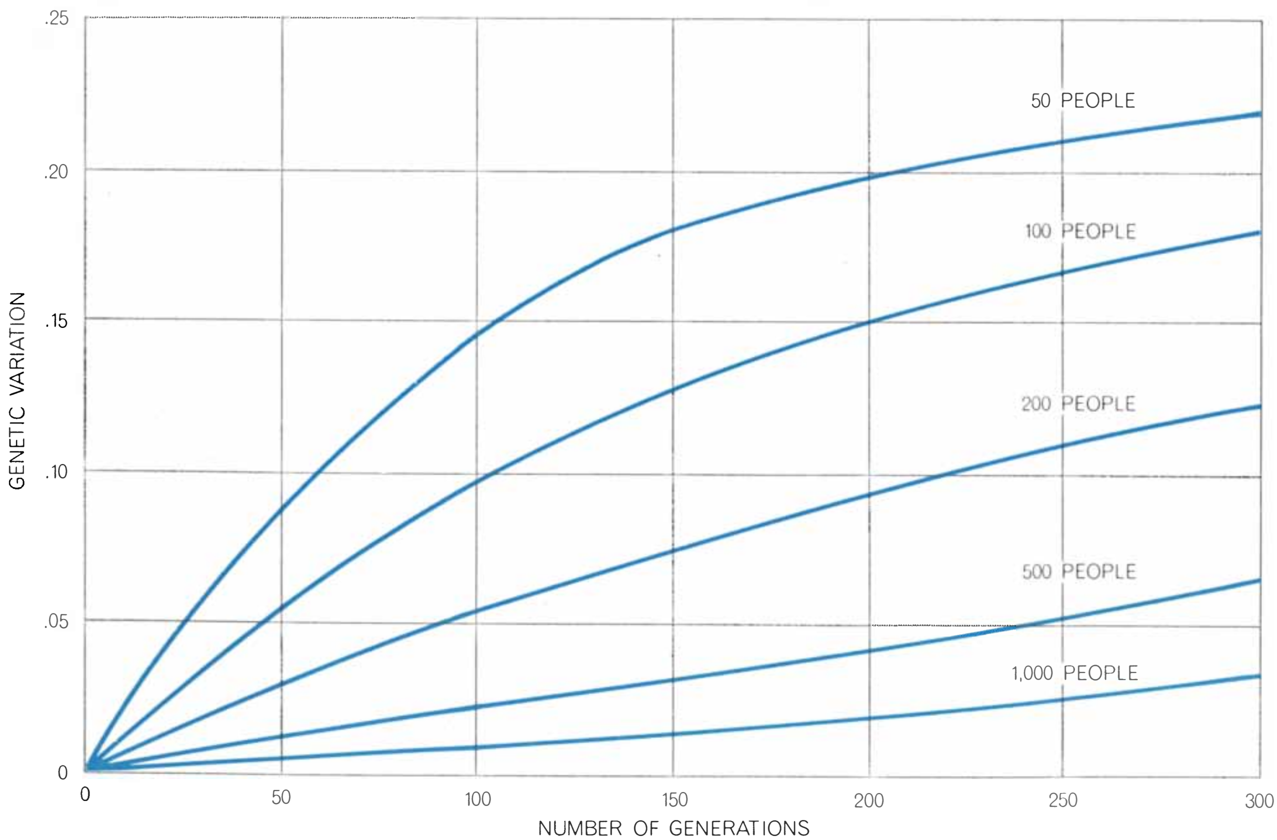
It is the availability of parish books in the Parma Valley that makes it possible to carry on this investigation in parallel with the study of genes. Unfortunately the reconstruction of pedigrees from parish books is a laborious task, and it has not yet been completed. We do, however, have data on relationships from another source: records of consanguineous marriages. We found it particularly interesting to test the validity of this method as an alternative to the direct study of the effects of drift on genes. In the Parma Valley we could compare all these approaches. We could see, for instance, if we could predict drift from consanguinity or vice versa, or better still, predict both from a common source: simple demographic data.

Consanguineous marriages and genetic drift are similarly affected by common factors: population size and migration. A small population encourages consanguineous marriage because after a few generations most marriage partners would also be relatives. Migration, on the other hand, tends to decrease the frequency of consanguineous marriage by introducing new partners who are not relatives, or, if the flow is outward, by removing relatives who would otherwise be available.

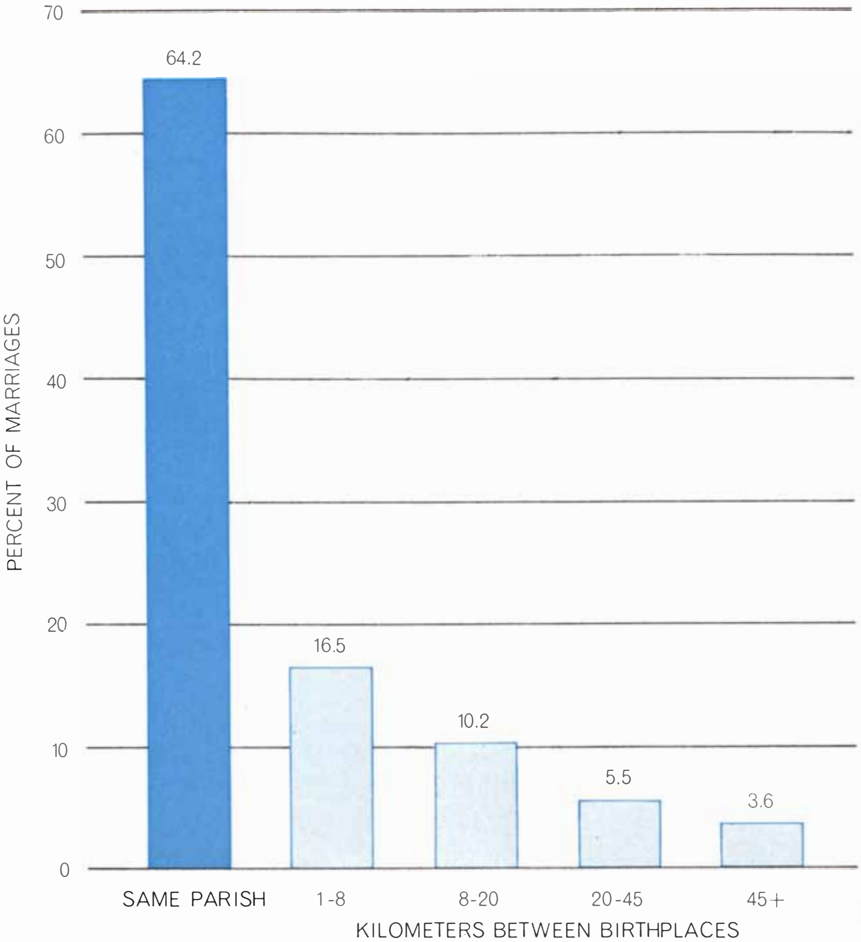
The mathematical model with which the frequency of consanguineous marriage can be predicted is relatively simple. It is based on the idea that the population whose size critically affects the frequency of consanguineous marriage and genetic drift is somewhat diffuse, geographically speaking. As the Swedish geneticist Gunnar Dahlberg pointed out in 1938, this population basically consists of a group of people who are potential marriage partners for one another. This population is therefore not identical with the marriageable population because there are social barriers that reduce marriage choice. A village or a town might also be so large that not all the available partners would know one another. Such factors tend to make the population of eligible partners smaller than it is in a smaller village. The group can, however, extend across political boundaries, so that marriages are made between people living in different villages.

Since a simple census will not yield the size of the population of marriageable individuals, the population must be determined mathematically. By definition the population consists of a circle of N people available to one another for marriage. Assume now that an individual is not prohibited from marrying a blood relative (provided they are not so closely related that the marriage is forbidden by law). In this case the probability that he (or she) will marry a relative will be equal to the ratio between the number of eligible blood relatives, c, and the number of candidates who are not relatives. The probability of consanguineous marriage, m, will therefore equal c/N. This probability is also identical with the overall frequency of consanguineous marriage; all other factors being excluded, the frequency of consanguineous marriage would depend only on the number of available partners who are also kinsmen. Thus if there were 40 available partners and 20 of them were blood relatives, the frequency of consanguineous marriage would be one in two, or 50 percent. Knowing the number of blood relatives from simple calculations, and the frequency m of consanguineous marriages from ecclesiastical records, we can determine the size N of the population of eligible mates because it is a function of m and c. In other words, if $m = c/N$, then $N = c/m$.

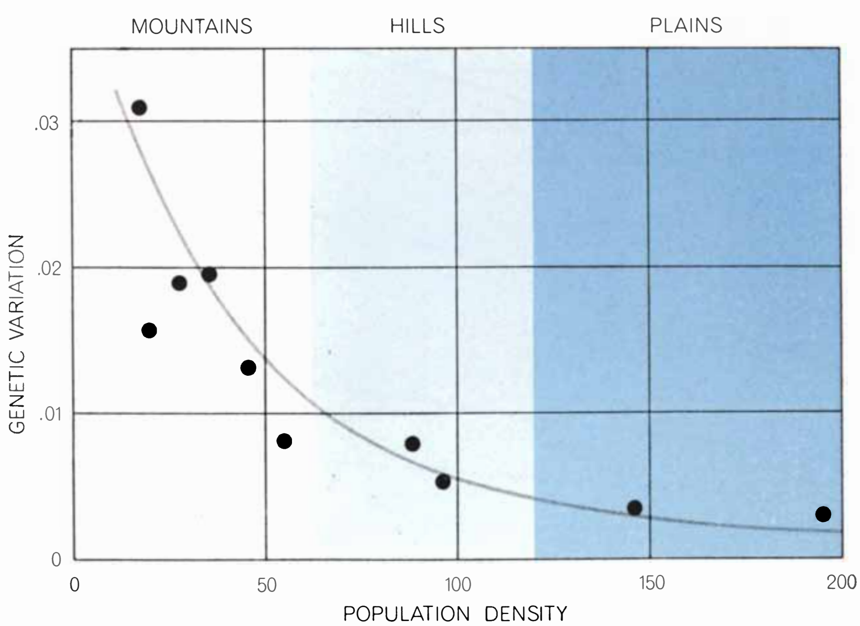
Having determined the population size, it would be convenient at this point if we could simply complete our



POPULATION AND DRIFT are closely related. The frequency of a particular gene in each population begins in the first generation at 50 percent, equivalent to zero on the vertical scale marked according to a measure of genetic drift called variance. After 300 generations the variation of the gene in the smallest population has increased to .22, almost as far as it can go, whereas in the largest population it has reached only .03. Genetic drift, then, is strongest in smaller populations and weakens as the population size increases.



MIGRATION in upper Parma Valley has been infrequent, a conclusion drawn from the fact that most marriages recorded from 1650 to 1950 in parish books unite men and women who are from the same village. The number falls as the distance separating birthplaces increases.

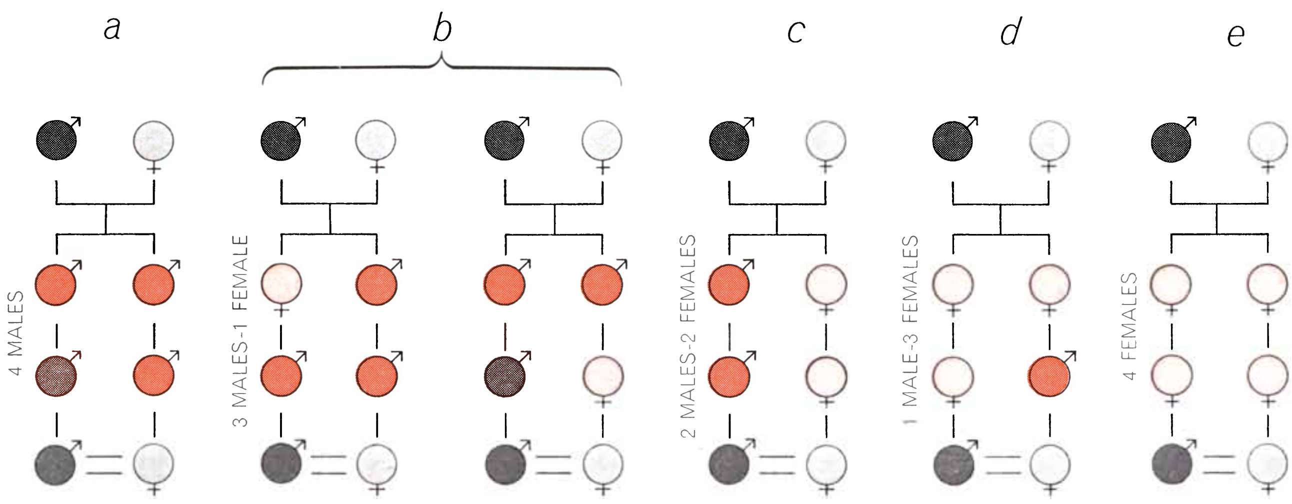


VARIATION in the frequency of a blood type between one village and another was greatest, as predicted, in the isolated upland hamlets, and declined as population density increased farther down the valley in the hill towns, on the plain and in the city of Parma.

model by taking into account the effect of migration on the available supply of marriage partners. We could then predict the frequency of consanguineous marriage and therefore estimate the amount of genetic drift. Reality, however, forces us to make a circuitous detour. It appears that there are certain factors (such as the tendency of people to marry people of a similar age, biases for or against certain kinds of consanguineous marriage and the fact that degrees of kinship too remote are recorded incompletely or not at all) that would distort the prediction of genetic drift because they affect only the frequency of consanguineous marriage without influencing the rate of variation for a gene. In order to calculate genetic drift on the basis of consanguineous marriage we must identify and compensate for such factors.

These factors can be inferred from the study of consanguineous marriages. The Vatican archives contain records of 590,000 dispensations for consanguineous marriages granted from 1911 to 1964, the year in which our gathering of data temporarily ended. Only the dispensations for the more distantly consanguineous marriages that were granted directly by the bishops in certain areas such as Sicily and Sardinia and a few other remote dioceses are excluded. This material provided our investigators (led by Moroni, who had been a student of mine at the University of Parma) with a mass of valuable information. The records provide, among other data, the given name and the family name of the parents of the couple and the degree to which the marriage is consanguineous.

These records exist because the Roman custom and religious belief that prohibited marriage between blood relatives was inherited and diffused by the Catholic church. During the Middle Ages only the Pope could grant dispensations from the prohibition, and the dispensations (at least those known to us) were not numerous. The degree of consanguinity eligible for dispensation, however, has varied through the centuries, and a progressively more liberal trend can be detected. In the 16th century the Council of Trent recommended that a special dispensation be required for marriages up to "the fourth degree" (third cousins). Since 1917 dispensation has been required only for marriages between second cousins. The Vatican Council has recently pushed the liberalization one degree further, so that today only a marriage between first cousins, or between uncle and niece, require dispen-



FAMILY TREES affect the frequency of consanguineous marriage. Different family trees associated with second-cousin marriages show that when spouses share only male ancestors (*a*), the number of consanguineous marriages reaches 774. As the number of female ancestors rises the number of marriages falls. Trees such as *b* (two varieties are shown) have produced 652 marriages, *c* 325 marriages, *d* 262 marriages and *e* 252 marriages, according to diocesan records from 1850 to 1950. Since land passes from father to son, men do not usually emigrate and marriages among relatives are therefore much more likely. There are 10 other trees for second-cousin marriages.

sation. Marriages between closer relatives are not and never have been eligible for dispensation.

The dispensation must be requested by the parish priest from the Curia before the marriage can be celebrated. In some cases the bishop can grant dispensations, but customarily he must forward a copy of the request to Rome (or, in other countries, to the representative of the Pope in that country). The Vatican will then reply to the bishop, and he will reply to the priest. The Vatican practically always grants dispensations in allowed cases, and therefore the dispensation request has constituted, or at least constitutes today, only a formal obstacle.

Of the factors that seem to alter the frequency of consanguineous marriage, one is the very closeness of the blood relationship. Apart from legal and religious restrictions, there is the widespread knowledge that consanguineous marriages may result in hereditary handicaps for the offspring.

Age can also have an important effect. As the archives show, marriages in which the consanguineous mates are a generation apart, such as those between uncle and niece and between first cousins once removed, are rarer than those between first cousins and between second cousins. We can explain this fact if we assume that in both consanguineous marriages and nonconsanguineous marriages age affects the choice of mates in the same way. In Italy, for instance, there is a mean age difference between husband and wife of about five years; the difference is smaller for young spouses and larger and more variable for older spouses. Therefore by considering the age differences among children of the same family, between parents and children and between normal spouses we can predict that marriages between uncle and niece will be only 3 percent of what would be expected on the basis of the frequency of this relationship, and marriages between aunt and nephew will be still rarer. By the same token one could expect a higher frequency of marriage between first cousins, because they tend to be of a similar age.

Migration also tends to reduce the frequency of consanguineous marriage. In places where the population is small and migration is low, blood relatives remain in contact. Hence we are not surprised to find most of the consanguineous marriages in rural areas whose populations have been rooted in the same soil for many generations. In industrial areas migration tends to disperse blood relatives so that they may not even meet, much less marry. Therefore consanguineous marriage tends to be diluted in frequency, or even to vanish, just as genetic drift does.

The effects of migration become more complex when we study specific types of genealogical trees. From data gathered in northern Emilia, a broad region including the Parma Valley, it appears that in the past century the number of consanguineous marriages diminishes when among the immediate ancestors of husband and wife there are more females than males. The reason is that men and women migrate in different patterns. In a largely rural area such as the one we were examining, where land is inherited by the male child, fewer sons emigrate than daughters. Moreover, a woman who marries someone from another village will emigrate in the process, and the distance between the villages will make her descendants a little less available for marriage with descendants remaining in the original place. The more women there are in the genealogical tree, the more significant this pattern is [*see illustration above*].

We have also isolated a factor of a sociological nature. It is scarcely important enough on the statistical level to merit consideration, but I shall cite it as a curiosity. It is the tendency of children of consanguineous mothers to intermarry. Probably this is because the mothers, when related, have a greater tendency to conserve bonds that favor marriage among the children.

With the help of Kimura (who is with the National Institute of Genetics in Japan) and Barrai, we incorporated these inhibiting factors and the effects of migration into a mathematical model that predicts the incidence of consanguineous marriage with some accuracy.

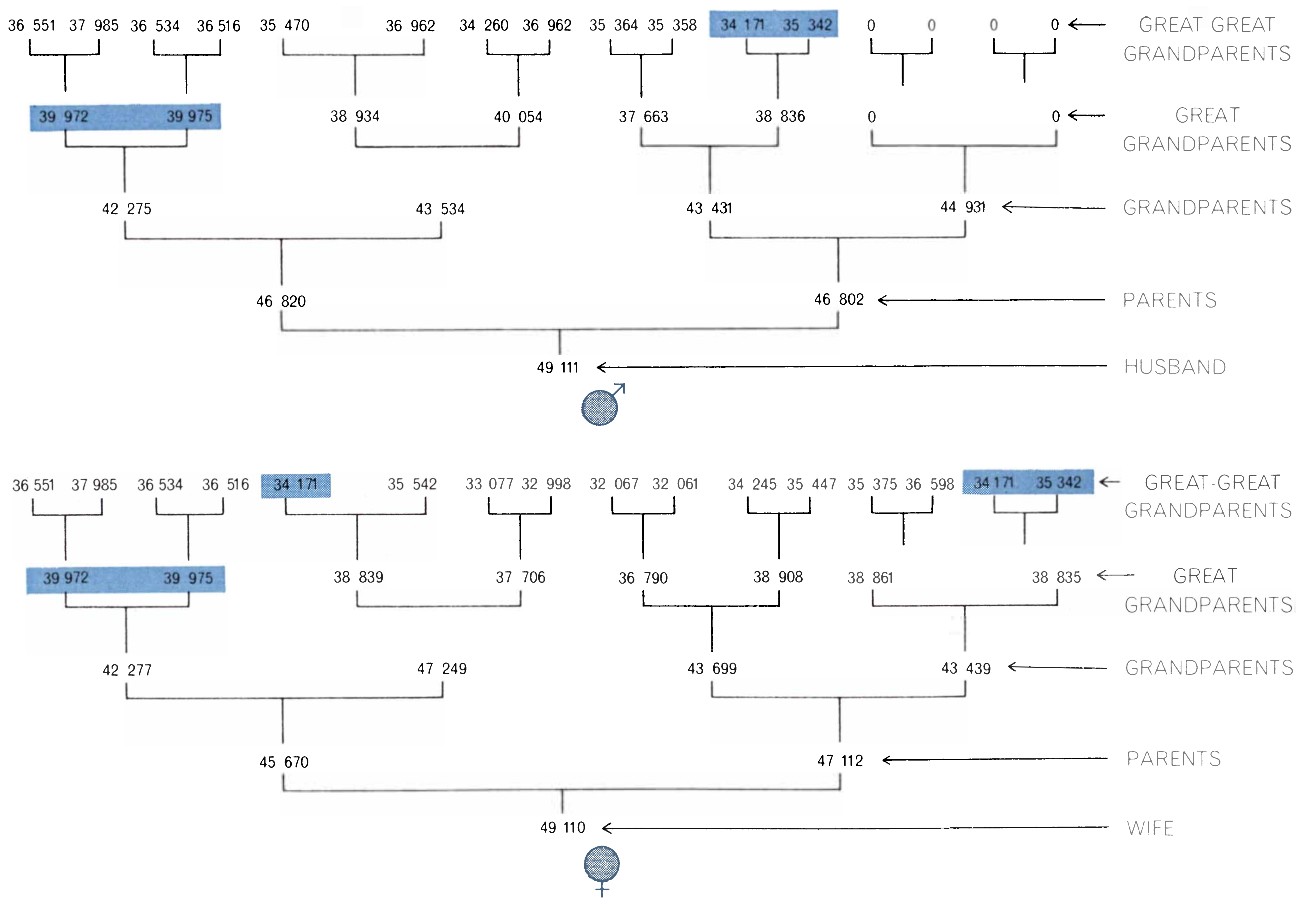
Even this carefully derived mathematical theory suffers, as all applications of mathematics in biology do, from the necessity of simplifying reality in order to make results calculable. The difficulty lies not only in solving complicated mathematical problems but also in successfully simplifying the terms of the problem to allow the use of appropriate mathematical instruments without losing essential features of the problem. Computers, however, make possible another technique for attacking these problems. If we have enough data available on the population under examination, we can reconstruct it in the computer and see what happens in experiments. The repetition of these experiments a sufficient number of times gives us a view of what we can expect in reality. In this way we can, without the use of higher mathematics and with the employment of real data, forecast the complicated effects of the genetic structure of a population on phenomena such as the frequency of consanguineous marriages or genetic drift.

Naturally we simplify the artificial population that we reconstruct in a computer as much as we can. Our computer "men" and "women" do not have hands and brains, only a number (0 or 1) that indicates sex, a number of several digits representing a name, and other numbers that identify the father and the mother of each individual, his distant ancestors and his descendants. If we study the effects of age, we must give our artificial subjects an age, and by the same token we can characterize their geographic location and social class [*see illustration below*].

An artificial population so constructed marries, reproduces and dies according to certain probability tables drawn from reality. For reasons of economy we make the time advance crudely, in steps of 10 years. Thus if we have a 30-year-old person, we ask the probability of his dying before 40, and on the basis of tables of the real population we make him die according to a random procedure that has a probability equal to the real one. To determine whether a 30-year-old man dies before 40, for example, we calculate from the program a random number between 00 and 99. Such a number has equal probability of being one of the 100 numbers from 00 to 99. Since the probability of a man of this age dying is 12 percent, the individual dies if the chosen number lies between 00 and 11 and survives if it falls between 12 and 99. In the same way, using a random-number table that gives real probabilities, we decide if he marries and whom he marries by making the choice based on age, social class and geographical location. Finally, computer-generated marriages can be analyzed to determine the degree of consanguinity.

The results obtained by simulating the population of the Parma Valley confirm the impressions we had gained from our mathematical analysis of the actual population. That is, taking into account age, migration and the number of blood relatives of a given degree, the frequency of consanguineous marriages is about the same as what one would expect it to be if such marriages happened randomly. This is particularly true for the more remote degrees of consanguinity. Among first cousins, however, it seems that the frequency of consanguineous marriages is only half what it would have been if such marriages had been random.

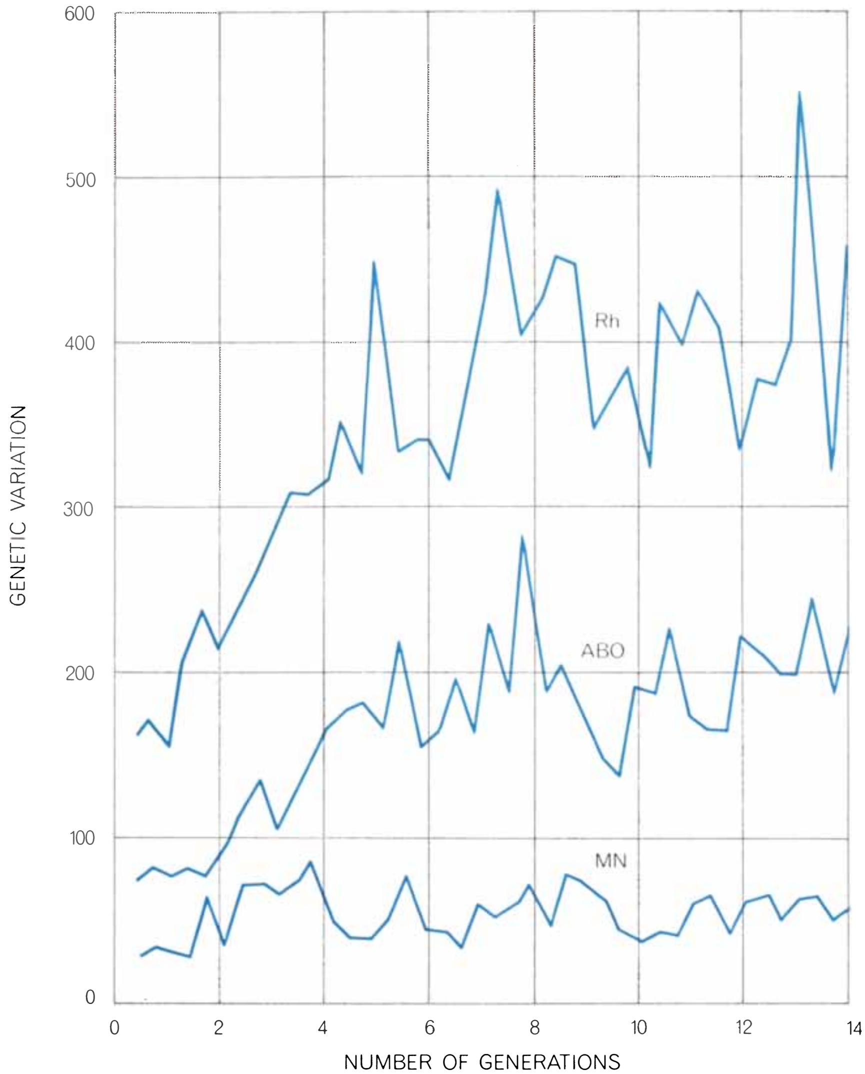
Our final experiment consisted in sim-



MAN AND WIFE are programmed for a computer run that simulates populations in order to determine the rate of genetic drift through the frequency of consanguineous marriage. The male's name is encoded in the digits 49 111. Other numbers (not shown) can be programmed that indicate the person's sex, age, social class and genetic endowment. Shading indicates ancestors shared by man and wife.

ulating the populations of 22 villages in order to test our hypotheses of genetic drift. The 5,000 individuals in the test population were given genes of the three blood-group systems: ABO (governed by the gene types A_1, A_2, B and O), Rh (seven gene types) and MN (two gene types). At the beginning of the simulation the frequencies of these genes in each simulated village were the same as the average of the frequencies in the actual villages. With the passage of each simulated generation, however, the frequencies began to change by a factor that was approximately the same for each blood-group system. They finally leveled off when the number of people belonging to a blood group was two or three times the original value. This equilibrium was reached fairly soon (after about 15 generations) as a consequence of the establishment of a balance between drift, which tends to make villages different, and migration, which tends to make them more nearly the same.

We found that the variations among the simulated villages quite closely matched the variations predicted by Bodmer's theory, thus confirming its capacity to represent real data. We also found that the variations among the simulated villages matched, although less closely, the variations observed among the real villages. In both the simulated and the real villages no single gene type either vanished or became predominant; the drift was not strong enough to achieve this result. Fairly divergent proportions of genes could be found in the different villages. Which gene increased in which particular village was of course a matter of chance. We did not expect a real village to show the same proportion of a given gene as its artificial counterpart [*see illustration on this page*].



GENE FREQUENCIES for the ABO, Rh and MN blood-group systems in the populations of the villages from the upper part of the Parma Valley vary with the passing of each generation. Differences in gene frequency between villages simulated in a computer program and measured by "chi square" were similar to the differences in the real villages. Therefore migration, population size and other demographic influences that control the simulated genetic drift are probably equivalent to the forces at work in the real villages.

It is clear that since the observed variation corresponds—within limits that we are now investigating—to the expected one it is not necessary to invoke explanations other than the action of genetic drift. The methods we have used in our study of the Parma Valley have now been applied to other populations as diverse as African pygmies, New Guinea tribesmen and the descendants of the Maya Indians. The results so far have confirmed the concept that genetic drift is the principal agent responsible for the variations among villages, tribes or clans. In fact, at the microgeographical and microevolutionary levels on which we worked the differences attributable to natural selection were not large (apart from some minor ones we intend to reexamine). Furthermore, there is evidence that drift can operate on a macroevolutionary scale extending over millions of years. Recent comparisons of the sequences of amino acids in the proteins of separate species show differences that are at least in part caused by drift, although the evidence is still controversial.

In seeking to extend these conclusions one could study populations distributed over large areas. In that case, however, it would be easier to encounter disturbing factors. For example, in a large enough region one might encounter selective pressures whose diversifying effects would be added to the effects of drift on certain genes, whereas in another area or in the case of another gene natural selection might oppose drift and reduce the variations.

In any case, the discovery that genetic drift can affect evolution on a small scale over a short period of time gives the phenomenon a more important role in evolution than was once thought. It would be an error to assume, however, that evolution is almost entirely random. Only natural selection can bring about adaptation to the environment, and its importance must not be underestimated. The relative importance of drift and natural selection in determining the course of evolution remains to be assessed.

METALLIDING

With an electrolytic bath consisting of a hot molten fluoride, atoms of one metal can be made to diffuse deep into the surface of another. The result is a wide range of materials with new surface properties

by Newell C. Cook

The association of one metal with another often results in properties that are superior to those of either metal alone. Such associations have traditionally been accomplished by two processes: alloying, which mixes the metals in the molten state, and plating, which attaches one of the metals to the surface of the other. My colleagues and I have now developed a third process called metalliding. It involves diffusing the atoms of one metal into the surface of another. The diffused metal becomes an integral part of the surface of the other metal, instead of being only mechanically attached to the surface as in plating. The new process is in fact a form of alloying, except that the alloy is only at the surface.

The diffusion is achieved by means of a high-temperature electrolytic process. The diffusing metal, serving as an anode, and the receptor metal, serving as a cathode, are suspended in a bath of molten fluoride salt. When a direct current passes from the anode to the cathode, the anode material dissolves and is transported to the cathode. There the anode material diffuses into the cathode, giving rise to an alloyed surface [*see illustration on page 44*].

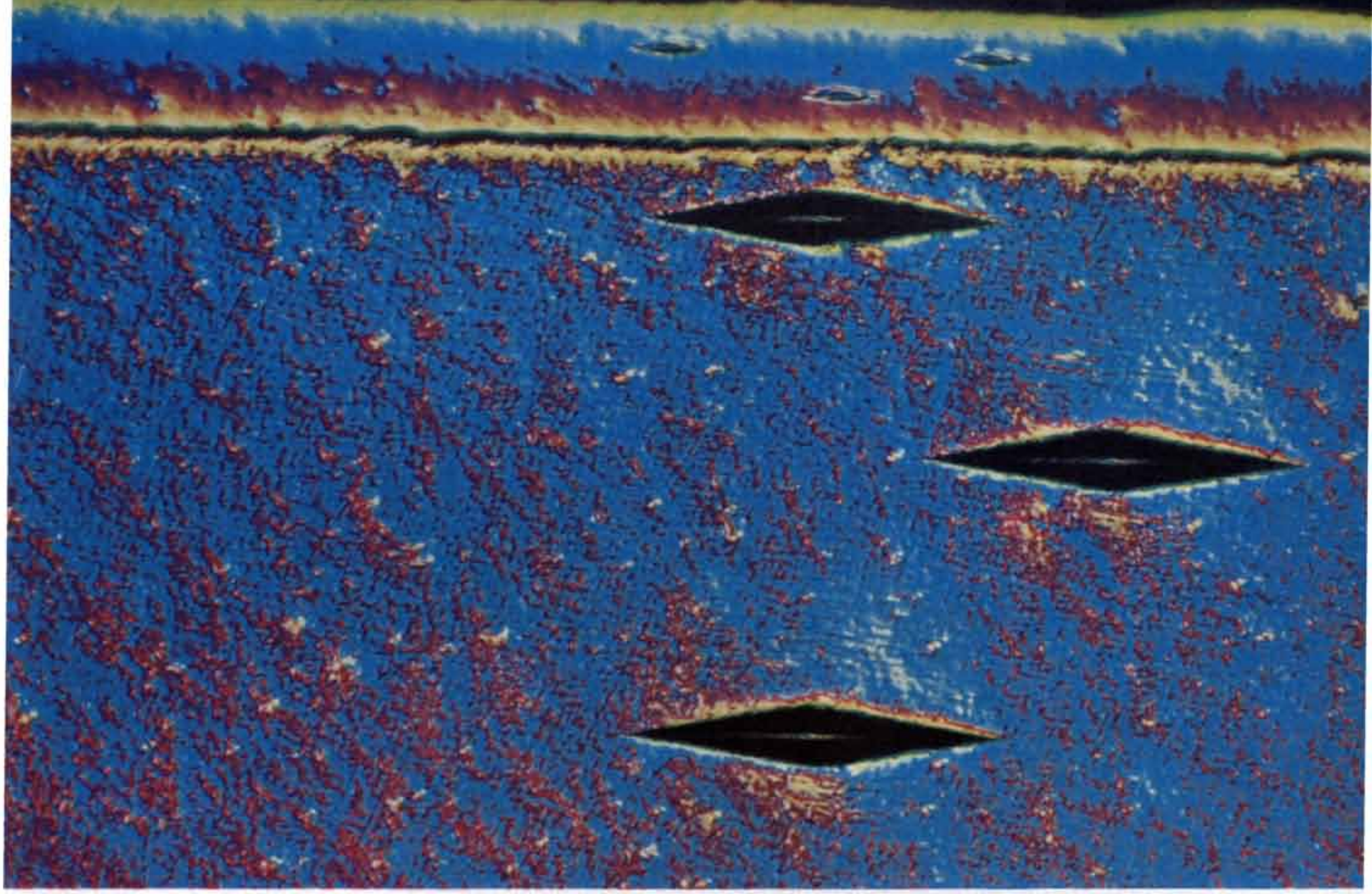
In this way one can achieve a number of desirable changes in properties. For example, the diffusion of boron into the surface of molybdenum produces a surface with a hardness approaching that of diamond. If silicon is diffused into molybdenum, the resulting material can be used in air for hundreds of hours at white heat, whereas untreated molybdenum burns in air at dull red heat and is rapidly destroyed. When beryllium is diffused into copper, the copper is made stronger, more resilient, harder and more resistant to oxidation while retaining its excellent electrical conductivity. By the same token borided steel is as hard as tungsten carbide, titanided copper resists boiling nitric acid and corrosion in air and tantalided nickel becomes almost as resistant to corrosive oxidation as pure tantalum. All together we have been able to use 25 different metals as alloying agents (diffusing metals) and more than 40 different metals as the substrates (receptor metals) on which the alloy surfaces are formed; the total number of combinations we have produced so far is about 400, not counting instances where the substrates are themselves alloys.

Surface alloying is by no means a new concept. In a familiar example, galvanized steel, surface alloying bonds a thin layer of zinc to the steel surface. More recently aluminized steel for trash burners and aluminized and chromized steel for the turbine blades of jet engines have appeared. Most surface alloying, however, has been accomplished by immersing steel in molten zinc, tin or aluminum, all of which melt at comparatively low temperature. Many benefits could be achieved if steel and other metals could be immersed in molten boron, silicon, chromium, titanium, tantalum and so on, but all of these metals melt at such high temperatures that the steel itself would melt on immersion in them. What was needed was a simple, practical and broadly applicable means of alloying metal surfaces. It is such a process that was discovered at the General Electric Research and Development Center, more by good fortune than by design.

We were engaged in a research program aimed at making fluorocarbons by electrolytically fluorinating graphite in molten alkali fluorides. (An alkali fluoride is a combination of one or more of the alkali metals, such as sodium or potassium, with fluorine; in such compounds the name of the more metallic element is given first and is followed by the name of the less metallic one, with the ending changed by the suffix "-ide," so that the compound of sodium and fluorine is called sodium fluoride. We have adopted a similar custom in describing our process: the diffusion of silicon into a metal is siliciding; of beryllium, beryllding, and so forth, so that "metalliding" arises naturally as a generic term for the process.) In the course of our research program we prepared some extremely pure fluoride salts. Silicon tetrafluoride, a gas, was dissolved in these salts to provide silicon cations (positively charged ions) that, on being deposited electrolytically, would not vaporize as the low-boiling alkali metals would.

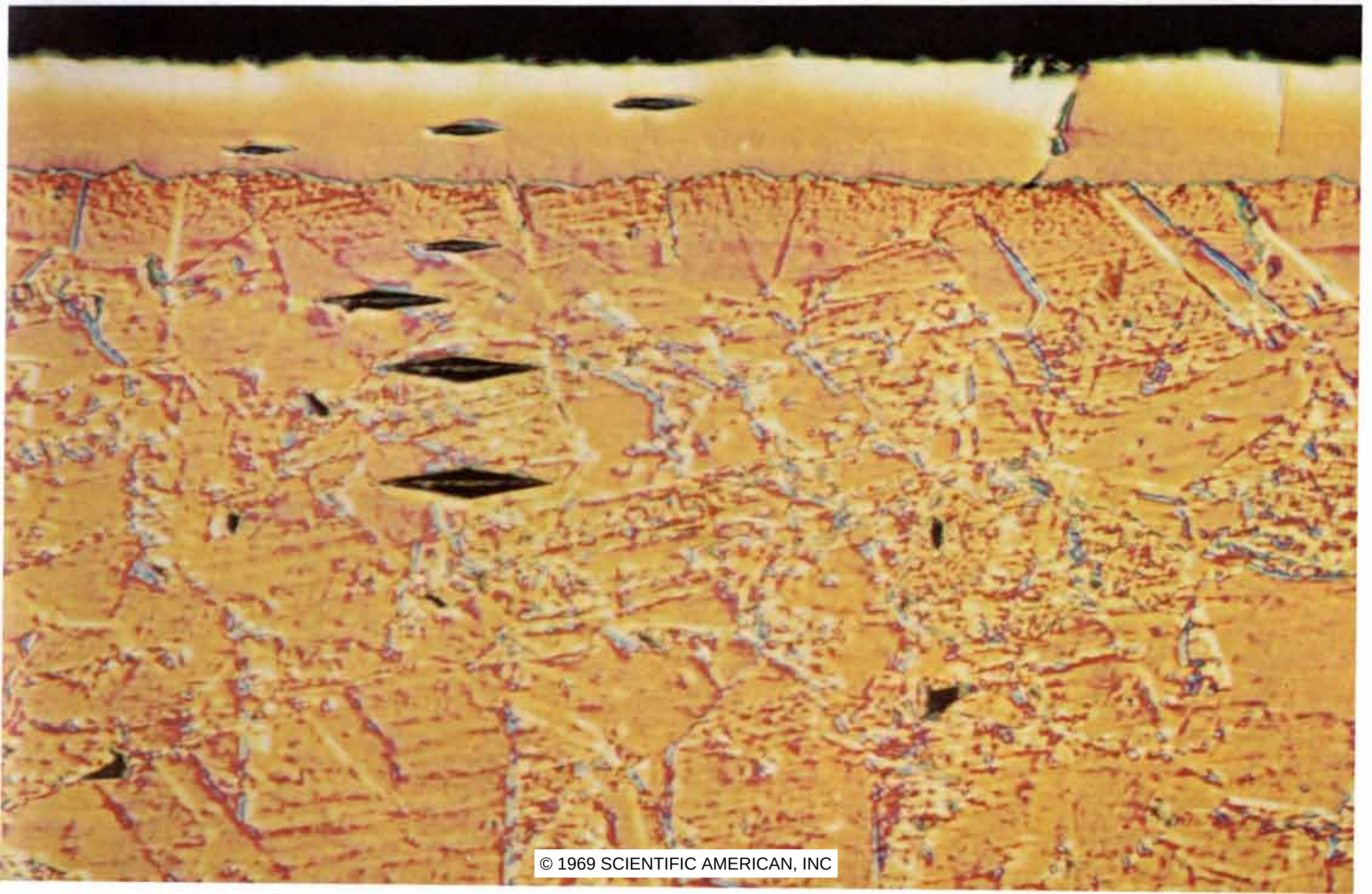
A graphite anode and a platinum cathode were then immersed in the salt, and in a partial vacuum a current was passed through the salt, which was at a temperature of 525 degrees Celsius. As we had wished, fluorocarbons were deposited at the anode, and as we had hoped, no alkali metal was deposited at the cathode. This indicated that silicon had been deposited on the cathode, which was the result we had intended in order to prevent the vaporizing of the alkali metal out of the fluoride salt. We had expected that the silicon would form on the surface of the platinum in a film or dendritic growth. Instead the surface was very bright and had become quite hard. Analysis showed that the deposited silicon had diffused smoothly into the platinum surface.

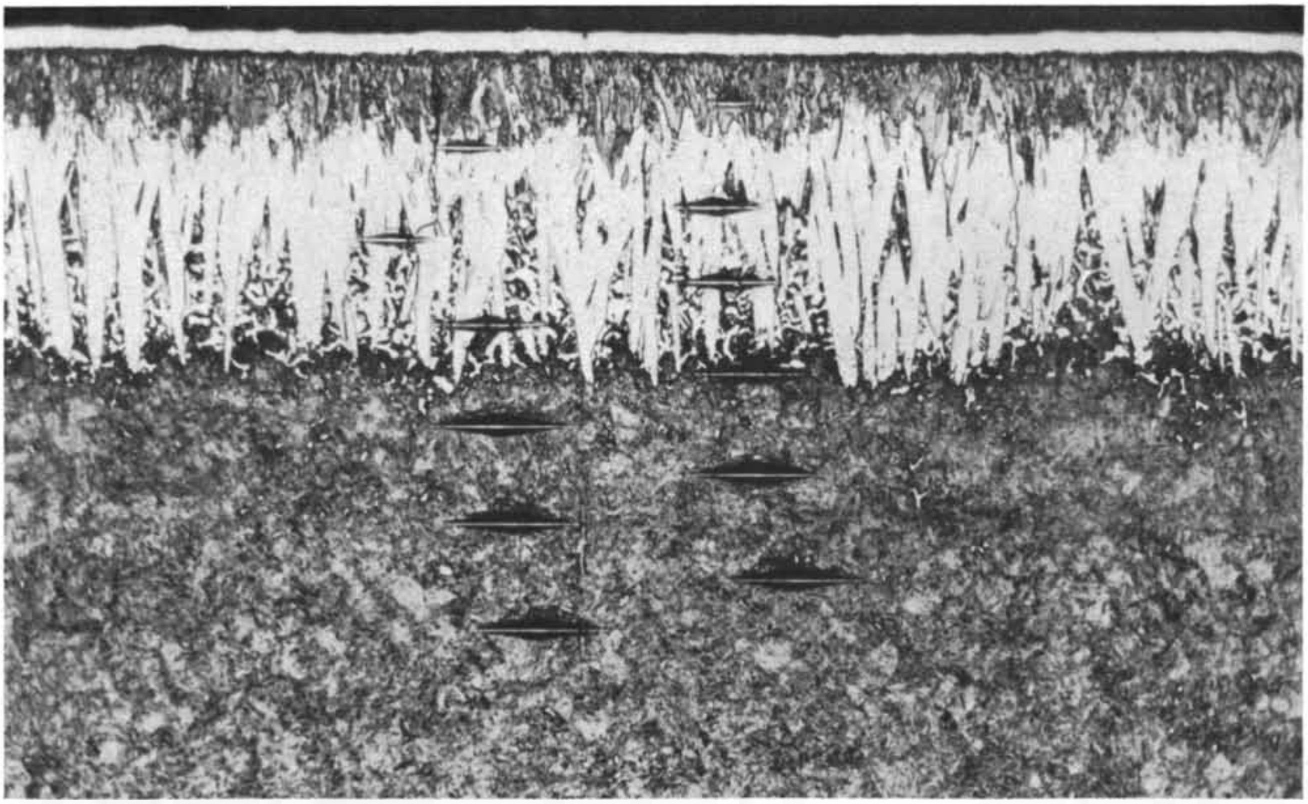
This diffusion was so unexpected at such a comparatively low temperature that we decided to see what would happen if we tried the same thing with other metals. When a molybdenum cathode was substituted for the platinum



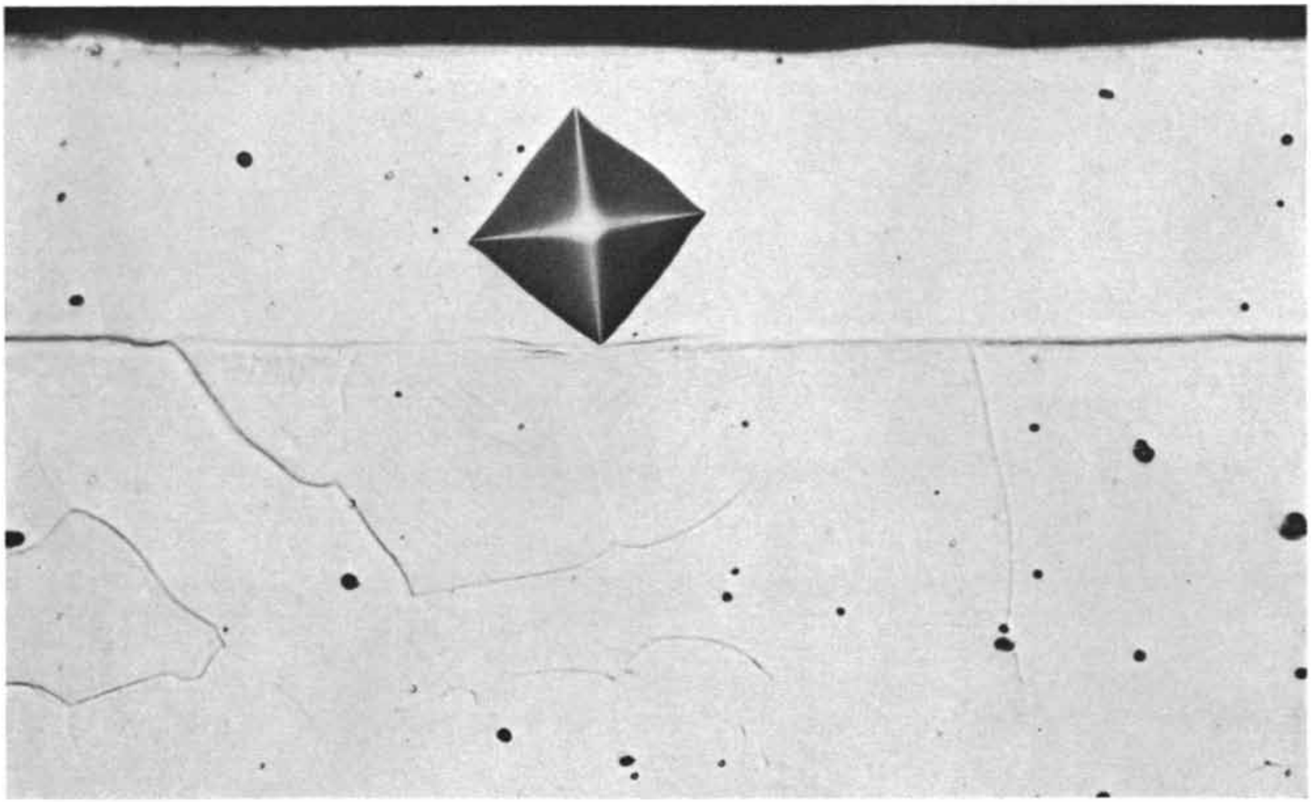
YTTRIDED ALUMINUM coating .003 inch thick (*above*) has formed on an aluminum substrate by diffusion of yttrium atoms into the surface of the aluminum. Thin dark line separating the layers arises in polishing because coating is harder than substrate.

TITANIDED COPPER layer (*below*) was formed similarly. Micrographs were made by Nomarski interference-contrast technique; color is due to interference between waves reflected from surfaces at different depths. Diamond-shaped pits show relative hardness.





BORIDED STEEL has a surface consisting of an intermetallic compound, with a white inner portion of iron boride (Fe_2B) and a darker outer portion of another iron boride (FeB). The hard coating has formed a rootlike penetration of the steel substrate. The coating, which is four mils, or .004 inch, thick, developed in two hours in molten fluorides at a temperature of 900 degrees Celsius.



CHROMIDED STEEL develops a coating at a rate of .5 mil in two minutes in molten fluoride at 1,130 degrees C. Hardness indentation is visible at top. The diffusion of chromium into steel produces a smooth, bright finish. The enlargement is 500 diameters.

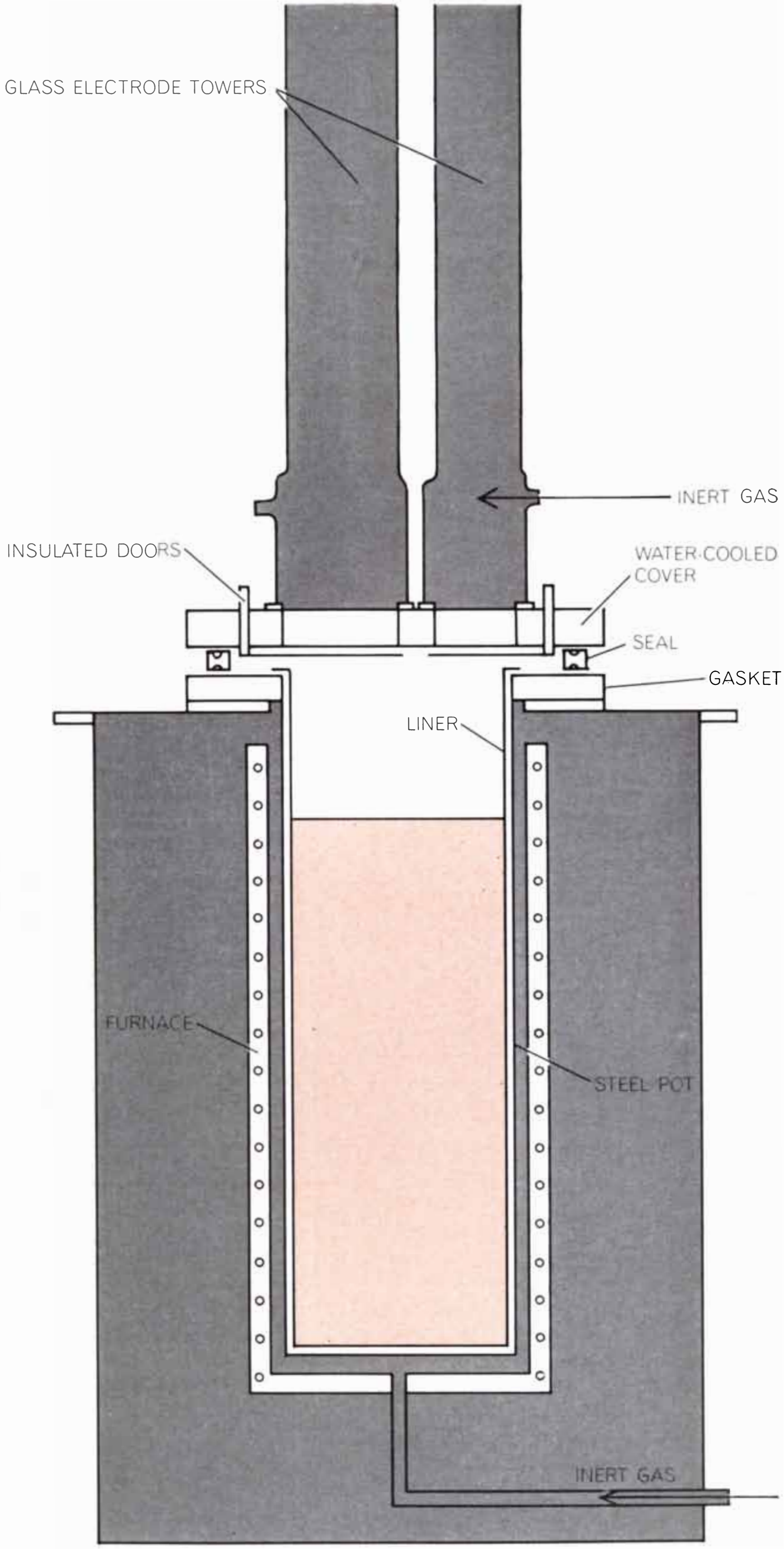
one, diffusion was slow at 525 degrees C. but was quite rapid at somewhat higher temperatures. Such metals as iron, nickel, vanadium and copper were also tried, and all gave similar results.

The possibilities of extending the use of the alkali fluorides as solvents to other surface alloying reactions were intriguing, and we began to explore the process more fully. We wanted to see how generally it would work and what would happen when we manipulated the variables as we applied the process to various systems of metals and baths. These studies have shown that the molten-salt technique can be used with most of the metals on the periodic table as either the diffusing metal or the substrate.

Superficially our process resembles electroplating. In a plating process, however, the electrolytic solution is generally aqueous, and it is at room temperature or slightly above it. The metal of the anode is deposited on the cathode in a somewhat porous layer that adheres primarily by being mechanically interlocked with a film of oxide (or hydrated oxide) on the surface of the cathode. Plating works best in depositing the nobler, or less reactive, metals such as gold, silver, copper and nickel; it cannot be done with the more reactive metals such as aluminum, titanium and silicon because the hydrogen ions always present in aqueous solutions are more readily plated out than the ions of the metals.

In the metalliding process the molten fluorides of alkali metals and alkaline-earth metals take the place of water as the solvent. The fluorides have melting points ranging from about 400 degrees C. to 1,350 degrees. These salts are stabler chemically than any of the fluorides dissolved in them for metalliding. They are ideally suited both as a solvent and as a medium from which metal ions can be deposited on metal surfaces. Moreover, the fluxing action of the molten fluorides dissolves from the surface of the cathode metal the oxide film that forms in air on all metals except gold and possibly platinum. An oxide film on the surface of a metal is always a barrier to the diffusion of other metal atoms into the substrate. The clean surfaces created by the fluoride solvents enable the atoms being electrolytically deposited to make direct contact with the atoms of the cathode's surface and allow diffusion to proceed at the maximum rate.

Since most of the structural metals of the periodic table dissolve in other metals or react with them, and since molten-



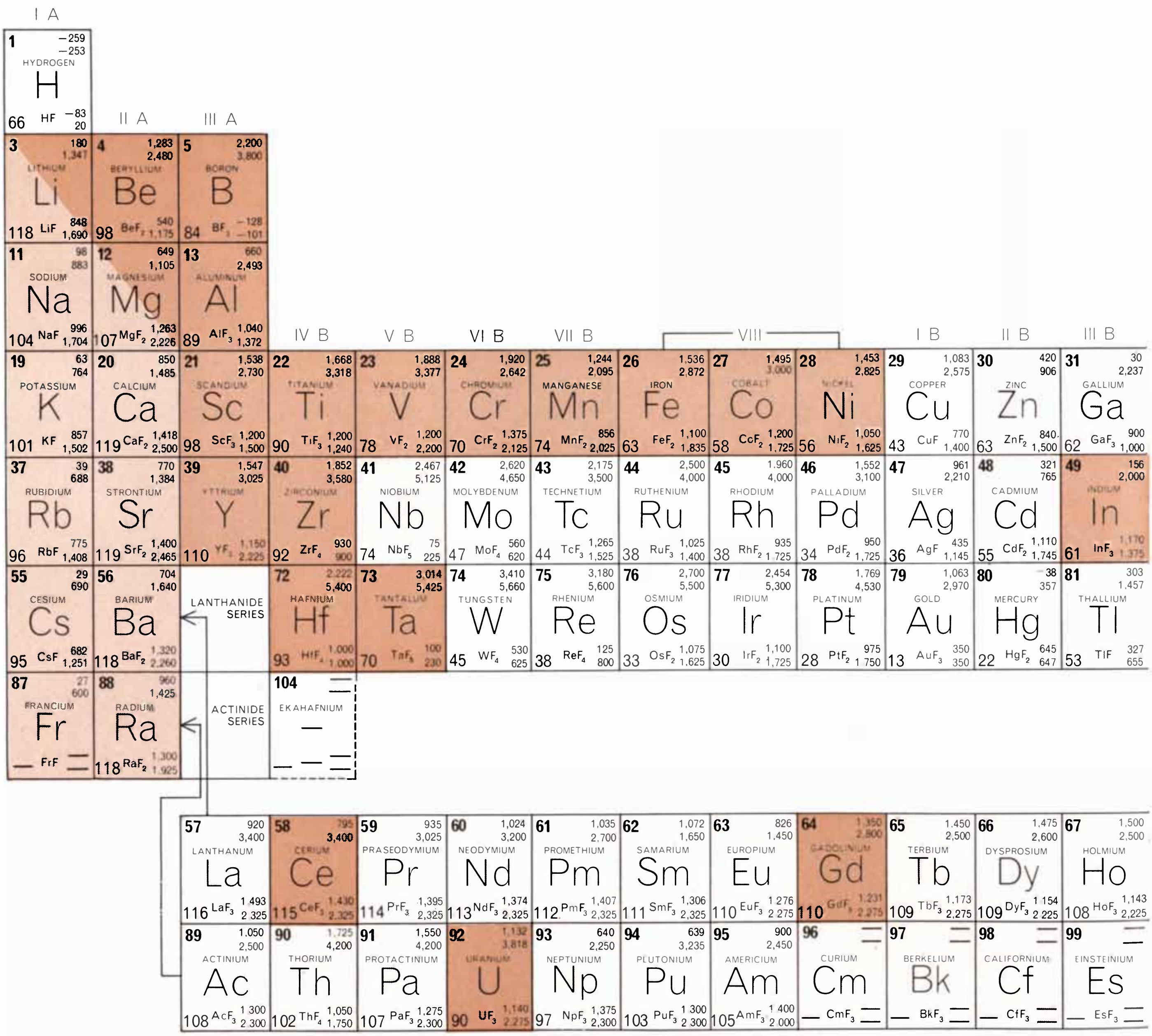
METALLIDING APPARATUS usually has vessels of steel, but they often need liners of special materials so that the electrolytic bath will not be polluted by metal ions arising from the vessel. The inert gas provides an unreactive atmosphere for metalliding. The two metals involved in a metalliding operation are dipped into the hot bath and serve as electrodes; the diffusing metal is the anode and the receiving metal is the cathode.

fluoride solvents can be maintained at temperatures high enough to permit diffusion and reaction, surface alloying takes place. The alloy surfaces are firmly bonded because the diffusing atoms penetrate the original structure and become part of it. The coatings are never porous because the original surface of the completely dense substrate is nonporous, and in accommodating the new atoms the structure of the substrate is only rearranged and expanded.

The alloy coating can usually be formed with a high degree of electrolytic efficiency. Control of the coating's thickness can be quite precise. Most of the coatings are formed in thicknesses of from one mil (.001 inch) to five mils in two to three hours. Some coatings develop more rapidly, becoming several mils thick in only a few minutes, and others form quite slowly, taking two or three days to attain a thickness of one or two mils.

Almost without exception, increasing the temperature has speeded up the coating process. The alloys that are formed at the higher temperatures often have different properties, and sometimes less desirable ones, than the alloys formed at a lower temperature. As the temperature approaches the melting point of the substrate metal or of the alloy surface being formed, the rate of diffusion usually increases rapidly.

Let us examine the process more closely by briefly considering the fluoride chemistry involved. Fluorine is the most electronegative element, that is, the most active in acquiring electrons, usually to form a negative ion. It removes electrons



METALLIDING CHART shows the fluorides, or combinations of metal and fluorine, that have been used as metalliding agents (*dark color*), those used as solvents in which the metalliding fluorides dissolve (*medium color*) and those that are potentially suitable as metalliding agents (*light color*). The arrangement is that of the periodic table of elements. The chemical symbol for the element is in the center of each box, which also includes the atomic number of the element at top left, the element's name at top center, the melting and boiling points of the element in degrees Celsius at top right, the formula of the metalliding fluoride at bottom center, the melt-from the atoms of metals (which are electropositive elements and usually lose electrons to form positive ions) more effectively than any other element on the periodic table. The products of these reactions are metal fluorides (salts), which are generally more ionic (made up of positively and negatively charged species) and more thermodynamically stable than any other group of salts.

The most electropositive metals are on the left side of the periodic table, and they are the metals that form the most stable fluorides [*see illustration on these two pages*]. This is why the alkali and alkaline-earth metals, which are in Group IA and Group IIA at the left on the table, are the ones we use as solvent systems. With the exception of beryllium fluoride, these fluorides can generally be regarded as inert and versatile solvent systems in which other metal ions can be electrolytically transported and manipulated with little or no interference from the solvent medium.

The fluoride solvent systems have a number of other advantages. First, they hold metalliding ions in solution. The alkali and alkaline-earth fluorides combine with the fluorides of all other metals to produce soluble and highly stable fluometallate anions (negative ions). Hence the "-iding" agents dissolve in the molten fluorides whether those agents are a solid with a high melting point or a gas. Usually only a small amount (less than 1 percent) of the "-iding" fluoride needs to be dissolved in the solvent fluoride for the metalliding reaction to take place. The solvent system can be varied according to the type of reaction desired. For example, it is usually advantageous to include potassium fluoride in the solvent system for the siliciding and boriding reactions; fluorosilicate and fluoroborate ions are held much more tightly by potassium ions than by sodium and lithium ions.

Second, the alkali and alkaline-earth fluorides do not form solvent cations that interfere with the alloying reaction. In general the Group IA and Group IIA metals do not dissolve in or form compounds with the structural metals, primarily because the IA and IIA metals have atoms of comparatively large diameter. Therefore fluoride salts of these metals are inert solvents for most metalliding reactions, because metal atoms that are generated electrolytically from the salts do not dissolve in the surface of the cathode or react with it. Before they move many atomic diameters from the surface of the cathode they collide with fluometallate anions and promptly take away fluorine atoms. This liberates "-iding" atoms, which then diffuse into the surface of the cathode.

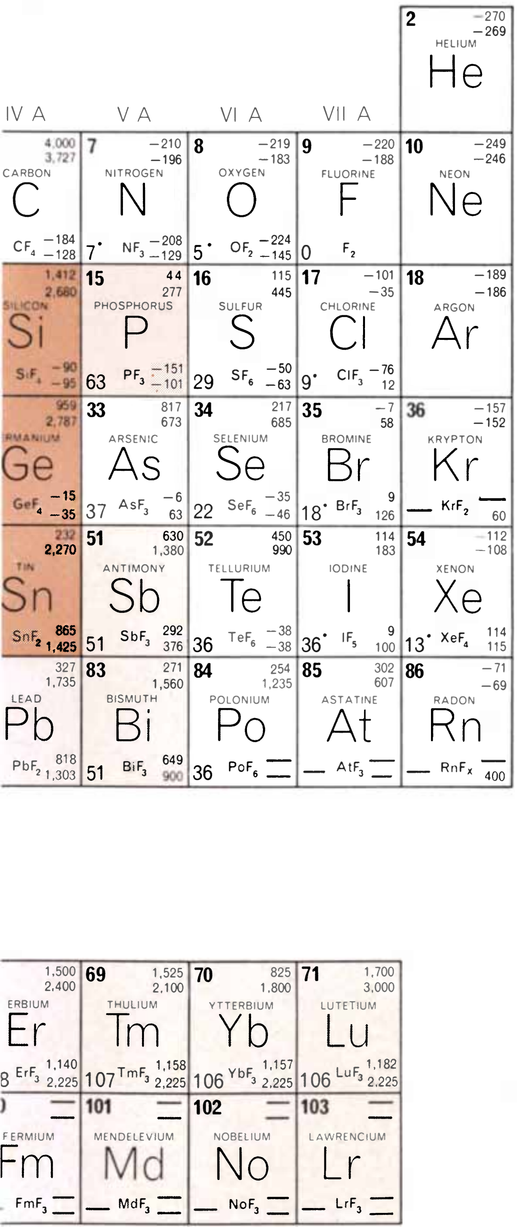
Third, the fluoride solvents are excellent electrical conductors. They are so completely ionized in the molten state that current-carrying capacity has never been a limiting consideration in forming diffusion coatings. Moreover, the solvent fluorides are essentially noncorrosive, particularly when they are largely free of oxygen. They have still other advantages: they have low vapor pressure at operating temperatures, they resist displacement reactions by anode metals and they have a high surface tension (so that little of the salt is removed when a coated piece is taken out of the metalliding bath).

The properties and functions of the fluoride solvents are the salient technical features of the metalliding process. The other features are related primarily to the electrolytic characteristics of the process. Since most of the metalliding reactions involve the diffusion of a more reactive metal into a less reactive one, the system constitutes a galvanic cell—a battery. When the anode is shorted to the cathode by a connection outside the electrolytic bath, electrons flow from the anode to the cathode. As the current flows the anode goes into solution, forming the "-iding" ions that are fundamental to the process. At the cathode the electrons reduce the "-iding" ions to atoms of the anode metal, which then diffuse into the surface of the cathode.

The electrons continue to flow as long as there is diffusion, because the activity of the cathode remains below the activity of the anode. Thus most metalliding reactions will sustain themselves through the battery action of the internally generated electromotive force. In practice, however, an external electric current is usually imposed on the internal electromotive force, with the same direction of flow, in order to achieve a more uniform and higher current density than the battery action will provide. In this way metalliding can proceed from three to 10 times faster than with the self-generated battery action without exceeding the rate at which the alloying agent can diffuse into the cathode substrate.

When the metalliding cell is operating as a battery, the polarity of the cathode is actually positive compared with the anode, whereas in plating the cathode is always more negative than the anode. When in metalliding an additional current is applied from an external source at a sufficiently low amperage, and diffusion occurs rapidly, the entire reaction can be run without the cathode's becoming negative. If the flow of current is interrupted during the applied-current reaction, a rapid return of the cathode to positive polarity indicates that diffusion is keeping up with deposition. Failure of the cathode to return to a positive polarity indicates that the anode metal is starting to plate the cathode instead of diffusing into it.

Metals that are more reactive than the anode metal can often be metallided if the activities of the metals are not too dissimilar, and if the alloying agent diffuses rapidly into the cathode and low-



ing and boiling points of the fluorides at bottom right and bond strength of the fluoride at 1,000 degrees C. at bottom left. The bond strengths marked with a dot are for a temperature of 25 degrees C.

ers its activity. Examples include the aluminiding of titanium, zirconium and hafnium and the berylliding of yttrium. Such reactions involve going against the electromotive series, which lists metals according to their reactivity, beginning with the most reactive (lithium). In such cases a fairly high negative voltage is usually applied to the sample as it is immersed.

Metalliding reagents are generally more active electrochemically than the metals being treated. Hence each reagent has a specific group of metals and alloys on which it will form diffusion coatings, provided that it is soluble in those metals or forms compounds with them. The most active metals offer a wide range of metalliding possibilities, and the less active metals have a correspondingly limited range.

The range of berylliding is broader than that of any other metalliding reaction because beryllium dissolves in more metals and forms compounds with more metals than any other metal does. In addition beryllium forms a very stable fluoride, so that it can be used as an "-iding" agent to coat even the reactive rare-earth metals. Indeed, of the 40-odd structural metals ranging from magnesium to uranium only magnesium and aluminum do not form beryllide coatings. The most promising coatings include those formed on copper, titanium, nickel, cobalt, iron and yttrium. The beryllide coatings are quite hard: they range from 500 to 2,500 on the Knoop scale, on which pure copper is 50 and diamond is more than 6,000. Many of the beryllide coatings are highly resistant to corrosion, owing to the formation of a beryllia-rich protective oxide film on their surface.

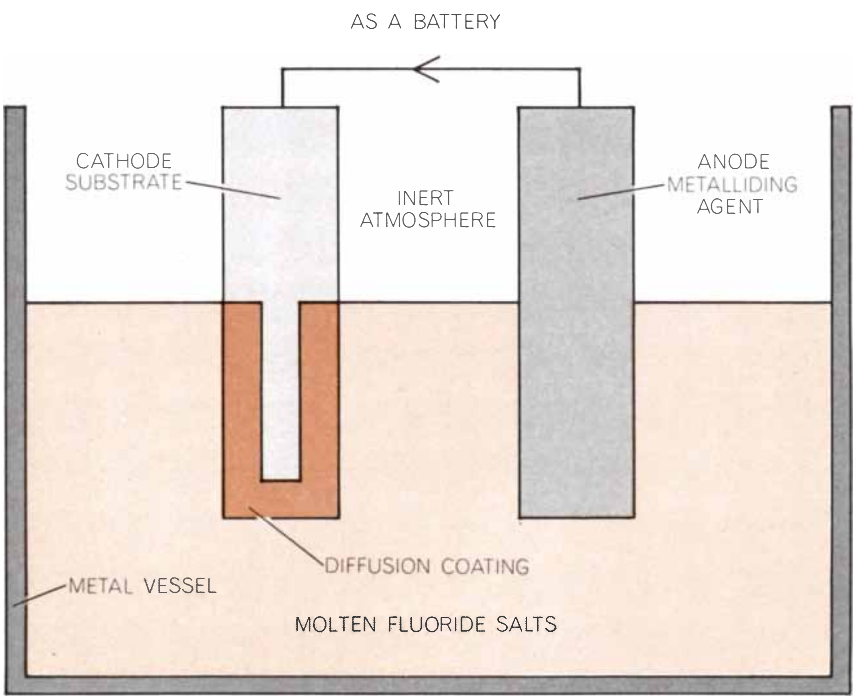
A few examples will indicate the usefulness of the beryllide coatings. Copper can be beryllided readily and has the same properties as bulk alloys of copper and beryllium. After being beryllided for an hour at 750 degrees C. a copper spring made of 1/8-inch wire develops a beryllide coating 10 mils thick with a beryllium content of 2 to 4 percent. The spring has excellent fatigue properties and will support far more weight than a similar spring made only of copper.

Titanium coated with beryllide is as hard as many of the Carboloy materials (about 1,500 on the Knoop scale) and has a much lower coefficient of friction against steel than titanium alone. Beryllided titanium shows no sign of corrosion in air at 1,000 degrees C. for eight hours; untreated titanium oxidizes rapidly in air at 700 degrees. Beryllided uranium is protected from burning at a red heat of 700 to 800 degrees. Similar results have been obtained by berylliding many other pure metals and alloys.

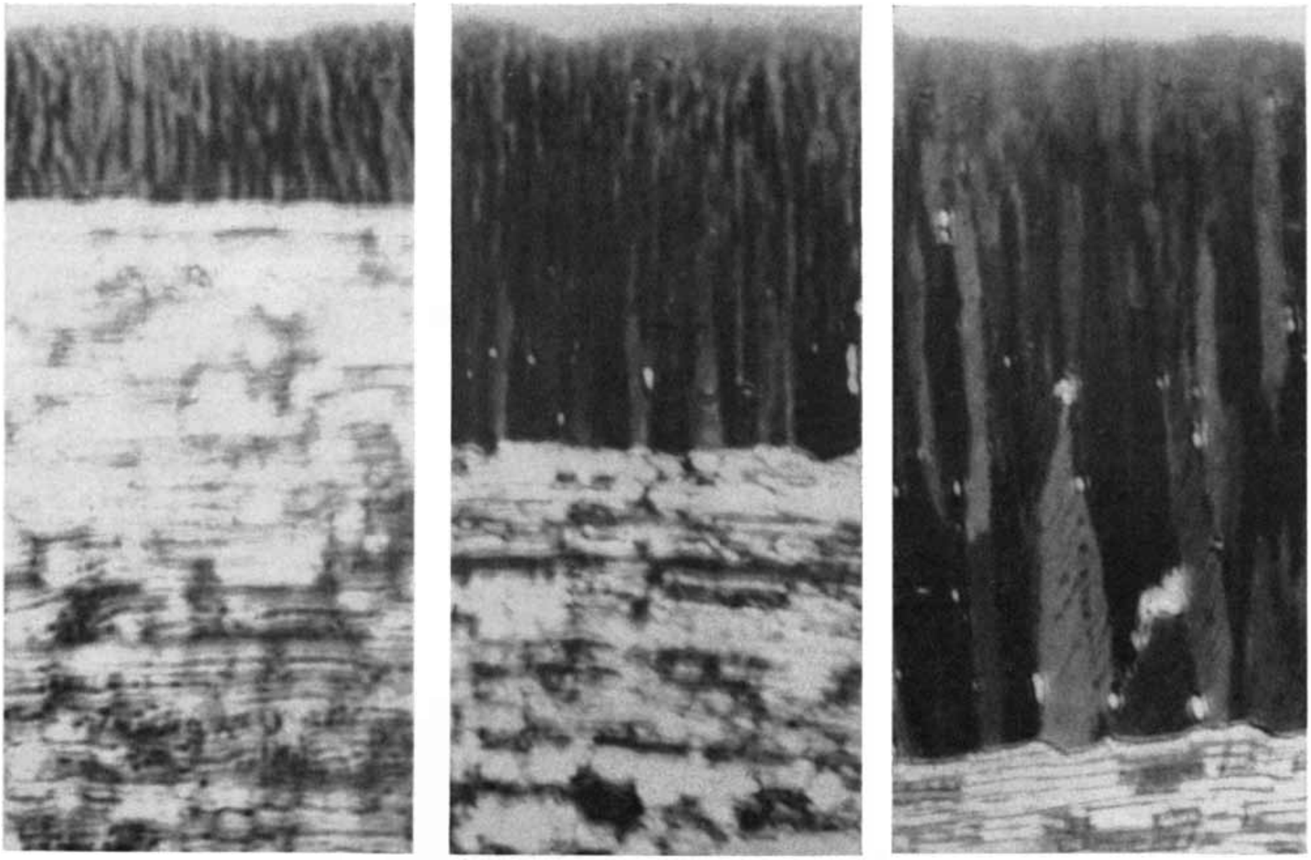
Boron and silicon are similar in reactivity, and so they are similar in the range of their applications as metalliding agents. The metals that can be borided and silicided include vanadium, chromium, manganese, iron, cobalt, nickel, copper, niobium, molybdenum, technetium, ruthenium, rhodium, palladium, silver, tantalum, tungsten, rhenium, osmium, iridium, platinum and gold—a list that contains most of the familiar structural metals. Boriding and siliciding can be accomplished in a large number of salt mixtures but is usually done in a ternary composition of lithium fluoride, sodium fluoride and potassium fluoride.

Once the coating has been applied the similarity of boriding and siliciding ends; the coatings are significantly different in properties and potential applications. Boride coatings are exceptionally hard: on steel they usually fall between 1,500 and 2,500 on the Knoop scale, and often they exceed 3,000. On simple steels and many alloy steels the coating develops a rootlike attachment as the boron diffuses in; the coating is tightly anchored and maintains its integrity even when the material is considerably deformed. The boride coatings usually have poor resistance to corrosion (except on stainless steels), but this can be remedied by lightly chromiding and siliciding the boride layer. Borided steels show great promise for bearings and for dies. At their present stage of development they are too brittle to be used as cutting tools.

Cobalt does even better than steel at forming well-anchored and hard boride coatings, which give the material good properties as a spring. On molybdenum, boride coatings have measured above 4,000 on the Knoop scale; they are so hard that they can be polished only slowly even with diamond compounds. These coatings have proved to be exceptionally good bearing surfaces for pumps handling liquid sodium and potassium at 650 degrees C. The only metals of the group listed above that do not form good boride coatings are copper and silver; boron is not sufficiently soluble in and reactive with them to form good compounds. There is, however, a compensating factor: for the same reason screens of copper and silver can be used as basket anodes to



ELECTROLYTIC SYSTEM employed in metalliding reactions is depicted schematically. The metalliding agent, serving as the anode, dissolves in the molten fluorides, becoming positive ions because of the tendency of the fluorine in the solvent to capture electrons. At the cathode, which consists of a piece of the metal to be coated, electrons from current flowing externally through the system reduce the ions to atoms of the anode metal, which then diffuse into the surface of the cathode, giving the cathode metal new surface properties.



GROWTH OF COATING of silicide on molybdenum in a metalliding process carried out at a temperature of 675 degrees C. is rapid. Coating attains a thickness of .5 mil in 50 minutes (*left*), one mil in 100 minutes (*middle*) and two mils in 200 minutes (*right*).

hold chunks of elemental boron in the boriding process.

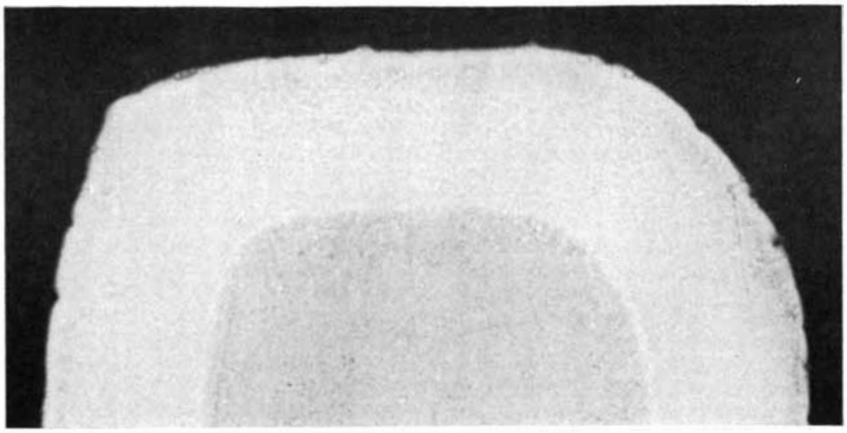
Well-defined silicide coatings can be formed on all the metals of the group except silver, in which silicon has low solubility and forms no compounds. Here again there is a compensation: a silver basket holding chunks of silicon makes an excellent anode up to 775 degrees C. (Above this temperature a silicon rod becomes sufficiently conductive to serve as an anode itself.)

Silicide coatings are generally more brittle and less hard (500 to 1,500 on the Knoop scale) than borides, but they strongly resist oxidation. Silicided molybdenum provides the most dramatic and useful example of such behavior. This resistance to oxidation has great potential; molybdenum is one of the most important refractory metals (having a melting point of 2,610 degrees C.), but it cannot be used in air at temperatures above 500 degrees because it oxidizes to form volatile molybdenum trioxide. The silica film that develops on the surface when silicided molybdenum is heated forms a remarkably protective envelope, as can be seen in the photograph on the cover of this issue of *Scientific American*. Silicided molybdenum successfully resists burning in air for more than 750 hours at 1,500 degrees C., more than 100 hours at 1,600 degrees and 25 hours at 1,700 degrees. Modified silicide coatings have lasted for an hour on being cycled to temperatures above 2,200 degrees.

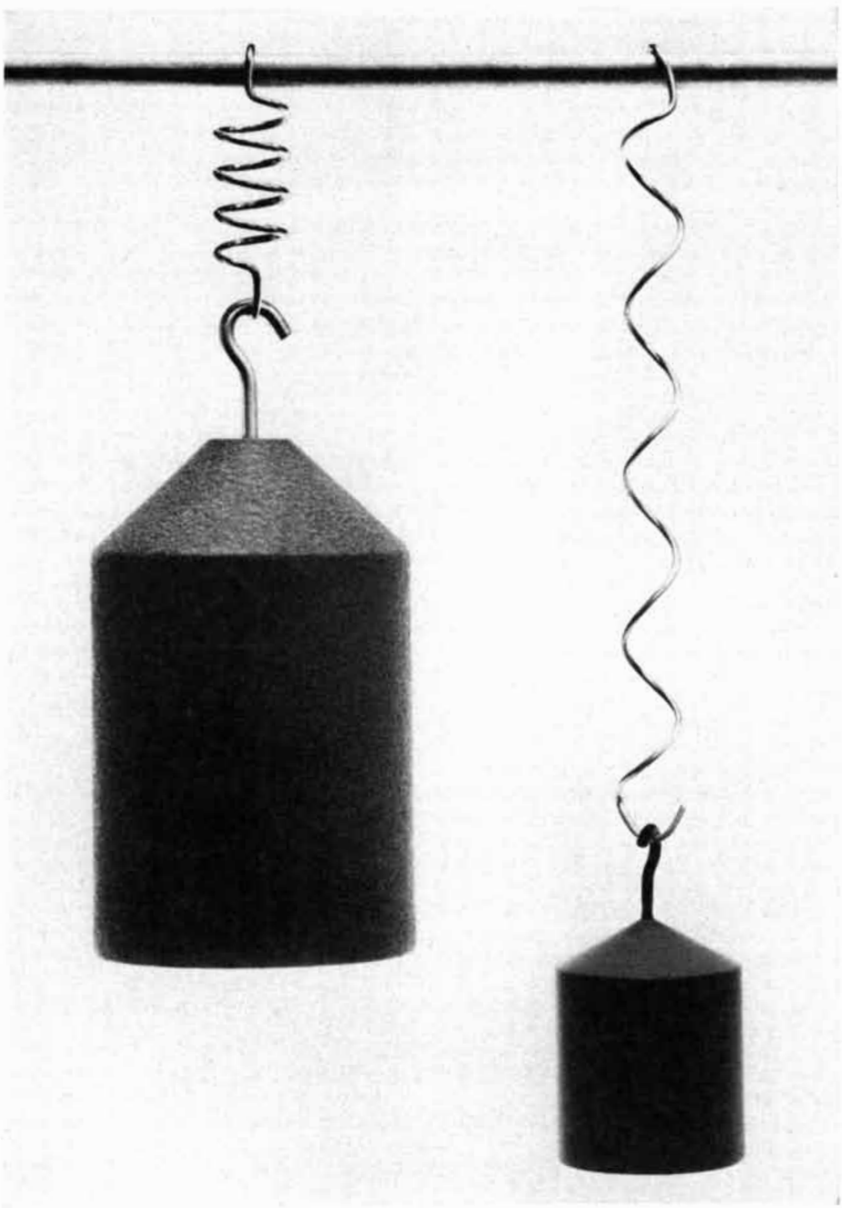
Titaniding has been one of the most difficult processes to master because of titanium's extreme sensitivity to traces of oxygen. When the salts are clean, however, titanium consistently and readily forms excellent diffusion coatings on the same general group of metal substrates used in boriding. The coatings on mild steels, nickel-based alloys and copper form up to many mils thick and are highly resistant to corrosion, particularly by acids. Titaniding in conjunction with other "-idings" such as berylliding, boriding and aluminiding offers a multitude of sequentially or simultaneously formed double-alloy coatings with tremendous potential as wear-resistant and corrosion-resistant surfaces.

Aluminiding electrolytically in molten fluorides offers many advantages lacked by the conventional method of immersing metals in molten aluminum. The technique should therefore extend aluminiding into new areas. Copper, for example, cannot be aluminided by immersion in molten aluminum because it would rapidly dissolve, but it can be aluminided in molten fluorides, forming alloyed surfaces up to many mils thick in a few minutes. The metals that can be aluminided are the same ones that can be borided, with the addition of the metals in Group IVB of the periodic table. The principal benefit is improved resistance to oxidation.

Yttrium and the rare-earth elements are the most reactive structural metals on the periodic table, with activities equal to or just below those of the alkali metals and the alkaline-earth metals. Thus they can form diffusion coatings on most metals through the metalliding process. Yttride, ceride and gadolinide coatings have been made with good yields on nickel, cobalt, iron, copper, platinum, rhenium, chrome steel, zirconium and other metals. Yttrium and cerium have also been diffused into aluminum to form hard coatings. Better results were obtained with chloride salts than with fluoride ones, however, because the chloride salts melt at lower temperatures and are more stable with respect to aluminum. More effort needs to be made in



BERYLLIDING OF COPPER is shown in this photomicrograph at an enlargement of 145 diameters. Beryllided copper has a number of properties not possessed by copper alone.



NEW PROPERTY of beryllided copper is strength. The beryllided spring at left holds a 25-pound weight with little deformation; a five-pound weight straightens a copper spring.

this direction, since wear-resistant coatings on aluminum should find wide application.

Vanadium, chromium and manganese are similar in reactivity and are high enough in the electromotive series to form diffusion coatings on most of the "boride group" of metals. At 1,130 degrees C. chromium diffuses rapidly into low-carbon steels, forming coatings four mils thick in less than an hour. The coatings are notably ductile, free of "pinholes" and resistant to corrosion.

Tantalide coatings can also be formed on most of the metals in the boride group. On nickel the coating is readily formed up to three mils thick. Even in coatings less than a mil thick the surfaces have essentially the resistance of pure tantalum to chemical corrosion.

As one moves farther to the right on the periodic table the scope of the metalliding processes becomes much more limited. The diffusion of iron, cobalt and nickel is restricted to about 12 metals. Iron can be diffused into cobalt and nickel with good results, but attempts to do the reverse have been unsuccessful. The diffusion of nickel into molybdenum, tungsten and copper has been quite successful and should find many uses. Germanium, which occupies a slightly higher position than nickel in the electromotive series, has readily coated nickel but has formed no coatings on iron. Germanium is readily diffused, however, into molybdenum, palladium, copper, platinum and gold.

Several metals clustered around germanium on the periodic table are potentially useful as metalliding reagents, but most of them are so low in activity that their range will be quite limited. Only cerium and gadolinium of the lanthanide series have been widely tested in metalliding processes; nonetheless, all these metals have a large range, and others should serve well as surface alloying agents. It is always dangerous to predict that theoretically possible events will not come to pass, but it does seem unlikely that the highly radioactive metals of the actinide series will be employed as metalliding reagents.

In sum, as is shown in the accompanying chart [*pages 42 and 43*], 24 metalliding reactions have been studied, and there are about 36 more metals whose fluorides are stable enough for the purpose. Hence metalliding appears to have the broadest range of all the surface-alloying methods for getting the most out of both the substrate metals and their specially tailored surfaces.

Founded on faith

"Lightly screened neutral donors at the inner edge of the depletion region account for intrinsic Appelbaum-Anderson zero bias anomalies in metal-semiconductor tunnel junctions. A negative g-shift and a natural Zeeman level width $\hbar/T_1$ for the localized magnetic moment are implied by the Appelbaum theory. These features are observed in the experimental results. . . . Since the same s-d exchange Hamiltonian applies both to the tunneling experiment and the dilute alloy problem, it appears that tunneling experiments . . . may permit direct measurement of previously inaccessible magnetic properties predicted for the Anderson model of dilute alloys. It is possible that our measured g-shift, in this sense, is the first direct observation of a large first-order Yosida g-shift."

—E. L. Wolf and D. L. Losee, Physics Division, Kodak Research Laboratories, in *Solid-State Communications 7:* 665 (1969)

Wolf and Losee need fields up to 150 kilogauss for much of this "zero-bias anomaly" work of theirs. For respectable fields you have to go to the Francis Bitter National Magnet Laboratory. After a while you sort of assume that no serious person would need to be told what a negative g-shift is. Actually, of course, not very many laboratories are known yet to be concentrating as hard as we are on experimental clarification and theoretical explication of these zero-bias anomalies in tunneling through a Schottky barrier.

—Laboratory Head, Solid State Laboratory

A Schottky barrier is a name given long ago to an insulating region that forms in a semiconductor when you put a metal of much different work function on the surface. Tunneling through it becomes a very simple idea with a little background in quantum mechanics and kind of a waste of time to talk about otherwise. It has had a play in the electronics business. Some day it could turn out to be useful in recording and displaying information.

—Physics Division Head

We are pleased to find ourselves in a position of some leadership in the exploration of these very fundamental phenomena. We don't know if or how this particular fraction of our work will bear a significant payout for the shareowners. But we need to understand how things behave to help in finding a more direct route to really new marketable technology.

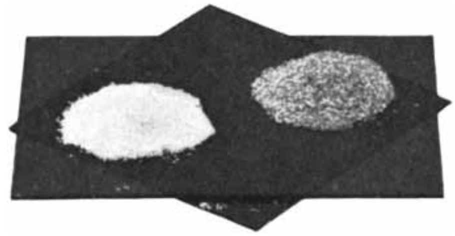
—Research Director

You have to have faith.

—General management

Kodak

Humanism in plastic

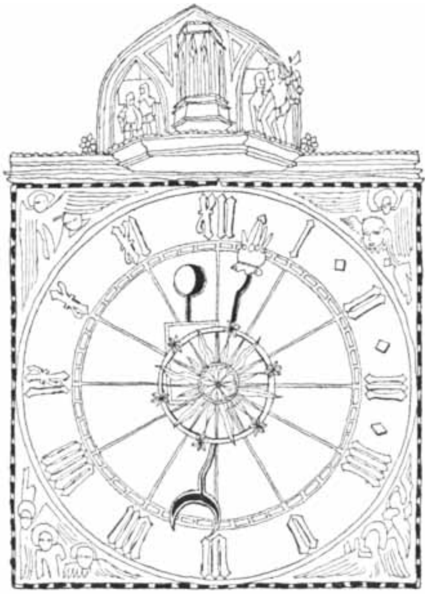


Plastics. We make them. In some quarters we are known better for them than for our famous photographic goods. People take Tenite and Uvex plastics from us and shape them into things of use and beauty. Rarely one of a kind, though. Replication for widespread possession spoils the beauty for some eyes. Against that, credit plastics with helping keep pride in the making of something serviceable and tangible for one's living. Use of tools more sophisticated than hammer and chisel need not diminish the pride.

That man's hand is an extension of his mind is a basic belief among the educators who specialize in "industrial arts." To reveal to youth the humanistic aspect of all-pervasive technology is their mission. Readers who see some merit in realistically directing the attention of the schools to the work of man's wondrous digits—in plastics or other materials—can get a steer from the professional organization of teachers known as the American Industrial Arts Association, 1201 16th Street, N.W., Washington, D.C. 20036.

EASTMAN KODAK COMPANY
An equal-opportunity employer

SCIENCE AND



The UN and the Environment

Warning that the future of life on the earth could be endangered by the continued deterioration of the human environment, Secretary-General U Thant has submitted to the UN plans for a global conference in 1972 that would "provide a focus for worldwide action to avoid a possible crisis." The convening of such a conference was approved last December by the General Assembly, and the Secretary-General has now submitted a report on the nature of the problem, the proposed scope of the conference and preparations for holding it.

The report relates the deterioration of the environment to three basic causes: population growth, urbanization and the advent of new technology. Each of these trends has advanced exponentially in recent decades, with the result that for the first time in human history "all areas on the earth's surface have been to some degree modified by man." The prospective doubling of the world population in less than half a century will intensify crowding, urban problems and damage to the natural environment, and "no nation can any longer be isolated from these global pressures. It has become clear that we all live in one biosphere within which space and resources, though vast, are limited."

The Secretary-General has suggested that the conference, for which Sweden has offered to serve as host, should avoid detailed and technical consideration of specialized topics. It should concentrate, rather, on calling attention to broad problems and mobilizing public action to cope with them. U Thant identified four groups of "substantive" problems and four groups of "strategic" problems or modes of action. The substantive problems are (1) human settlements and industrial development, (2) use and development of natural resources, (3) environmental pollution and (4) maintenance of human environmental values. The strategic problems are (1) economic and social planning, (2) financial policy, (3) public administration and legislation and (4) regional and international cooperation. Individual commissions would be established to deal with each of these areas of concern. U Thant emphasized the necessity of broad participation and asked the General Assembly to take steps to ensure adequate representation from the developing countries.

Cold Power

A recently completed study, sponsored by the Edison Electric Institute and conducted by the Linde Division of the Union Carbide Corporation, has demonstrated the feasibility of transmitting high-voltage, alternating-current electric power through underground cables made of superconducting niobium tubes immersed in liquid helium at −452 degrees Fahrenheit. The study is one of seven such research projects initiated by the Edison Institute to explore various new approaches to the problem of distributing the large blocks of electric power that will be needed in major urban areas in the next few decades.

At present, and for the foreseeable future, the most economic means of distributing electric power for long distances are overhead transmission lines. When such lines reach the limits of the growing metropolitan centers, however, they must go underground, and there's the rub: conventional underground cables are not only more expensive and less efficient than overhead lines but also limited in their power-carrying capacity. Energy losses in the form of heat increase with the power load, eventually reaching a limit above which the heat cannot be dissipated fast enough into the surrounding soil.

Refrigerating conventional underground conductors by surrounding them with liquid hydrogen at −420 degrees F.

in order to absorb heat losses and cut resistance has already been tested and has been found to extend the power-carrying capacity of the cable about ten-fold. Reducing the temperature still further in order to take advantage of the phenomenon of superconductivity (the resistanceless flow of electricity in certain materials at temperatures near absolute zero) initially ran into the problem that the known superconducting materials exhibited zero resistance only for direct-current transmission; in alternating-current cables they were subject to reactive energy losses. Hence it was believed that in a superconducting distribution system the a.c. power carried by the overhead lines would have to be converted to d.c. for underground transmission and converted back into a.c. at the point of use—a procedure that would be prohibitively expensive.

What the Linde research effort has shown is that these a.c. reactive losses could be reduced to an acceptable level by constructing a superconducting cable of extremely pure niobium. The niobium is electrolytically deposited as a thin film on the surface of a tubular conductor in a molten-salt bath (see "Metalliding," page 38).

The power-carrying capacity of such a superconducting cable would be about 25 times the capacity of a conventional underground cable. For example, it is estimated that one 20-inch superconducting pipe system could handle up to 10,000 megavolt-amperes of power at a voltage of 345 kilovolts, which is more power than New York City currently consumes. The projected costs of such a high-capacity, low-loss transmission line are substantially less than the present costs for the large number of conventional high-voltage cables that would otherwise be needed. Linde has proposed a 12-year, $8-million research program to develop and refine a superconducting distribution system based on this principle, which is at present "only slightly beyond the conceptual stage."

The Arrival of Mitosis

One of the major events in the evolution of life on the earth was the appearance of the first eucaryotic cells: cells that have well-defined nuclei and multiply by mitosis, the process in which a succession of precise nuclear changes precedes cell division. Before that event life had developed to the level of the procaryote, a primitive cell that multiplied by simple fission. The procaryote is represented still by the bacteria and the blue-green algae; all higher organisms are descended from the first eucaryotic cells. Now a team of paleontologists and geologists has reported finding microscopic fossils that represent the oldest eucaryotes yet known: primitive algae that appear to be between 1.2 and 1.4 billion years old.

The fossils were collected in 1966 in the southern Nopah Range east of Death Valley in California but their age was not clearly established until last year, write Preston E. Cloud, Jr., of the University of California at Santa Barbara, G. R. Licari of the University of California at Los Angeles, Lauren A. Wright of Pennsylvania State University and Bennie W. Troxel of the California Division of Mines and Geology. Their report is in the *Proceedings of the National Academy of Sciences.*

In both the Nopah Range and the nearby Kingston Range the fossils are embedded in chert, a fine-grained siliceous sedimentary rock. Several varieties are described. Some are globular, with dark spots that suggest various kinds of subcellular structure, including nuclei, and may be related to living green algae. These varieties are associated with chainlike filaments of procaryotic blue-green algae. Two other groups of minute globular cells appear to be eucaryotic green and yellow-brown algae. The authors point out that the presence of eucaryotes implies the availability of at least some free oxygen in the earth's atmosphere, a necessary precondition for the emergence of higher, multicelled organisms.

Pest Plagues

As it becomes increasingly clear that potent but unspecific chemical pesticides such as DDT have unpredictable effects far beyond their original purpose, the search is pressed for alternatives to such pesticides. As far as the control of insects is concerned, one of the most attractive alternatives has been the insects' own bacterial and viral parasites. These organisms can be mass-produced, mixed with a liquid or powder "carrier" and sprayed as if they were ordinary insecticides. In principle such biological agents have two major advantages over chemical pesticides; first, they attack only one species or genus and do not harm other organisms, and second, whereas some chemical pesticides accumulate in the soil and water, bacteria and viruses simply join populations already in existence. Infectious agents may have another advantage: in both laboratory and field tests insects have so far failed to evolve resistance to them. Some species of insects can become immune to chemical agents within a few generations because resistant strains replace the susceptible ones.

Bacterial agents have already fulfilled some of this promise. Lettuce and cabbage farmers now use bacterial insecticides (derived from *Bacillus thuringiensis* and manufactured by the International Minerals & Chemical Corporation and Nutrilite Products, Inc.) that attack a larva called the cabbage looper. Fungi can also be used against insects. Entomologists working with the World Health Organization have shown that one fungus, *Coelomomyces*, may be able to reduce mosquito populations. Now viral agents may soon be on the market. Several have already been developed and tested successfully by the U.S. Department of Agriculture and industrial firms. Nutrilite and International Minerals have had applications pending with the Food and Drug Administration for more than a year for permission to produce viral insecticides capable of controlling the corn earworm, the cotton bollworm and the tobacco budworm.

Biologists and virologists have noted, however, that viral and bacterial agents have some technical limitations that may prevent them from replacing chemical pesticides. First, their specificity is not an unmixed blessing: a new agent must be found and developed for each insect. Furthermore, some viruses and bacteria may perform well in the laboratory but cannot be made into a pesticide at a reasonable cost. Finally, some agents may take several days to destroy their target, whereas chemicals act almost instantaneously.

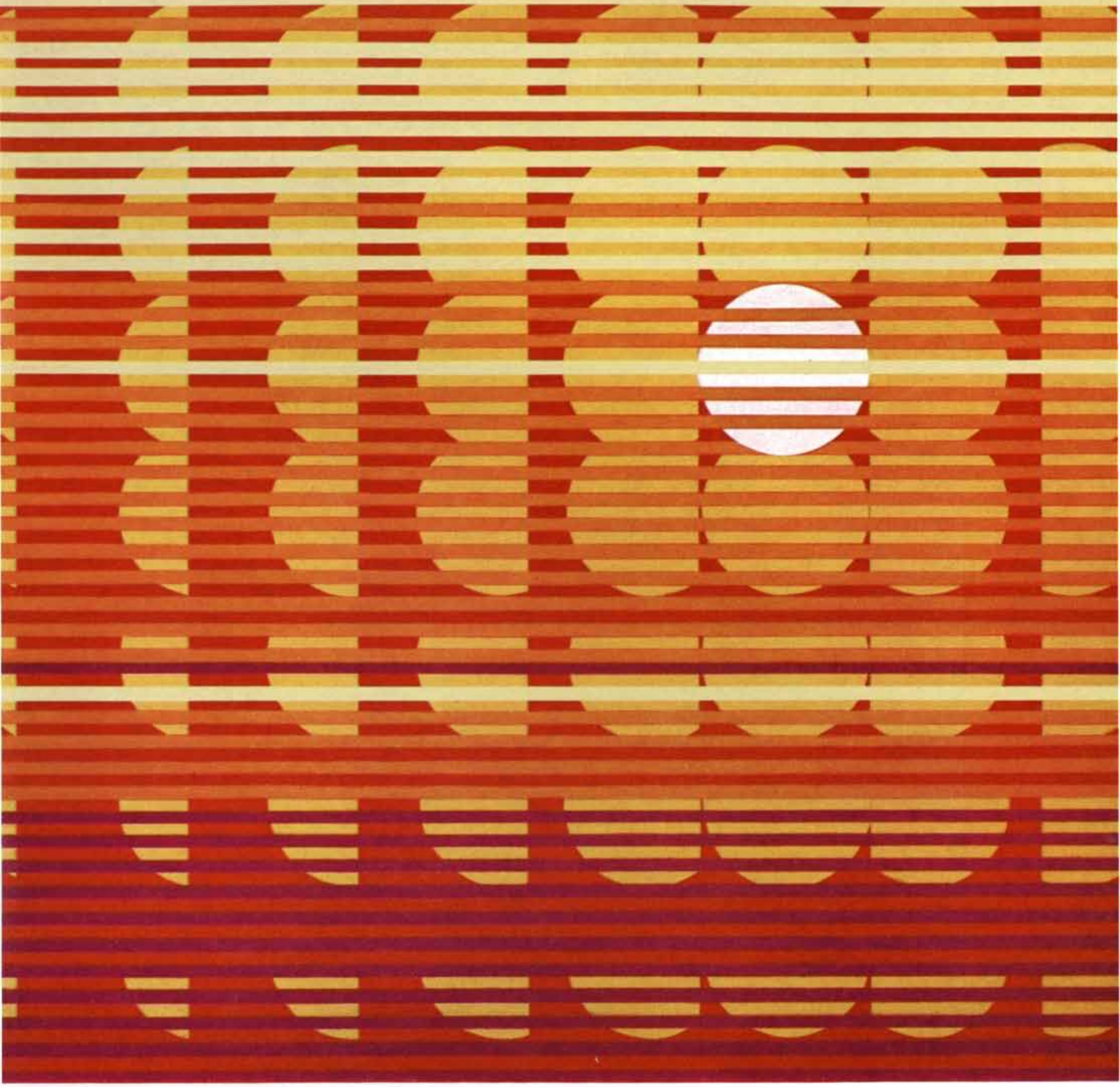
There is also the problem of whether the bacterial and viral agents themselves are safe. The fact that the FDA, which usually renders a decision in 90 days, has been holding the Nutrilite and International Minerals applications for many months indicates how seriously the agency takes the issue. Advocates of viral pesticides argue that, since insect tissue bears no serological relationship to human tissue, the viruses that infect it would have to undergo highly improbable mutations in order to threaten man. These workers feel that viral pesticides have been thoroughly tested for safety. They also point out that such organisms already exist in nature without infecting human beings. FDA officials, however, apparently want to be sure the viruses that attack insects will not, when used as pesticides, harm man as well.

Australarctica

More evidence has been adduced to support the concept of continental drift. Walter Sproll and Robert S. Dietz of the Atlantic Oceanographic Laboratories of the Environmental Sciences Service Administration, report they have succeeded in demonstrating that Antarctica and Australia, now separated by 2,000 miles of ocean, were once a single land mass.

Sproll and Dietz compared bathymetric data, much of it gathered by the ship *Oceanographer* over southern Australia's continental slope, with similar information about the contours of Antarctica's continental slope. Concentrating on the 1,000-fathom isobath (a line around each continent at that depth), which they believe represents the true edge of each land mass, they fed their data into a computer at the University of Miami. The computer then compared segments of the two continental margins until it found the best fit between them by matching the convex coast of Wilkes Land in eastern Antarctica with Australia's Great Bight, the concave stretch of the southern Australian coast that runs from a point south of Perth eastward to Melbourne and Tasmania. After the computer had plotted the two isobaths on a chart, Sproll and Dietz put in three submerged "continental fragments"—the Naturaliste and Bruce plateaus in the west and the Iselin plateau in the east—and found that they neatly bracketed Australia's southern coast. These land masses had intentionally been left out of the earlier programming so that they would not bias the trial runs. Although stretches of the continental edges on each side of the break are badly slumped, and sediments may have extended the continental slope, the fit between the two margins is so good that the total area of misfit is smaller than the state of Illinois.

The original land mass probably broke up because of activity along the southeast Indian ridge, which runs between Australia and Antarctica. Such ridges are actually rifts in the earth's crust through which the mantle rises and flows out across the ocean floor. The ridges seem to mark the cracks dividing the coast into plates that float away from each other on the plastic mantle under them. According to the evidence of the magnetic anomalies on each side of the southeast Indian ridge, Australia and Antarctica began drifting apart 40 mil-



what is creativity?

Creativity, innovation, imagination . . . whatever you call it . . . is the ability to form new relationships between previously unrelated things. Whether you are composing a symphony, baking a cake or synthesizing a new chemical, the process is the same. At Celanese we put a premium on creativity. We endeavor to apply imagination boldly to the vigorous pursuit of ideas for new and improved chemical products. Products that will fill the needs of people everywhere. The key word is "needs." For Celanese is market oriented. We do not invent cures for non-existent diseases. Our scientists and technologists apply creative input to the rigorous disciplines of polymer chemistry. This we do in chemicals, fibers, plastics and petroleum products. We consider creativity a marketable commodity. And a reason our sales have grown to 1.3 billion dollars last year.

CELANESE

Now we can get rid of mosquitoes

Mosquitoes are more than a pest. They are a menace. The diseases they carry kill more people than cancer.

The most effective way of killing these mini-monsters is to attack them where they breed. In water. And man has tried many weapons. Among them, diesel fuels and insecticides. But all have a serious snag. They can pollute the water or harm plants and wildlife.

Last year, Humble Oil & Refining Company (a Jersey affiliate) introduced a new weapon – FLIT® MLO. Mosquito control experts are getting more and more excited about it. And so are the conservationists. With good reasons on both sides.

A thimbleful of FLIT MLO can kill several million mosquito larvae. Yet, even when used in large quantities, FLIT MLO will not harm fish, birds or plants. Nor will it pollute the water.

FLIT MLO is a unique water-white liquid which spreads an imperceptibly thin film over the water's surface. It kills larvae by a non-poisonous process.

without hurting anything else.

When the larvae come up to breathe, they actually ingest the product deep into their bodies. So deep, they can't wash it out. They die from lack of air. It is difficult to build up resistance to that.

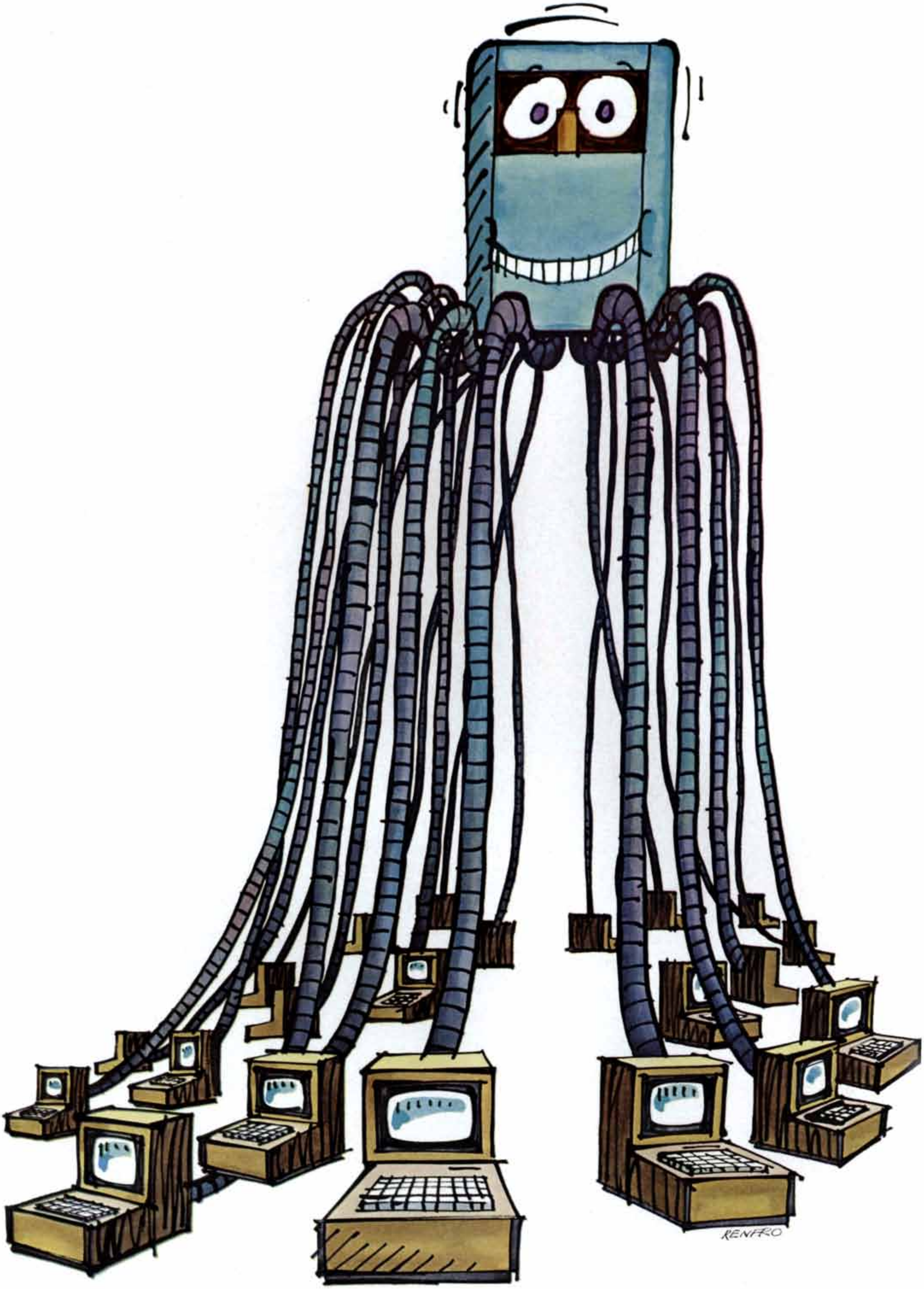
FLIT MLO can also give the adult female mosquito quite a shock. It reduces the surface tension of the water. When she swoops down to lay her eggs, her feet prick through the surface as she lands. She sinks and drowns.

And a similar thing happens in the unlikely event a new mosquito is born in a treated area. It clambers out of the pupa to dry its wings, puts its feet in the water and down it goes.

Mosquito control crews are now being trained to use FLIT MLO in ten different countries on three different continents. Mosquito legions beware. This may seem like a skirmish. But it could be the start of a global war.

Standard Oil Company (New Jersey)

Esso


RENFRO

The Octoputer

RCA's many-tentacled computer does time sharing plus regular computing. It's a generation ahead of its major competitor.

Once there were only monster computers that did big batch jobs like payrolls. Problem? They kept people waiting.

Then came the whirling dervishes of time sharing that let a lot of people work at once. Inefficient. The machines were only busy part time.

Now there's a new kind of creature that does time sharing and batch work together. So lots of people can use it—efficiently.

It's the Spectra 70/46. The Octoputer. There's nothing quite like it on earth or under the sea.

Remote computing is its basic concept. That means it works from way over here where you are, to way over there where it is. The Octoputer's arms are long and strong.

It sits in the middle of your company and reaches helping hands out in all directions. Suddenly, your company works harder. More of your people use the computer—solving more problems, finding more facts, writing more programs. And it does your big batch jobs in its spare time.

The Octoputer does a real armload of work for a handful of change. Check the bills from your time-sharing services. See if it's not more efficient to do the same work on your own Octoputer. And get batch processing, too.

One more thing. The Octoputer concentrates on remote computing because that's what you're going to need—that's where the industry is going. We got there first because communications is what RCA is famous for.

It'll keep us ahead of our competition. It can keep you ahead of yours. Step up to the Octoputer and shake hands hands hands hands hands...

We will arrange an introduction. Call 609-424-2385. Or write RCA Computers, Cherry Hill, N.J. 08034.

RCA COMPUTERS

lion years ago, 60 to 100 million years after the other continental blocks had sundered.

The Sproll and Dietz findings do not, however, help to settle the disagreement between the geologists who think there was only one vast land mass they call Pangaea and those who believe all the continents were once parts of two supercontinents: Gondwanaland (comprising Australia, Antarctica, Africa, India, South America and Malagasy) and Laurasia (consisting of North America and Eurasia). Resolution of the controversy will have to await more extensive surveys and comparisons of the continental margins.

A Pox on Pox

Chicken pox, which is usually a mild disease, can be serious for certain children, and the problem of immunizing them against it has been a difficult one for a number of reasons. These difficulties appear to have been surmounted by the ingenious technique of inoculating such children with globulin obtained from adults who were recovering from shingles, which is caused by the same virus that causes chicken pox. The technique was developed in experiments at the New York University Medical Center and is described in *The New England Journal of Medicine* by Philip A. Brunell of the New York University School of Medicine and several co-workers.

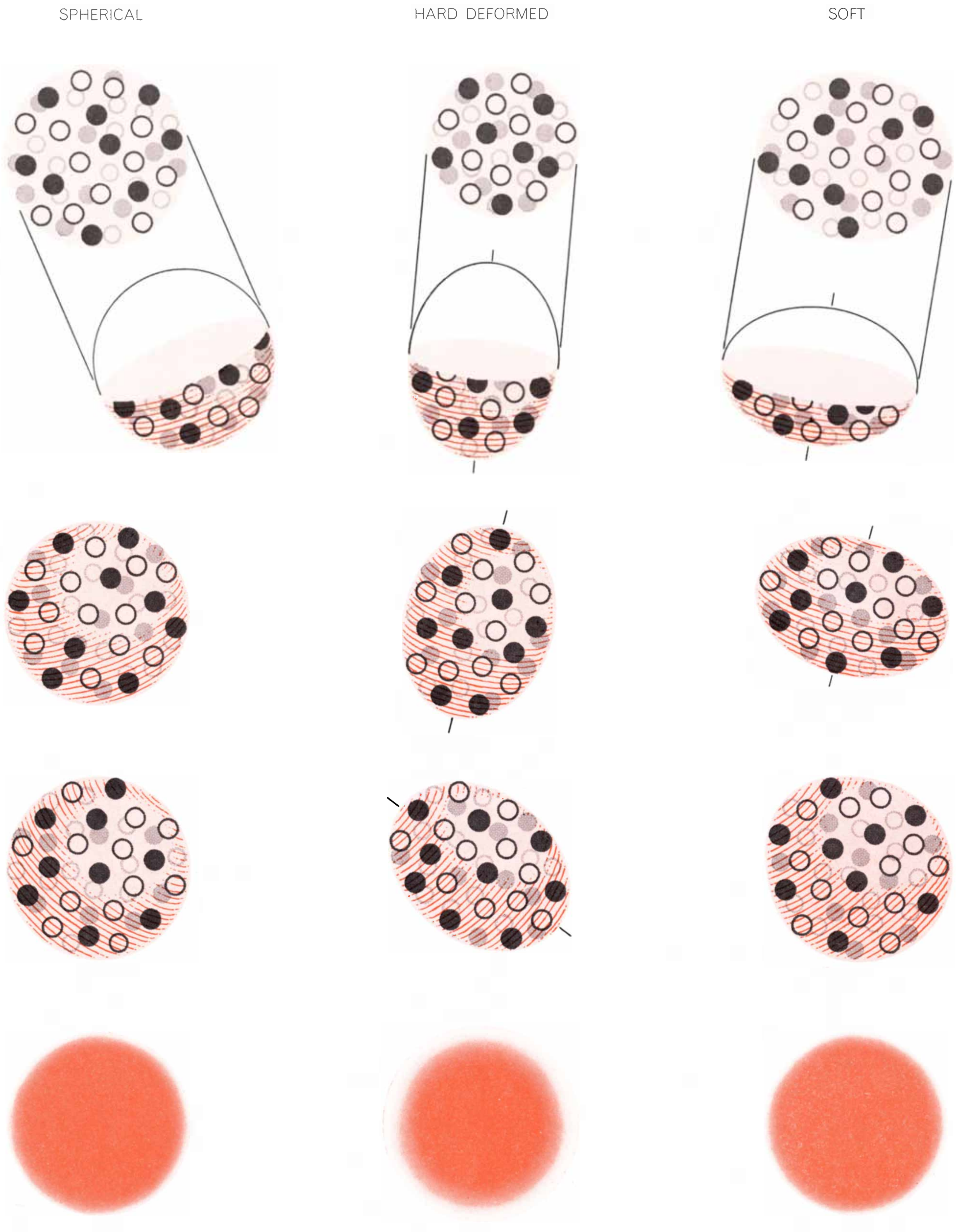
One might think that children for whom chicken pox (also called varicella) was potentially dangerous could be immunized with globulin from children who have had the disease or with a live-virus vaccine. The first approach encounters both legal and physiological difficulties; the second one is undesirable because the virus in question is a latent one, so that a child vaccinated against chicken pox might later develop shingles (also called herpes zoster). Efforts to immunize children against chicken pox by means of human serum globulin had proved unsuccessful (although large doses did lessen the severity of the disease). Brunell's group therefore prepared globulin from patients who were convalescing from shingles, because the serum had a higher content of varicella-zoster antibody than is found in immune serum globulin. Working with siblings who had been exposed to chicken pox, the group administered immune serum globulin to one child in each of six families and zoster immune globulin to the other. In each of the six families varicella developed only in the child who had received immune serum globulin; none of the children who had received zoster immune globulin contracted the disease. Brunell and his associates are now seeking to determine the minimum amounts of varicella-zoster antibody to prevent chicken pox in children and to ascertain the dose of zoster immune globulin that would protect adults.

Time-Zone Effects

It has been clear for some time that long-distance flights across several time zones put a significant stress on the air traveler by interfering with the normal 24-hour cycle of physiological processes. In an article in *Science* Peter V. Siegel, Siegfried J. Gerathewohl and Stanley R. Mohler of the Federal Aviation Administration review recent investigations into these time-zone effects and make some recommendations for minimizing them.

Many human physiological functions fluctuate rhythmically in the course of a day: wakefulness, hunger, psychological and motor performance, temperature, heart rate and a number of metabolic processes. In the jet age these circadian (approximately 24-hour) oscillations are put out of phase with the environment. For example, a traveler who leaves New York in the evening crosses seven time zones in a seven-hour flight to Rome and arrives in the morning. It is then 1:30 A.M. New York time, however; the traveler is ready to sleep and is not hungry. By pushing himself he can get through the day but his efficiency will be low for the first 24 hours. The return trip may be even more debilitating. Leaving Rome at 11:00 A.M., he arrives in New York in the early afternoon after a 10-hour flight. He may be able to stay awake for a while, but soon after he finally goes to sleep his biological clock and his hunger for breakfast (which is "due" at what is now 1:00 A.M.) wake him up.

The authors recommend a formula worked out by the International Civil Aviation Organization to determine how much rest time should be provided after a business flight. The rest time depends on the travel time, the number of time zones crossed in excess of four, the departure time and the arrival time. As it works out that might mean a day of rest after a flight from Montreal to London, two and a half days for Montreal to Karachi, one and a half days for Montreal to Lima. The authors also recommend that "superimposed stresses," particularly heavy eating and drinking, be kept to a minimum in the rephasing period after a flight.



NUCLEI, made up of protons *(gray)* and neutrons *(open circles)*, are classified as spherical, hard deformed or soft, defined in terms of "snapshots" that reveal shape and orientation. The geometry of the shapes is shown in the top row. The hard deformed nuclei are prolate spheroids, ellipsoids with two equal short axes and one long axis. Soft nuclei change shape; the one at top right is an ellipsoid with three different axes. The two middle rows show snapshots of "even-even" nuclei in the ground (unexcited) state. Spherical and hard deformed nuclei maintain their shapes through time; only the orientations change, with all orientations equally likely. A soft nucleus may become oblate (two equal long axes and one short axis), spherical, prolate or pear-shaped. The bottom row shows "time exposures" of the same nuclei, which are equivalent to superpositions of many snapshots. All of them have spherical symmetry.

The Size and Shape of Atomic Nuclei

Some nuclei are spherical, some are ellipsoidal and some fluctuate in between. Measurements of radius and precise shape provide tools with which theoretical physicists can refine models of the nucleus

by Michel Baranger and Raymond A. Sorensen

If an atom were the size of a house, the nucleus would be a pinhead at its center. In actuality atoms are a few hundred-millionths of a centimeter in diameter, and on the atomic scale the nucleus seems virtually a point without structure. Nonetheless, its size and shape can be measured. In recent years experiments with powerful and sensitive instruments have yielded increasingly refined measurements for the nuclear physicist to ponder. In this article we shall describe some of the experimental techniques and their results. We shall also undertake to show how the internal motions and structure of nuclei can be understood by supplementing the concepts that govern the motions of ordinary matter with the concepts of quantum mechanics, which was invented to describe the much coarser motions of electrons in atoms.

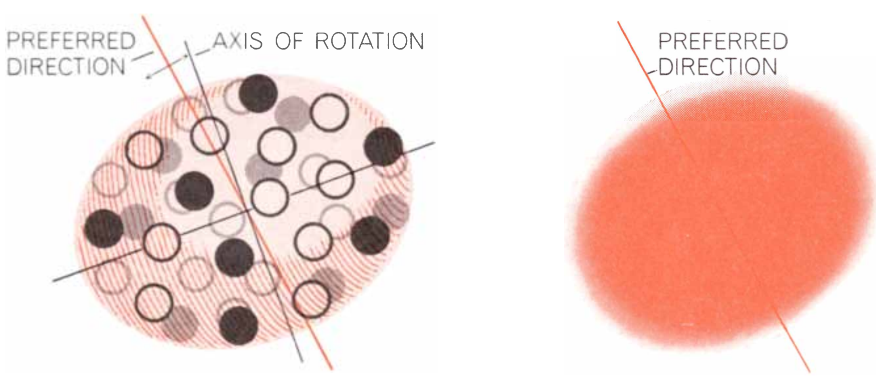
The nucleus directly manifests itself in one natural terrestrial phenomenon, radioactivity, and it is of supreme importance as the source of stellar energy. In other respects its properties are seldom in evidence; the structure of substances and their chemical transformations depend on the outer part of the electron cloud that fills most of the atom. The nucleus consists of two kinds of particle: protons and neutrons, together called nucleons. Their total number—the mass number—is designated *A*. The proton has a positive charge equal to the negative charge of an electron, and so, since the atom is neutral, the number of protons, *Z*, is the electron number. The neutrons are neutral and their number is denoted *N*, hence $A - N = Z$. Usually there are several kinds of nuclei for each element. They have the same number of protons (which establishes their complement of electrons and therefore their chemical behavior) but a different number of neutrons; each such nuclear species, or isotope, is identified by its chemical name followed by its mass number, as in carbon 13 or uranium 238.

Atoms are a few angstroms (10^{-8} centimeter, or a hundred-millionth of a centimeter) in diameter. The nucleus is some five orders of magnitude smaller—a few ten-trillionths of a centimeter in diameter. Its size and shape can nevertheless be defined much as they would be for a larger object, and we shall first discuss these definitions in some detail.

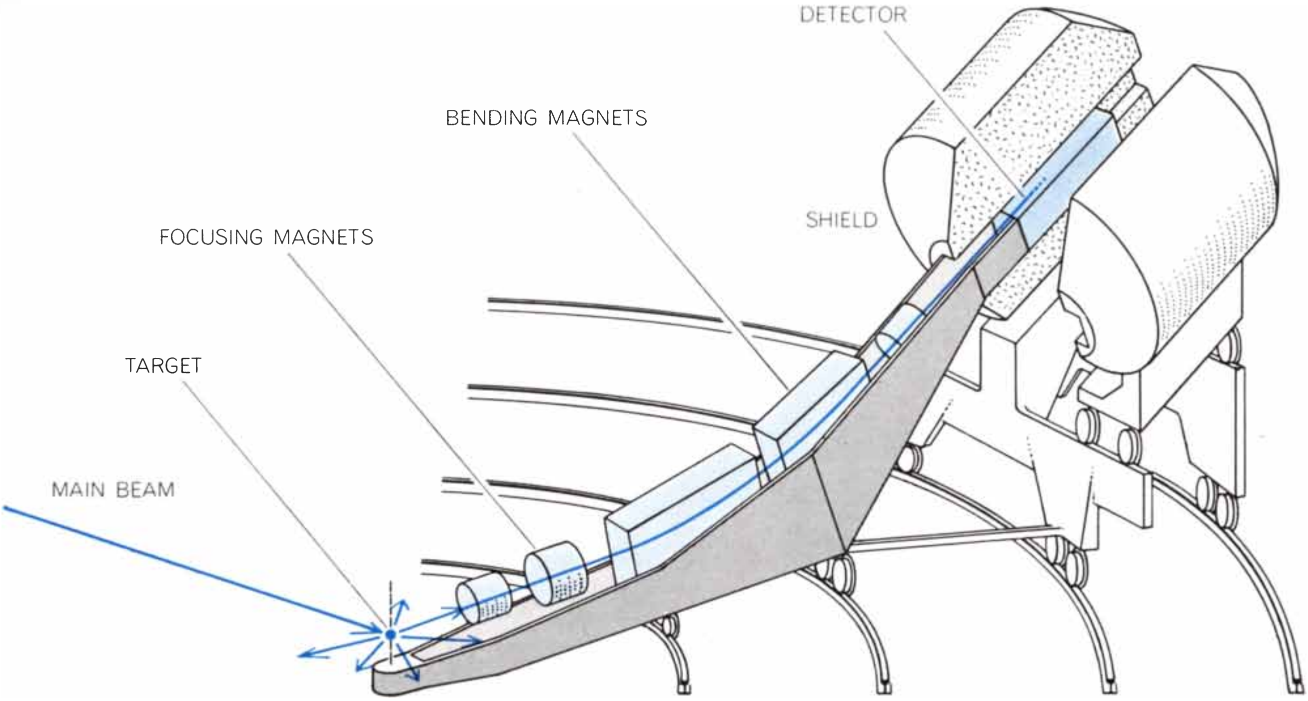
What Are Size and Shape?

The notions of the size and shape of a solid object at rest are familiar enough, but the nucleus consists of particles in rapid and complicated motion. Does it make sense, then, to refer to its size and shape? A good analogy is provided by an airplane propeller. When it is rotating, it appears as a blurred circle, and a casual observer might take the circle as its intrinsic shape. A careful observer with a fast enough camera, however, can make a photograph that reveals the propeller's true shape. This "snapshot" agrees with our customary notion of shape; the "time exposure" the first observer depended on reveals only an average shape, which may be quite different. Now suppose we take a second snapshot of the propeller, or of another propeller on an identical airplane. The result is the same shape as before, identical as to length of blades, angle between them and so on; the only possible difference is the orientation of the propeller in its plane of rotation. The shape of the propeller is permanent, and so we can say that the propeller is "hard." It would be different if we took snapshots of two identical octopuses, for instance. The two shapes would almost certainly be different; the octopus is "soft."

The shapes and sizes of nuclei must be defined in terms of snapshots. Moreover, we can distinguish between "hard" nuclei, whose shape is permanent, and "soft" nuclei, whose shape is changeable. In addition it is possible to make vari-



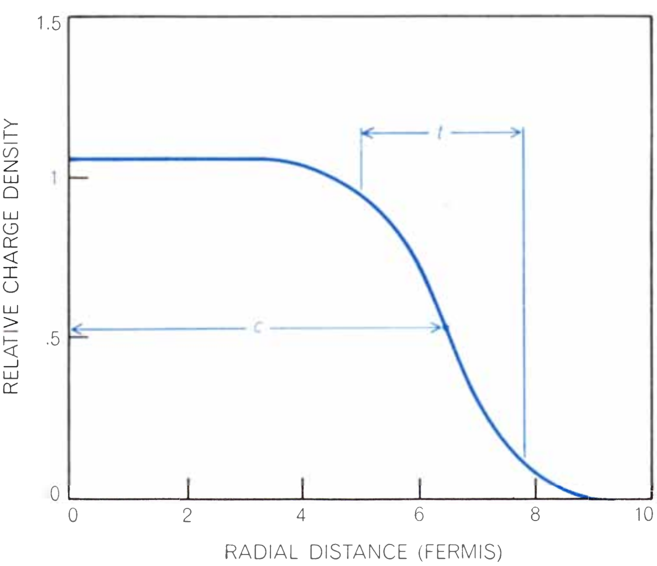
EXCITED STATE of an "even-even" hard deformed nucleus yields a different kind of time exposure. In a low-lying excited state the nucleus is rotating about an axis perpendicular to the axis of the spheroid. When the nucleus is oriented by an external field, the axis of rotation wobbles slightly about a preferred direction (*left*). Time exposure of the nucleus in this state is thus an oblate spheroid the axis of which is the preferred direction (*right*).



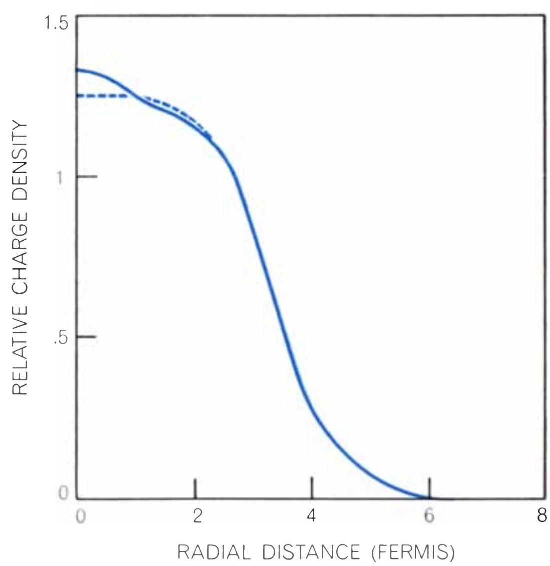
ELECTRON-SCATTERING EXPERIMENT uses a beam of high-energy electrons. The electrons strike the target and are deflected by its nuclei through various angles (*arrows*). The spectrometer counts the number of electrons scattered at each angle. Bending magnets determine the electrons' energy to see how much energy, if any, has been lost to the target nuclei in inelastic scattering.

ous time exposures of nuclei. These do not reveal a true shape but only a time average, like the blurred view of a propeller.

How are we going to perform the snapshot experiment in practice? When applied to nuclei, the snapshot concept is theoretical; although there is nothing in the laws of physics that says a snapshot experiment is impossible, it is far beyond present techniques. One would have to illuminate the nucleus with a beam of radiation of very short wavelength, sufficiently intense to interact simultaneously with every nucleon, and then focus the scattered radiation. None of this will be feasible for a long time. The experiments that can actually be done, and that we shall describe, are indirect. Yet with the help of the hypothetical snapshot definition they lead



CHARGE DISTRIBUTION of a spherical nucleus is characterized by a radius parameter *c*, the distance at which the charge has fallen to half of its central value, and a thickness parameter *t*, within which the density falls from .9 to .1 of its central value. This is a model of the bismuth nucleus, which has 126 neutrons and 83 protons; the number of neutrons is a "magic number," the number of protons nearly magic and therefore the nucleus is spherical.



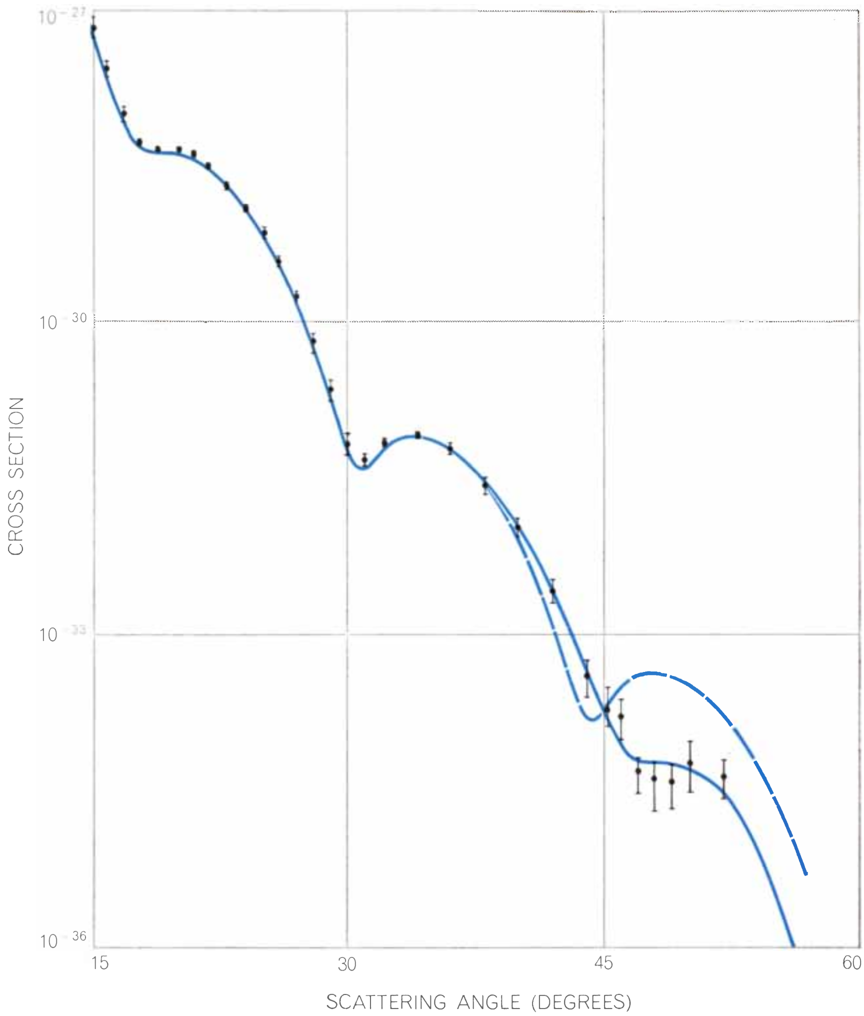
TWO MODELS are shown for calcium 40, another spherical nucleus. The broken portion of the curve is based on low-energy electron-scattering data. The wiggly solid curve in that region is a better fit to high-energy results (*see next illustration*). This illustration and the next are based on experiments by J. B. Bellicard and his colleagues.

unambiguously to a detailed knowledge of nuclear sizes and shapes.

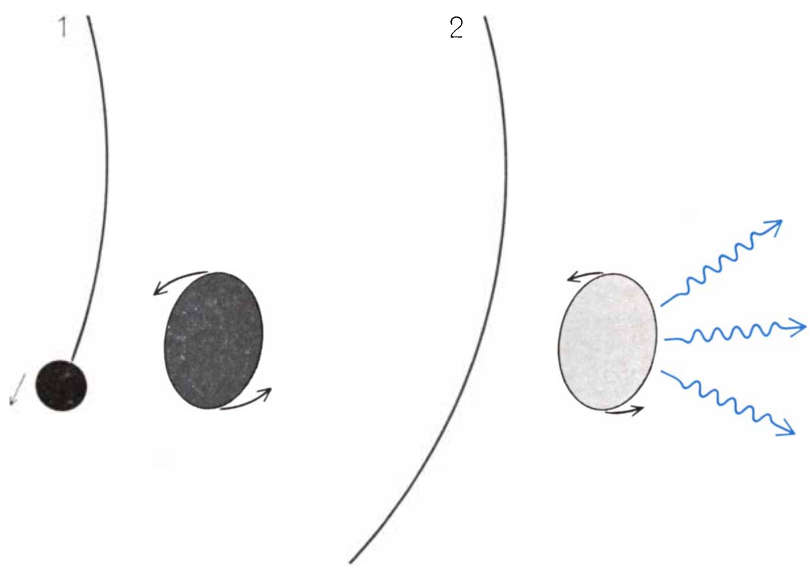
When the snapshot definition is applied to the shapes of atoms rather than nuclei, they are found to be essentially spherical. It is hard to imagine that nuclei could be anything else, but the truth is that few nuclei are spherical. For most nuclei the snapshots, if they could be made, would reveal a more or less ellipsoidal picture, often a bit pear-shaped in addition, with the ratio of longest to shortest diameter lying between 1 and 1.4. The only really spherical nuclei, those whose snapshots would yield consistently spherical pictures, are those whose Z and N are near special "magic numbers" such that the nucleons are able to arrange themselves in a particularly symmetrical configuration. For the majority of nuclei only an occasional snapshot would be spherical, and the interesting questions are: How deformed, and how hard or how soft, is each nucleus? The nonspherical nuclei seem to fall rather neatly into two categories. First there are "hard deformed" nuclei. These have an essentially permanent cigar-like shape: they are prolate spheroids, with one long axis and two equal short axes. Then there are "soft" nuclei, whose shape is highly changeable. Snapshots of many identical soft nuclei would include mostly a variety of asymmetric ellipsoids (in which each of the three axes is of a different length) and a sprinkling of spherical and prolate shapes and oblate spheroids (disks with one short axis and two equal long axes).

The degree of deformation of a nucleus can be defined by a rough formula: To obtain the deformation subtract the smallest diameter from the largest and divide the difference by the average diameter. In these terms the hard deformed nuclei have deformations in the vicinity of .3. The deformations of the soft nuclei fluctuate and are smaller, with typical values around .1.

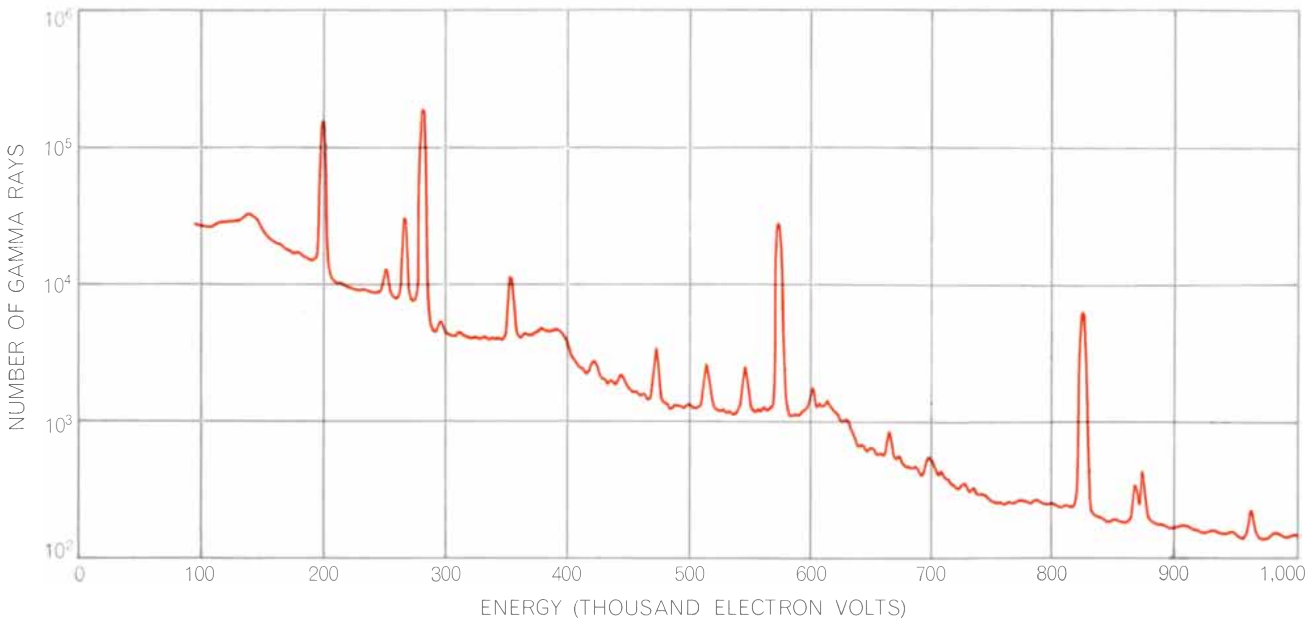
Like an atom, a nucleus can exist in a large number of quantum states [see "The Three Spectroscopies," by Victor F. Weisskopf; SCIENTIFIC AMERICAN, May, 1968]. Each state corresponds to a different mode of motion of the constituent nucleons, and the various states differ in such properties as energy, angular momentum, parity and size and shape. In talking about size and shape one must therefore specify the state of the nucleus. This is usually done in terms of the energy, which is the easiest property to measure. The ground state, or state of lowest energy, is usually the best known because it is the only one that can be stable, but there is much information to



EXPERIMENTAL CROSS SECTION (*points with error bars*) and two theoretical cross sections are shown for 750-MeV electron-scattering by calcium-40 nuclei. (The cross section is in effect the ratio of scattered to incident electrons at each angle.) The data fit the theoretical cross section (*solid curve*) predicted by the wiggly distribution in the preceding illustration better than the one predicted by a smooth distribution (*broken curve*).



INELASTIC SCATTERING, in which projectiles transfer some energy to nuclei, can be accomplished by Coulomb excitation, with alpha particles as projectiles. The electrical force between the alpha and the nucleus sets the nucleus into motion: either vibration or, as here, rotation (*1*). As the nucleus returns to its ground state it emits gamma rays (*2*).



GAMMA RAY ENERGY SPECTRUM is shown for arsenic-75 nuclei excited by alpha-particle bombardment. Some of the peaks are gamma rays emitted by radioactive contaminants but most of them represent energy-level transitions in the nucleus under study and determine its nuclear spectrum (*see next illustration*). The curve is based on results obtained by R. L. Robinson and colleagues.

be obtained by raising the energy to an excited level.

Indeed, any experiment that reveals the true shape of a nucleus must necessarily be one that throws the nucleus out of its original energy level. This follows from Heisenberg's uncertainty relation, which says that the duration of an experiment multiplied by the uncertainty in the energy balance is at least equal to Planck's constant. For a snapshot the duration must be small and so the energy uncertainty must be large; in other words, the energy level must be changed. Such an experiment is called inelastic. On the other hand, any experiment in which the energy level of the nucleus stays undisturbed must last a very long time. Therefore such an "elastic" experiment can reveal only a time exposure.

This means that we cannot make two successive snapshots of the same nucleus in the same level. Does this mean that in considering a particular energy level we are confined to making time exposures? No, because we can make as many snapshots (perform as many experiments) as we like with different but identical nuclei in the same level. The statistics of shapes that emerge are characteristic of the level and contain much more information than a time exposure. In particular, depending on whether the shapes turn out to be all the same or not, the nucleus can be described as hard or soft.

In the case of a hard deformed nucleus, all snapshots made in the ground state and the low-lying excited states yield the same intrinsic cigar shape. The time exposures, however, are different. For the ground state of a nucleus with even numbers of protons and of neutrons there is no angular momentum and the axis of the spheroid has an equal probability of being in any direction; the time exposure is therefore spherical [*see illustration on page 58*]. In a low-lying excited state, on the other hand, the nucleus rotates about an axis perpendicular to the axis of the spheroid. With an external electric or magnetic field one can keep the orientation of that axis approximately fixed in space. As a result the time exposure is an oblate spheroid [*see illustration on page 59*].

For a soft nucleus the situation is less predictable. The different snapshots, as we have seen, reveal different shapes. If the level has zero angular momentum and the nucleus therefore has no favored orientation, the time exposure can only be spherical. When there is some angular momentum, the time exposure can be either prolate or oblate depending on the particular nucleus and the particular level.

Experimental Methods

Three general classes of experimental techniques are now being applied to this kind of work. First, one can observe the nucleus with a short-wavelength probe: accelerated electrons or other high-energy particles that are deflected by the nucleus. Second, one can observe it with a long-wavelength probe such as the atom's own electrons. Third, one can excite the nucleus and observe the radiation it emits.

The first method most closely resembles ordinary vision, in which we shine light on an object and see how it casts a shadow or diffuses or reflects the light. An object can be seen, however, only if it is larger than the wavelength of the light; to "see" a nucleus requires very short wavelengths indeed. Short wavelength means high energy, and electromagnetic radiations of such high energy—gamma rays—are difficult to work with. Electron rays, which can be deflected and focused with magnets, make better probes, but they too must be highly energetic: whereas an atom can in effect be seen by an electron microscope with an energy in tens of thousands of electron volts, seeing a nucleus requires electrons of many millions of electron volts (MeV). That calls for a powerful accelerator such as the two-mile linear accelerator at Stanford University, which delivers electrons of 20,000 MeV.

To study a nucleus one directs a beam of high-energy electrons at a target composed of atoms containing the nuclei under investigation. The electrons of the beam are little affected by the electrons in the atoms but they are scattered, or deflected, by the nuclei. A spectrometer records the number of electrons of a certain energy that are scattered through various angles [*see top illustration on page 60*]. In most such experiments one records only those electrons that have been simply scattered by the nucleus without exciting it; this is elastic scattering and, as we have explained, it yields a time exposure of the nucleus. From this one can determine the nuclear size to an accuracy of about 1 percent, the variation of the charge density with radius and even the distribution of magnetism in the nucleus.

The second method, like the first, involves the interaction of an electron with the nucleus, but in this case it is one of the nucleus's own atomic electrons. When an atom is excited by an input of energy (from heat or light, for example) and subsequently "relaxes" back to its ground state, its electrons change their orbits and the atom emits a spectrum of radiation at characteristic wavelengths. The size and shape of the nucleus play a very minor but detectable role in the electronic structure of an atom and therefore in its spectrum, and since atomic spectra can be measured with high precision these "hyperfine" effects can be observed. Since the nucleus itself remains unexcited, this method, like the first, yields a time exposure of the nucleus. Moreover, since the atomic spectral method does not involve a short-wavelength probe, its results are averaged in space as well as in time and it does not give a detailed distribution of charge and magnetism. Study of hyperfine splitting and shifting of optical spectral lines nevertheless has yielded many measurements of one kind of average value for the radius, the root-mean-square charge radius, as well as of electric and magnetic moments that carry information about the shape of the nucleus.

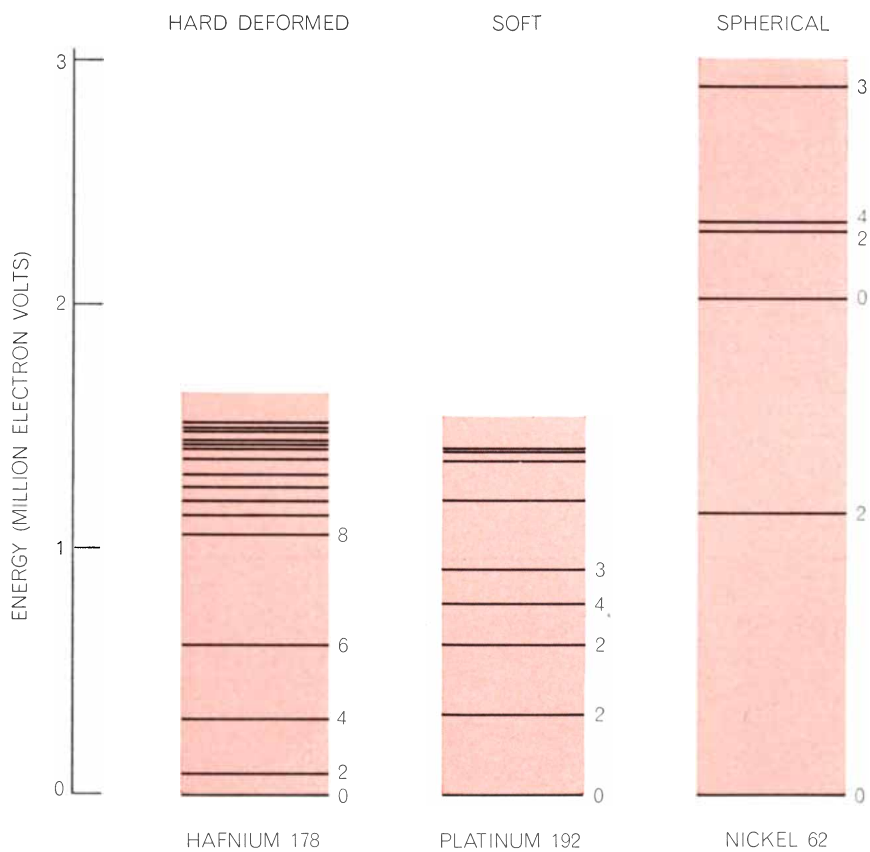
Some recent experiments of this type depend on X-ray spectra rather than optical spectra. Whereas optical spectra are generated by the outer electrons, X rays are emitted when one of the inner electrons changes its orbit; being closer to the nucleus, such an electron is influenced more by the nucleus and less by the other electrons. In the past few years such X-ray experiments have also been done with mu-mesonic atoms, in which a mu meson, or muon, is introduced into the atom to replace one of the innermost electrons [see "Mesonic Atoms," by Sergio DeBenedetti; SCIENTIFIC AMERICAN, October, 1956]. A muon is 207 times heavier than an electron (but still only a ninth as heavy as a nucleon), and so its innermost orbit lies much closer to the nucleus than any electron's orbit. The nucleus-muon interaction that is reflected in X-ray spectra of muon atoms is therefore quite sensitive to nuclear size, and the radius can be determined to about 1 percent.

The third method, unlike the other two, changes the energy of the nucleus and is capable of giving instantaneous information. One such technique, known as Coulomb excitation, is based on the inelastic scattering of a positively charged projectile such as an alpha particle [*see bottom illustration on page 61*]. The target nucleus is excited into a higher-energy state of internal motion and then emits electromagnetic radiation as it returns to its ground state. The distribution of excitation energy, as well as the spectrum and intensity of the subsequent gamma radiation, give information on the nuclear shape. Since in this case a nucleus is really being observed in motion, more detailed information is obtained about the shape and its changes over time than from the first two (time exposure) methods. The information is space-averaged, however, since the alpha particles, having too low an energy to penetrate the nucleus, behave as a long-wavelength probe.

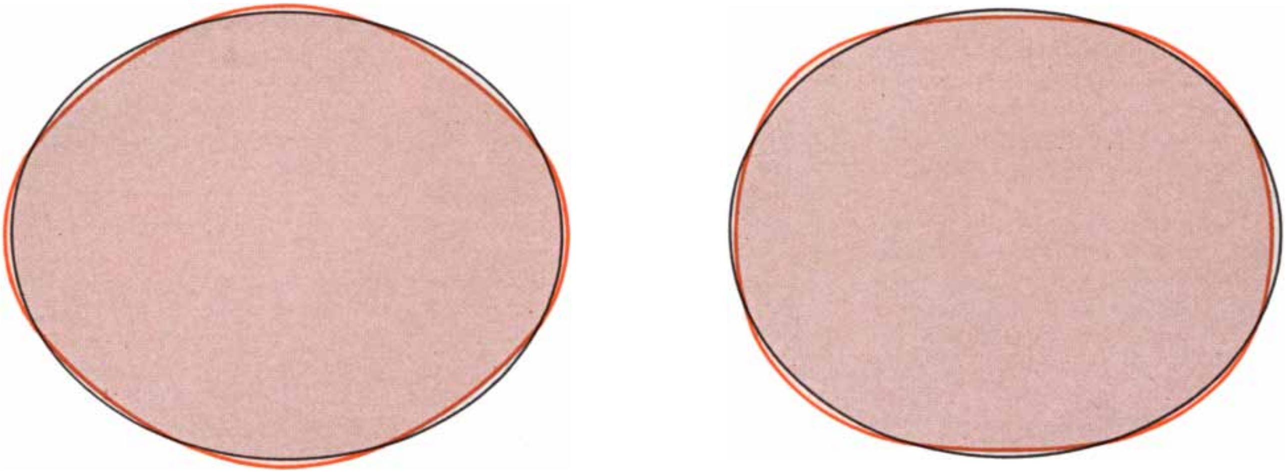
In summary, high-energy elastic scattering gives details of the spatial shape of a time average; low-energy inelastic scattering reveals the time behavior of a space average. A fourth method would be high-energy inelastic scattering, in which one measures the number of scattered electrons that excite nuclei to each energy level. This method can give both detailed space and detailed time information, but the experiments are difficult and their application to the study of nuclei has just begun.

Some Results

The results of time-exposure (elastic) experiments lead to determination of the radius of spherical nuclei and of the spherical image, as it were, of deformed nuclei. The distribution that is revealed, it should be pointed out, is essentially that of the protons. The reason is that the probes (electrons and muons) interact primarily with the charge, and only weakly with the magnetism, of the nucleus—and only the protons are charged. (Neutrons and protons do attract each other through the nuclear force, however, making it likely that their spatial distributions are similar, and some experiments confirm that this is generally true.) What is determined is the charge density as a function of distance from the center. Until recently only two parameters of that function could be determined: *c*, the distance at which the



ENERGY SPECTRA of three nuclei are shown, based on Coulomb excitation and other methods of nuclear spectroscopy. Number to right of level gives angular momentum. Low-lying levels of hard deformed nuclei are related to angular momenta by a precise formula; in soft nuclei the relation is different and less precise. Spherical nuclei lack low levels.



PRECISE SHAPES (but not the detailed charge distribution) are determined by inelastic-scattering experiments. Recently identified "hexadecupole" shape (*color*) is slightly sharper or slightly blunter than a typical "quadrupole" (ellipsoidal) deformed nucleus (*gray*).

charge density has fallen to half of its central value, and t, the surface thickness, which is usually defined as the radial distance from the point where the density is .9 of its central value to the point where it has dropped to .1 of the central value [*see illustration at bottom left on page 60*].

The results of the electron-scattering and the muon X-ray experiments show that, except for the lightest nuclei, the surface thickness is about the same for all spherical nuclei: approximately 2.5 fermis. (A fermi is 10^{-13} centimeter.) The radius parameter c, on the other hand, increases with increasing nuclear mass. In fact, the nuclear volume per nucleon is nearly constant for all nuclei, and the radius measurement c is about equal in fermis to 1.1 times the cube root of the mass number A. Absolute radius measurements can be made in favorable cases to an accuracy of about 1 percent and in many instances it is possible to determine differences between radii to an even higher accuracy. By direct comparison changes in radius between two isotopes or between two energy levels of the same nucleus have been determined to better than one part in 10,000. In other words, radius differences of less than .01 fermi have been detected!

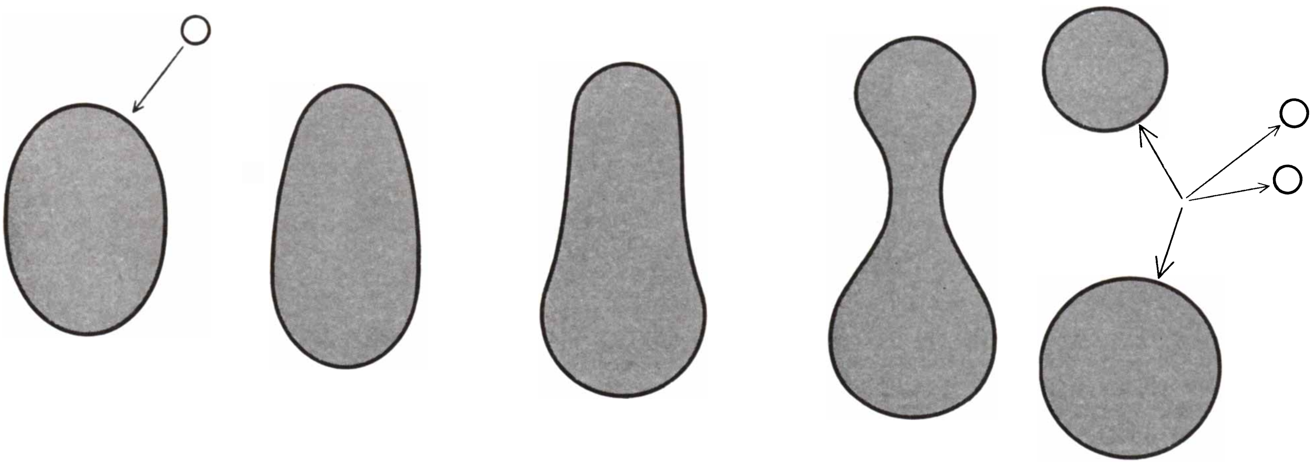
Three results of these difference measurements may be noted for their interesting implications. First, of two nuclei with the same mass number, the one with the more protons and the fewer neutrons has the larger charge radius. This suggests that if a few nucleons are added to a nucleus, they tend to accumulate near the nuclear surface.

Second, hard deformed nuclei are found to have larger radii than soft ones with about the same A. (We are talking about time-exposure experiments here, and the spherical effects of deformation.) A hard deformed nucleus appears as a fuzzy sphere with both c and t larger than for neighboring soft nuclei; the latter, having smaller deformations, look much like spherical nuclei in these time exposures.

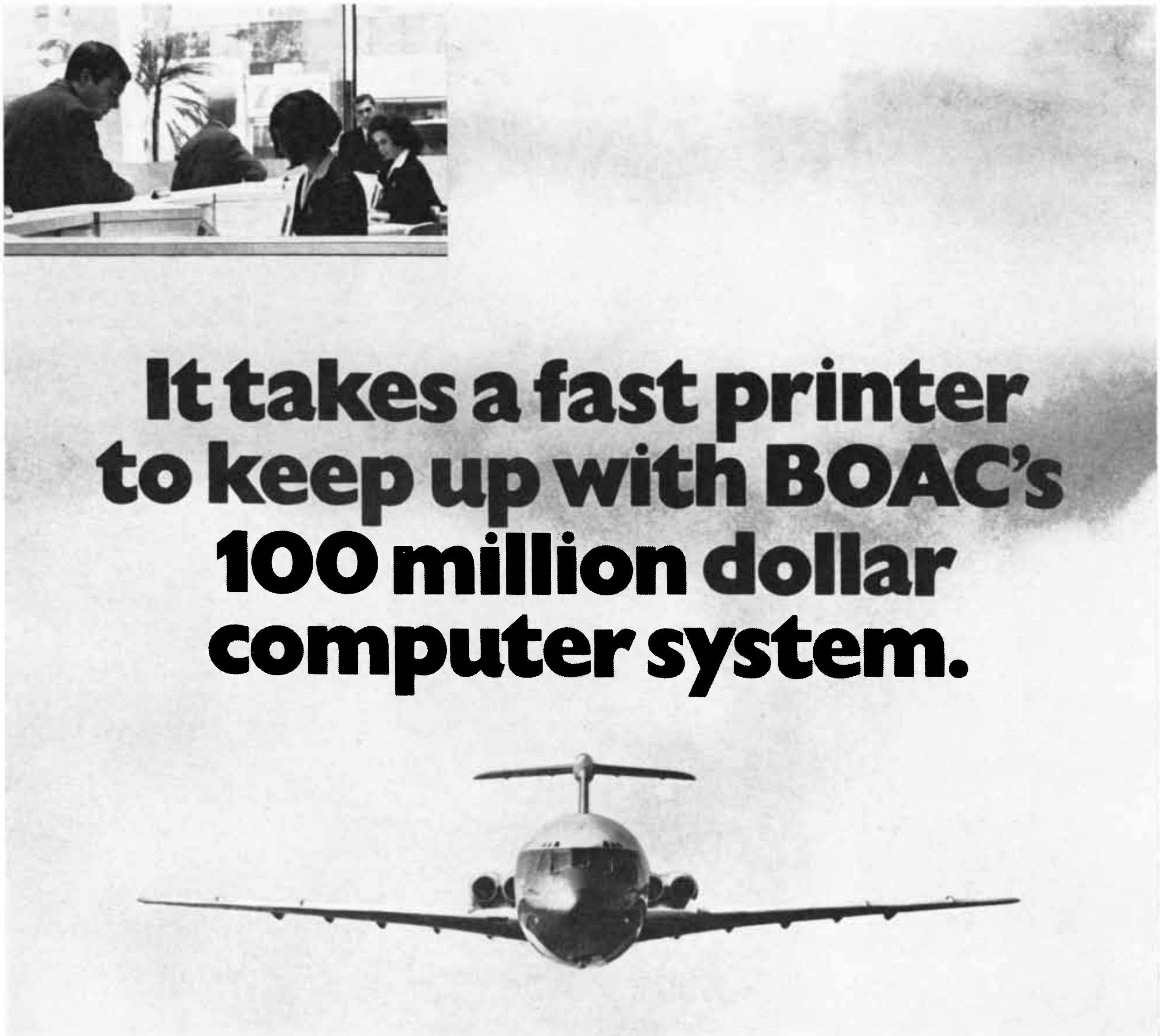
Third, nuclei of odd mass (for example nuclei with an even Z and an odd N) are a little smaller than the average of the neighboring even-even nuclei. The explanation seems to be that the even-even ones are somewhat more deformed.

The advent of higher-energy electron accelerators and correspondingly shorter-wavelength electrons has made it possible in the past few years to determine not only c and t but also some finer details of the time-exposed nuclear charge distribution. In spherical nuclei, it now appears, the charge density does not fall off smoothly with distance from the center. Instead it fluctuates [*see illustration at bottom right on page 60*]. This slight wiggle is direct evidence for the existence of nucleon shells.

The results of inelastic-scattering experiments provide more direct information on the shape of nuclei. When a hard deformed nucleus is excited into rotation



NUCLEAR FISSION occurs when a uranium or plutonium nucleus absorbs a neutron. The added energy causes the hard deformed nucleus to become more deformed and finally break into unequal pieces that repel each other, emitting energy and free neutrons.



Kleinschmidt.

The amazing Kleinschmidt 311™ Teleprinter is part of BOAC's 100 million dollar computer system called Boadicea. This system puts all the BOAC offices in the United States right next door to the home office in London.

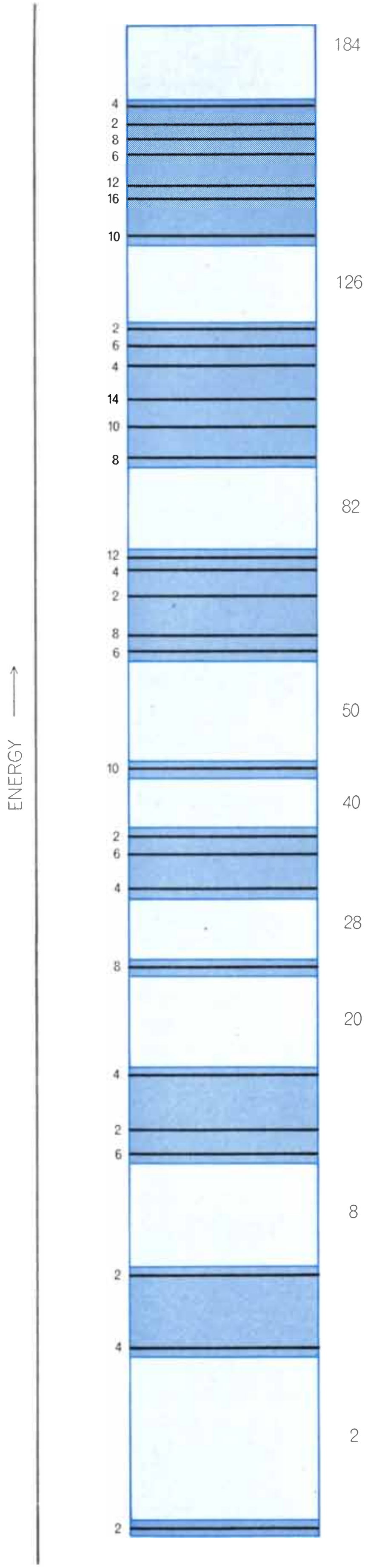
For example, if the BOAC office in Miami needs a passenger list for any BOAC flight, Kleinschmidt gets the answer from London in seconds. At 400 words a minute.

The Kleinschmidt 311 Teleprinter gives BOAC the ultimate in speed and reliability. Furthermore, the 311 Teleprinter can accept parallel information on demand at an average speed of 40 characters a second, which eliminates the

formatting problems of serial printers.

What Kleinschmidt is doing for BOAC it can do for any business that requires quick reception of the printed word. If you would like to learn more, please write Kleinschmidt Division of SCM Corporation, Deerfield, Ill. 60015. Or call us at 312-945-1000.

SCM®



"MAGIC NUMBERS" are derived from the shell model of the nucleus. A nucleon can assume certain energy levels (*horizontal lines*), each containing a certain number of orbits (*small numbers*). The levels are grouped in shells (*dark color*) separated by energy gaps. A magic number (*right*) is the number of orbits below a gap (28 and 40, with small gaps, are semimagic). A magic number of nucleons means a closed shell.

in such an experiment, its spectrum of excited states can be inferred from its gamma radiation [*see illustration on page 63*]. The kinetic energy of each state is found to be related to its angular momentum in exactly the same way as for a rotating top; this is the main observation leading to the conclusion that such a nucleus is hard. In addition to energy measurements, one can, in a careful experiment, observe how many of the nuclei are excited to each level and what radiation is emitted as the rotating nucleus slows down. From these data the detailed snapshot shape can be determined with great precision [*see top illustration on page 64*].

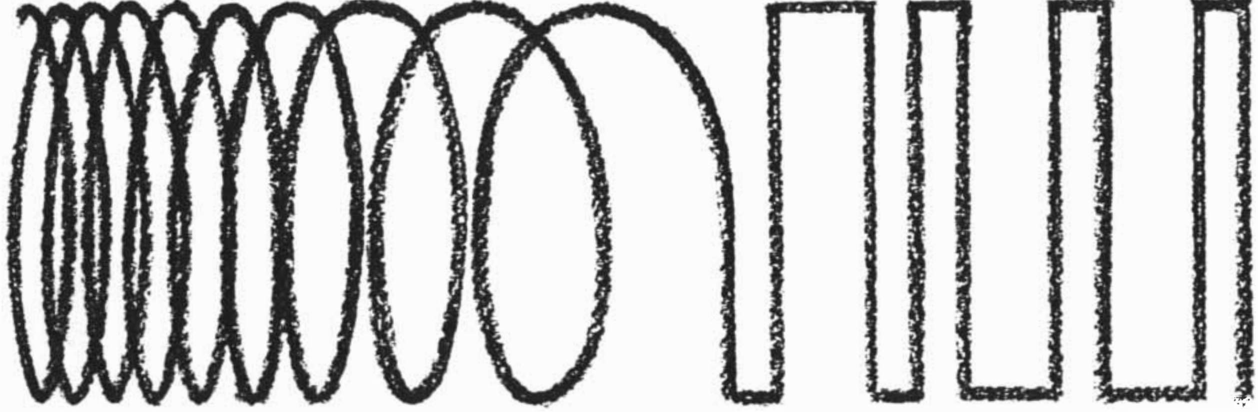
The soft nuclei can rotate too, but they also oscillate through different shapes. A simple kind of oscillation is one in which a nucleus is prolate, becomes spherical, goes on to the oblate shape and back through spherical to prolate once again. More complicated oscillations are also possible. These time-dependent shapes can be inferred from measurements of the spectrum of excited-state energies, which deviate from the simple relation established for the hard deformed nuclei. Measurements of the electromagnetic radiation emitted or absorbed when the oscillation and rotation slow down or speed up are also helpful. Finally, the time-averaged shapes of different excited states are different from what they would be for hard deformed nuclei, and their measurement has just begun.

The ability of nuclei to oscillate and change their shape is of the utmost importance for nuclear fission. In a nuclear reactor uranium or plutonium nuclei, which are ordinarily hard deformed nuclei, are set in motion by the absorption of a neutron. This motion is a change in shape: the nucleus becomes more and more deformed into a cigar shape, until finally it necks down near the middle and breaks into two pieces [*see bottom illustration on page 64*]. All of this takes place with the introduction of the very small amount of energy brought in by the neutron, but the two positively charged pieces repel each other strongly and fly apart, producing a large amount of energy. As the two pieces break apart, a few neutrons are freed, which serve to cause other such fission events, sustaining the energy output of the reactor [see "Nuclear Fission," by R. B. Leachman; SCIENTIFIC AMERICAN, August, 1965]. One of the interesting discoveries of the past several years is that heavy deformed nuclei sometimes have not one but two shapes, one more deformed than the other and both of them nearly stable. It had been observed that some nuclei did not fission at once when excited but waited a measurable time. A number of experiments now show quite convincingly that such nuclei are first knocked into a more deformed shape that is nearly stable but finally elongate and fission.

Fact and Theory

How can these results on nuclear sizes and shapes be understood on the basis of deeper principles? Can we start from a knowledge of the laws governing the motion and interactions of the individual nucleons and account for the foregoing findings? To a large extent we can, although many details are still not amenable to analysis. The difficulties one encounters when trying to connect fact and theory come from two sources. One is that the law of interaction of individual nucleons, the nuclear force, is incompletely known and understood. In this respect nuclei differ from atoms and molecules, for whose components the firmly established law of interaction is Coulomb's law of electricity: The force between two particles varies directly with the charge of each and inversely with the square of their separation. Our knowledge of the nuclear force comes from experiments involving only a few nucleons (usually two), and so far these experiments have not determined the interaction with nearly the precision of Coulomb's law. The other difficulty is the "many-body" aspect of the problem. Although physics has the reputation of being an exact science, most physical problems are mathematically intractable unless approximations are made. This is true even of celestial mechanics, which started physics on the road to exactness: as soon as three bodies are present rather than two, no exact analytical solution can be found. In general when it comes to dealing with many equivalent bodies, no progress can be made without drastic approximations.

There are various approximate views, or "models," of nuclei. Different models may be appropriate for different aspects



We uncoil the data transmitter

In data transmission, as in almost every other field of electronic engineering, the trend is towards micro-miniaturization. The objective is to produce cheaper, more reliable circuits. To transmit data signals on telephone channels, modulators and filters are required.

Unfortunately, when designers tried to miniaturize these circuits they met an apparently insoluble problem. Coils, which with capacitors determine the frequency response of a filter, are extremely difficult to reduce in size. So the designers examined ways of developing a filter without coils.

Dr. Leuthold of the ETH, Zurich, at that time working in our laboratories, evolved the theory of what is now known as the binary transversal filter. This consists of a resistor matrix connected to tappings on a flip-flop shift register. With a binary input signal it acts like a filter. Whilst not exactly equivalent to an LC filter, it can be made to act in a similar way in the desired frequency range. To do this, we make the shift frequency, which samples and shifts the input signal through the register, equal to a high multiple of the input bit frequency.

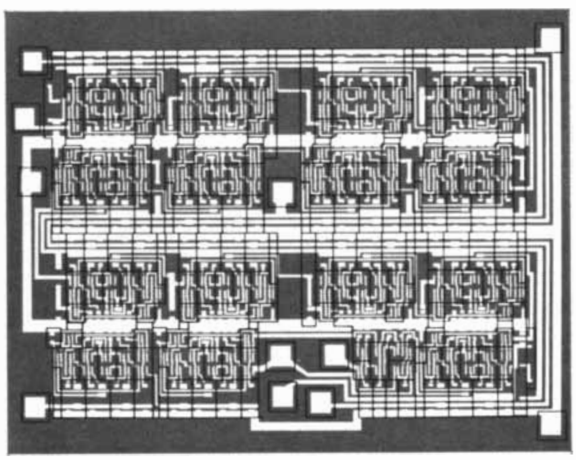
The new digital filter solved one problem but generated another - choosing a suitable method of modulation. To limit the spectrum of the modulated signal with a digital filter, the modulator output must take the form of a synchronous two level signal. This virtually ruled out conventional two-step (up and down) modulation, which would require analog filtering again.

The alternative, direct modulation on a carrier within the signal band, results in "fold-over" distortion. And, using square wave modulation, to other undesirable modulation products.

Nevertheless Mr. Van Gerwen, of our laboratories, decided to adopt the technique. He modulated the binary signals on a square wave carrier and got a mixture of the sidebands of all the odd carrier harmonics. Seemingly a complete mess. Yet, by making the carrier a multiple of half the bit frequency, both wanted and unwanted components occur at the same frequencies in the spectrum. Surprisingly, the unwanted components could be filtered out by adjusting the values of resistors in the filter matrix.

The new data transmitter, comprising flip-flops, logical AND-gates and resistors, worked and was, in principle, suitable for integration. Mr. Schmitz and Mr. Dijkmans of our laboratories, took up this challenge and concentrated all 203 transistors and 172 resistors on a single chip 6 mm^2, with resistance ratios accurate to within ± 1%. This chip can be used for 2400 bits per second in a 300 - 3400 Hz voice band channel. With a proportionately higher sampling frequency the same chip can also be applied to high speed data links. It operates at temperatures from -100°C to +100°C.

A similar technique can be used for an integrated data receiver. This needs an additional analog to digital encoder and consequently another chip.



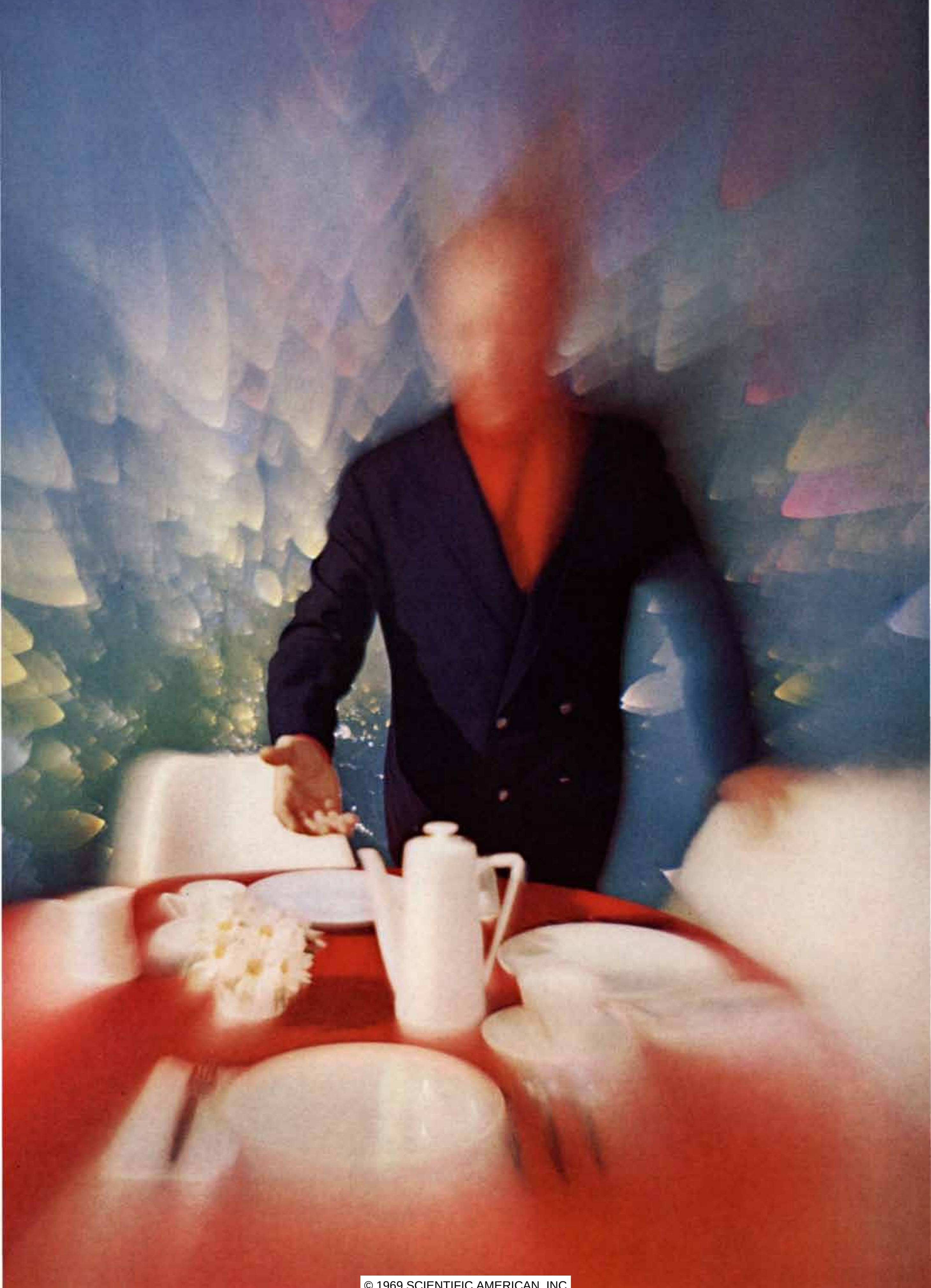
Fully integrated data transmitter - 203 transistors, 172 resistors on 6 mm^2 chip.

In the Research Laboratories of the Philips group of companies, scientists work together in many fields of science. Among these are: Acoustics, Cryogenics, Information Processing, Mechanics, Nuclear Physics, Perception, Solid State, Telecommunications and Television.

PHILIPS

This work was carried out at the Philips Research Laboratories, Eindhoven, the Netherlands.

PHILIPS



of nuclear physics. Much of the effort in theoretical nuclear physics is devoted to demonstrating the validity of each model on the basis of what we know about the nuclear force and about the laws of nucleon motion, which are the laws of quantum mechanics. The question can therefore be stated in this way: How do nuclear models account for empirical evidence on sizes and shapes and how can these models be justified?

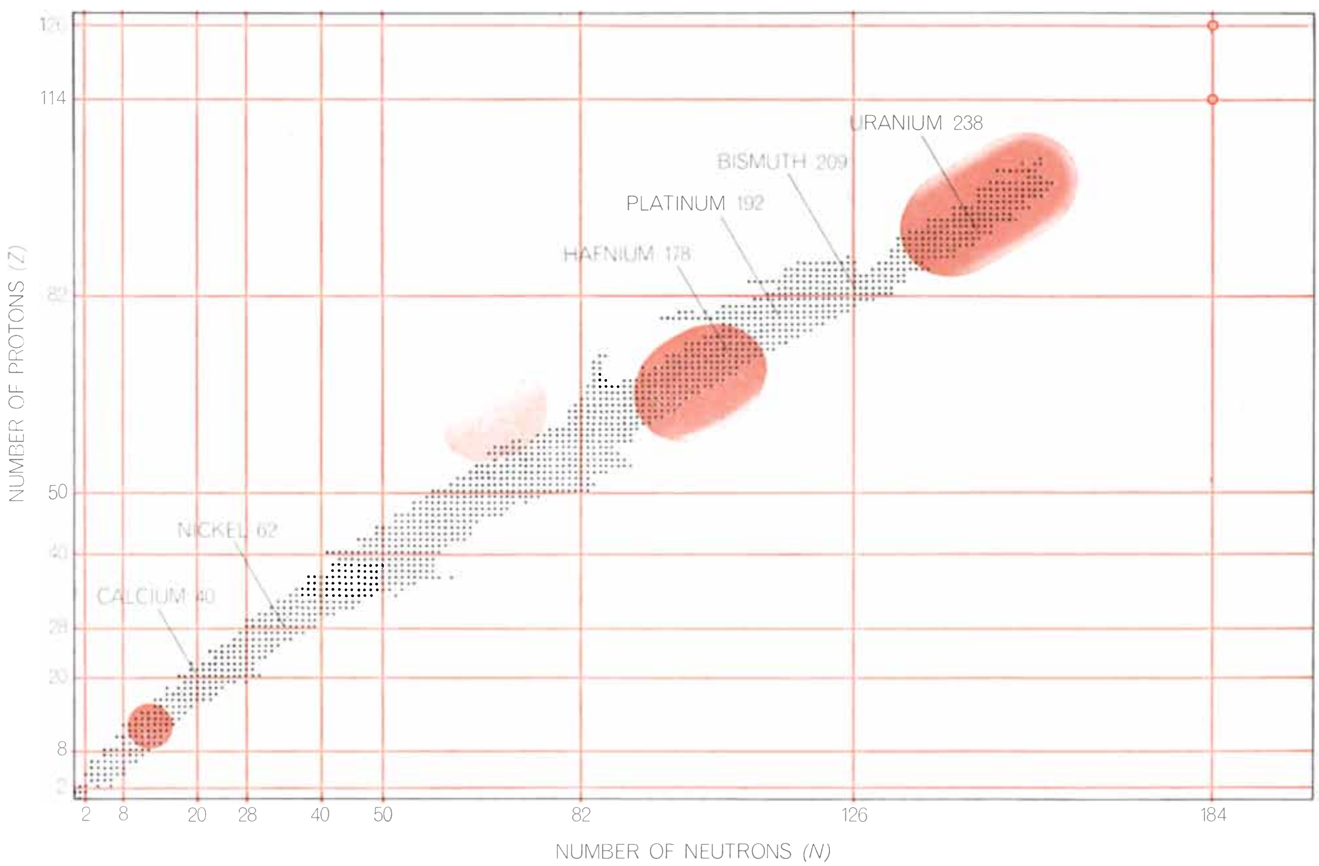
The fact that the volume per nucleon seems to be the same for all nuclei points the way to the simplest nuclear approximation, the liquid-drop model. According to this view, every nucleus is a small drop of a nearly incompressible fluid known as nuclear matter; the volume of the drop is simply proportional to the number of nucleons it contains. A nuclear drop, like one of water, can assume a variety of shapes, and each shape has a surface energy proportional to its surface area. Since for a given volume the shape of minimum area is a sphere, on the liquid-drop model the equilibrium (minimum energy) shape of all nuclei would be a sphere. This does not accord with the facts, and therefore refinements of the liquid-drop model are necessary.

The model does, however, lead to a good qualitative understanding of the fission phenomenon. Here the important factors are the nuclear force and the Coulomb force of electrical repulsion between the protons in the nucleus; if the liquid drop is large enough, the repulsion dominates and pushes the nucleus through the succession of increasingly distorted shapes until it divides. The liquid-drop model also explains why the surface thickness t is about the same for all nuclei: t is a property of nuclear matter and is approximately independent of shape.

In attempting to justify the liquid-drop model one finds that the crucial properties of the nuclear force are that it is short-range and that it is mostly attractive in the outer part of its range and repulsive at very short distances. These properties have been known to nuclear physicists for a long time, at least in a qualitative way. The same properties are true of the forces between molecules of an ordinary liquid, and they are sufficient to account for the liquid-drop model.

In order to reach a quantitative justification, however, it is necessary to introduce another model, a deeper and more sophisticated approximation of the nucleus. This is the shell model. With its help one can calculate the density of nuclear matter from a knowledge of the nuclear force; this yields nuclear radii that agree within a few percent with the experimentally derived radii. The shell model also corrects the inadequacies of the liquid-drop model and accounts for the observed variety of nuclear shapes. What, then, is the shell model?

The major difference between nuclear matter and an ordinary liquid is that quantum mechanics plays an essential role in nuclear matter. In quantum mechanics, as a consequence of the wave-particle duality, a particle restricted to a finite volume (such as a nucleon inside a nucleus) is limited to a discrete set of orbits, which can be classified according to their energy [*see illustration on page 66*]. These energy levels are in general not equidistant; they occur in shells, or



SHAPE OF NUCLEI is related to their shell structure. Every nucleus that has been observed is plotted here according to the number of its protons (Z) and neutrons (N). Deformation increases with distance from magic numbers (*colored lines*). (Protons, but not neutrons, have a magic number at 114.) Three main areas of hard deformation are shown (*dark color*), as is another area predicted by theory (*light color*). Some nuclei mentioned in the article are indicated; they are spherical or deformed and hard or soft as predicted by their Z and N. Doubly magic "superheavy" nuclei (*open circles*), now being searched for, should be relatively stable.

batches, that are separated by rather wide gaps. The shell model assumes that nucleons moving in their individual orbits behave approximately independently of one another. Moreover, according to the Pauli exclusion principle, only one nucleon of a given charge (neutron or proton) can occupy each orbit. To obtain the lowest-energy configuration of a nucleus the orbits are filled one by one with neutrons and one by one with protons, starting from the bottom in both cases. Most nuclei end up having their last neutron shell and their last proton shell partly filled. If, however, either N or Z is one of the magic numbers, the nucleus has a shell that is completely filled and is called a single-closed-shell nucleus; if both N and Z are magic, it is a double-closed-shell nucleus. The reason for the interest in closed shells is that they provide the nucleus with an extra measure of stability; this is important for the determination of the shape, as we shall see.

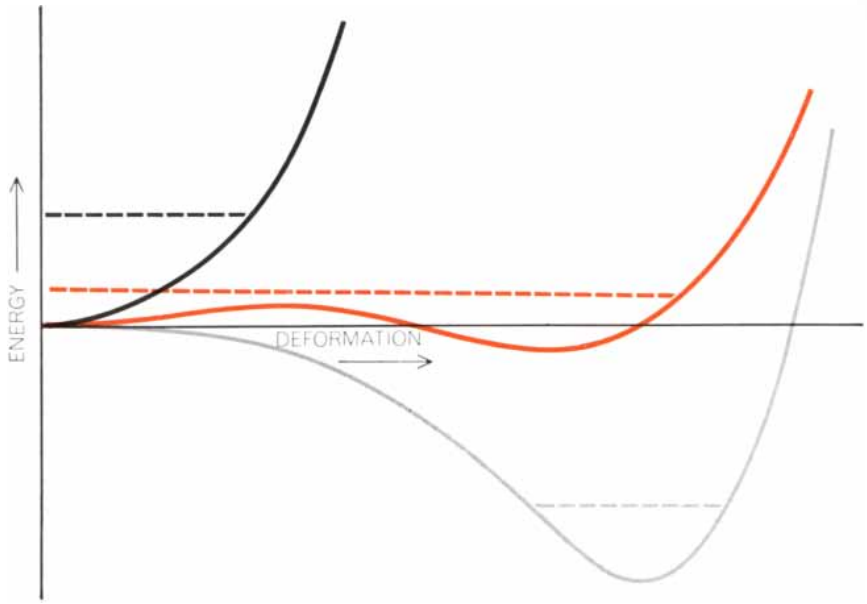
The justification of the shell model is a difficult problem that has received a great deal of attention. It is done through the intermediary of a "single-particle potential," which determines the single-particle orbits and the single-particle levels, and whose calculation is in itself a complicated process requiring knowledge of the nuclear force, the individual orbits and the individual energy levels. The entire procedure is a vicious circle through which one must run many times until a consistent solution emerges. Corrections to the shell model can also be worked out, with the hope that with enough corrections the exact situation can be approached. It is not yet clear that this process of successive approximations will converge well enough to yield a true quantitative description of nuclear phenomena, as opposed to the semiquantitative description that has already been achieved, but things are improving all the time.

Deformation Explained

Once the existence of shells has been established it is not too hard to understand the reason for the deformations. Each of the orbits composing a shell tends to favor certain directions over others. When the shell is completely filled with nucleons, all directions are equally favored and the shell is spherically symmetric. Hence double-closed-shell nuclei are spherical. In a partly filled shell, however, the orbits are not distributed evenly among all directions. We can think of the nucleons in a partly filled shell as applying pressure against the walls of the nucleus (the nuclear surface in which they are enclosed) in an asymmetric, or lopsided, way. The net result is to deform the nucleus. According to this argument we would expect the deformation to be roughly equal to the number of nucleons in unfilled shells divided by the total number of nucleons. This formula does indeed give the observed order of magnitude for the deformation.

If deformation is a consequence of partly filled shells, it is not surprising that the most deformed nuclei, the hard deformed ones, are found away from magic numbers [*see illustration on opposite page*]. Double-closed-shell nuclei are spherical and hard. Single-closed-shell nuclei are also spherical, although softer. As one moves away from magic lines the softness increases; then permanent deformation sets in and the softness decreases, until finally, far from magic numbers, we find hard deformed nuclei.

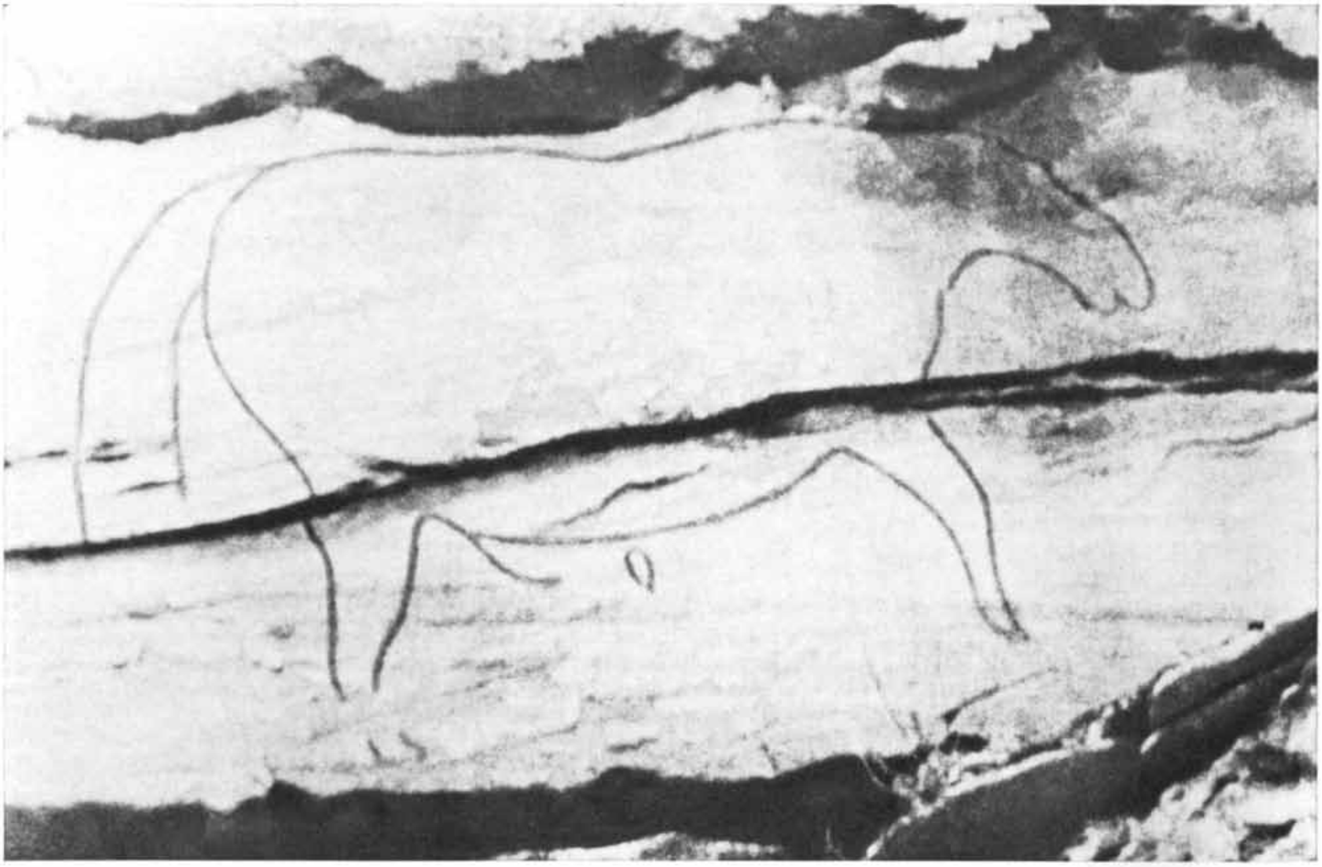
The calculation of permanent deformations in heavy nuclei on the basis of the shell model has been done in a semiquantitative way. The main obstacle to a fully quantitative calculation is its sheer size: it is too large for present-day electronic computers. The difficulty of the calculation is also related to the smallness of the deformation energy, which is only a few thousandths of the total energy of the nucleus; deformation is a very subtle effect.



COLLECTIVE MODEL of the nucleus deals with the energetics of deformation. Here potential energy is plotted against quadrupole deformation for three nuclei. The lowest total (potential plus kinetic) energy that quantum mechanics permits is the ground-state energy (*broken lines*). For a spherical nucleus (*black*) the lowest-potential shape is spherical and not much variation from it is permitted. For a hard deformed nucleus (*gray*) the shape is deformed, again with little variation permitted. For a soft nucleus (*color*) the ground state extends across a wide range of deformation and therefore no shape is particularly favored.

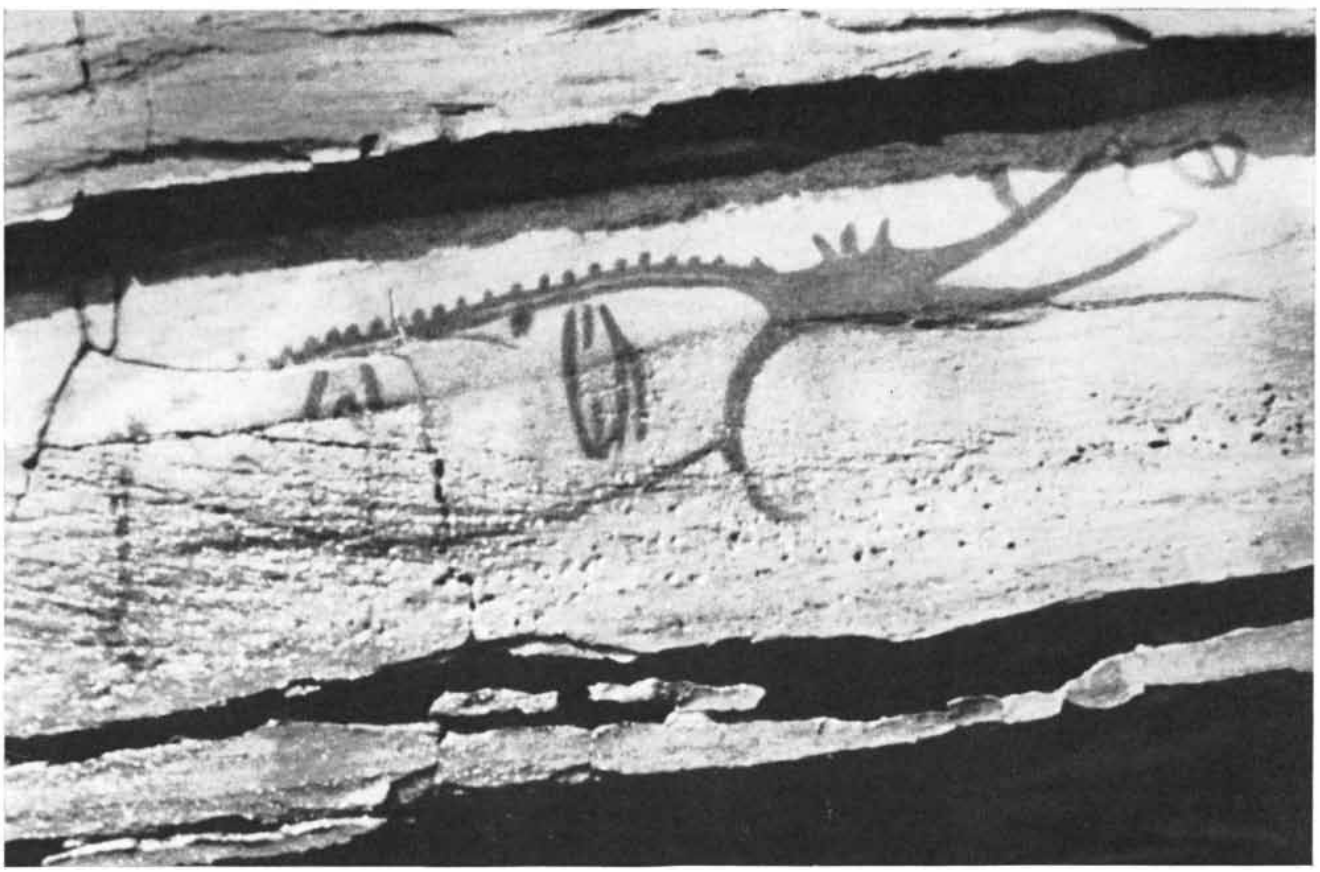
The need for understanding the time-dependence of nuclear shapes leads to the introduction of a third model—the collective model—that is partly a throwback to the liquid-drop model but with the added insight gained from the shell model. In the collective model one attributes a potential energy to every possible shape of the nucleus and a kinetic energy to every possible time variation of this shape. The motion of the nucleus is then reduced to a problem of quantum hydrodynamics. Typical plots of potential energy show how the steepness or flatness of the potential-energy curve determines the hardness or softness of the nucleus [*see illustration above*]. If it is steep, the nucleus departs only slightly from its equilibrium shape. If that shape is deformed, rotation is the most likely type of motion at low excitation energy. If the potential curve is fairly flat, however, the nucleus executes the slow oscillations typical of soft nuclei. To be sure, hardness and softness are matters of degree, and all intermediate stages are possible.

Since the collective model of the atomic nucleus seems to account nicely for many phenomena, the task of justifying it must be faced. This runs up against problems of principle as well as enormous problems of calculation. At the moment this justification is in a rather primitive stage and much work remains to be done.



PAINTED HORSE (*above*) on a rock near the Siberian village of Shishkino on the Lena River, a site rich in petroglyphs, resembles the horses in cave drawings of the Old Stone Age of France.

COSMIC MONSTER (*below*), also at the Shishkino site, exemplifies a theme of rock drawings in the age of metals. The monster is apparently trying to swallow the sun or moon (*circle at upper right*).



THE PETROGLYPHS OF SIBERIA

These rock drawings along Siberian rivers span a period of perhaps 20,000 years. They indicate connections among ancient peoples living as far apart as France and Japan

by A. P. Okladnikov

In the 17th and 18th centuries archaeologists from Russia and western Europe began to encounter surprising relics of ancient cultures in what had long been considered the barren wastes of Siberia. Old burial mounds on the steppes of southern Siberia were found to contain strikingly beautiful objects wrought in gold and sometimes even ornamented with precious stones. In the Altai Mountains diggers into prehistoric graves came on sophisticated tapestries and carvings in wood and leather [see "Frozen Tombs of the Scythians," by M. I. Artamonov; SCIENTIFIC AMERICAN, May, 1965]. These objects of art and fantasy, now assembled in the State Hermitage Museum in Leningrad, have attracted world attention. Many investigators of man's prehistory, however, have been even more intrigued by another series of discoveries, beginning about the same time, that has received less notice but is yielding a most interesting account of man's early life in the northern sector of the Eurasian land mass.

These discoveries are ancient rock drawings that have been found at river sites over the length and breadth of the northern continent all the way from Scandinavia to the Amur River basin in extreme eastern Siberia. Called *pisanitsy* by the Siberians, the drawings are generally painted or stamped in red mineral paint or raddle (red ocher) on rocks along riverbanks. They span thousands of years. The early drawings are mainly of animals; later ones include symbols and stylized figures of men.

The rock drawings first came to the notice of scholars in the middle of the 17th century. About 1645 a Russian chronicler reported that he had seen paintings of "wild animals," "cattle," birds and various other figures on rocks along the Tom River in the Altai highlands. In 1675 Nikolai Spafary, a Russian envoy on his way to China, discovered rock drawings at the Yenisei River. Repeated reports such as these from travelers aroused the interest of a number of German and Russian investigators, including Johann Georg Gmelin of the University of Tübingen; G. F. Miller, who is called "the father of Siberian history"; G. I. Spassky and I. T. Savenkov, an archaeologist who spent years studying rock drawings along the Yenisei and wrote a monumental book on the subject.

In recent years a number of scholars, including myself, have studied the Siberian rock drawings intensively and learned much about them and the people who created them. Among other things, we have sought to determine the age of the drawings and to interpret what they have to tell about the Siberian tribes of prehistoric times. The drawings show changes of style and subject over the centuries that suggest a pattern of evolution in the preoccupations and way of life of early man. Furthermore, they indicate that there was communication or commonalty of some kind among the nomadic northern tribes over the wide region from the Baltic to the Far East.

The earliest of the known Siberian rock drawings appear to be three red paintings that were found near the old Russian village of Shishkino on the Lena River. Two portray wild horses and the third a bison bull. In style they are like the Aurignacian cave drawings of the Old Stone Age of France. The Shishkino drawings resemble the Paleolithic art of western Europe not only in style and content but also in motif, particularly in emphasis on symbols of fertility. Just as the Aurignacian drawings are commonly marked by representations of the vulva, so does a huge drawing of a wild stallion at Shishkino have as a conspicuous feature an exaggerated erection of the phallus.

As the ice age waned and the land changed, so did the animals depicted in the rock drawings. At Shishkino, superposed on the figure of a bison of the Upper Paleolithic period, there is a graceful drawing of a reindeer stag, probably drawn (in a different technique) in the Mesolithic period that followed. Later, in the Neolithic period, the elk became the dominant figure in Siberian rock drawings. As the taiga (dense, wet forest) replaced the Pleistocene steppes and tundra, animal life changed radically. The elk became the forest hunters' main source of food and a prominent object in their religion and rituals.

Monumental figures of elks have been found on the rocks of the Shishkino area in the Lena valley, and other fine drawings of elks were stamped or painted on rocks along the Maya River in Yakutia and on the steep sandstone of the Kamennye islands in the Angara River. (These islands were recently flooded when the river was dammed to create a reservoir for the Bratsk hydroelectric station, but before the flooding nearly 70 drawings were removed and placed in the Irkutsk art museum.) The elks are often portrayed in pairs, one following the other. They are done in realistic detail, with special care lavished on the drawing of the male's magnificent branched antlers and the upper lip of the muzzle—the tastiest morsel of the animal's meat.

It is obviously no accident that the rock paintings were done in red, the universal symbol of blood, which has always been regarded as the carrier of the life-force. For the forest hunters of Siberia (and in the distant west as well) elk-hunting formed the core of livelihood and mythology. The Kamennye rocks depict hunters chasing the elks on skis, and some of the drawings show elks behind vertical red stripes, presumably representing enclosures in which the hunters trapped the animals and killed them with spears or arrows. Some of the peoples of Siberia and northern Europe (the Lapps, the Samoyeds, the Tungus) built ceremonies around the hunting and reproduction of the elk. They pictured the soul of the universe as a gigantic elk or reindeer with gold antlers, pursued through the sky and the nether world by a cosmic hunter.

In the Neolithic period the rock painters began to introduce many human figures instead of being almost exclusively preoccupied with animals. One significant development was that, whereas in Paleolithic times the female figure had been dominant in art and religion, the Neolithic artists focused mainly on the male hunter. They showed his figure

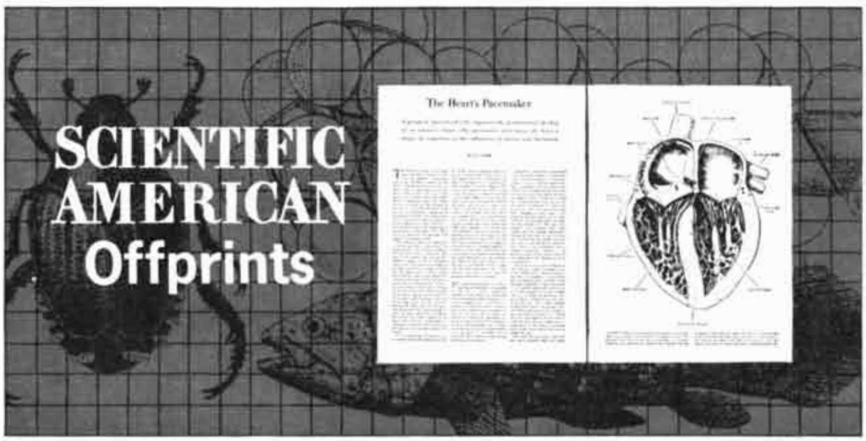
SIBERIAN RIVER SITES of rock drawings (*color*) came to the notice of scholars in the 17th century. The drawings reflect the changes in animal life as the steppes and tundra of the Pleistocene period were replaced by taiga (dense, wet forest). Whereas central

only in profile, with the body flexed in the posture of a skier going downhill. This curious convention, generally adopted not only for rock paintings but also for sculptures in bone, was characteristic of art in the Baltic area (for example Estonia and Karelia) as well as in Siberia. It suggests that by the end of the Neolithic period the peoples of the Siberian and Baltic regions had established connections.

With the advent of the age of metals (the Bronze Age and the early Iron Age) new themes emerged in the Siberian rock drawings. We can discern three different families of drawings, each distinct in style, content and mythology.

In the taiga areas along the Angara and Lena rivers and east of Lake Baikal the rock art of this period featured small human figures and fanciful monsters. The human figures were often depicted with horns and generally had an unmistakably phallic character. A common theme was a schematic picture of a boat with short vertical projections represent-

and western Siberia were inhabited by nomadic forest hunters, in the Amur River region near the sea there were settled societies of farmers and fishermen. The petroglyphs of the Amur region differ in style and content from those of the forest hunters to the west.



ing men standing or seated in it. At Shishkino there is a remarkable frieze depicting seven large boats in line. The boats are meticulously painted in red and are occupied by human figures, no less meticulously executed, who have horns on their heads and tails strangely protruding from their sides. Interestingly, this unusual combination of boats and little horned men with tails has also been found in ancient rock drawings in Scandinavia, which constitute religious monuments left by farmers and cattle herders whose religion was related to an agrarian cult of the Near East and Egypt. The similarity between the drawings in the Shishkino area and those in Scandinavia is not accidental. In many other regions across the continent between the Lena River and Scandinavia there are similar drawings of boats with seated rowers represented by short vertical lines.

A rock drawing near the boat procession at Shishkino shows a mythological monster with enormous toothed jaws that apparently is trying to swallow the sun or the moon, which is represented by a small circle with a stripe across it. Cosmic monsters of this kind were characteristic of the mythology of Turkish and Mongolian cattle herders on the steppes of central Asia. Their epic poetry generally called the monster the "mangus" and described battles in which it was fought by the legendary heroes called "bogatyrs" (mounted warriors).

Rock drawings of the same period in the Yenisei region also feature fantastic creatures. The tribes on the steppes of that area (the Okuner and Karasuk cultures) portrayed the gods of fertility as part animal and part man. They carved huge monuments in the river sandstone, commonly featuring a grotesque version of the human face that has been given the name *lichina,* from an old Russian word for "mask." I shall have more to say about *lichinas* later.

The second family of rock drawings that marked the early metal ages is found on the steppes of Transbaikalia (east of Lake Baikal) and northern Mongolia, some of them near the city of Ulan Bator. The dominant figures of these drawings, recurring with monotonous regularity, are soaring, eagle-like birds with outspread wings. The drawings also portray small, schematic human figures, often seemingly holding hands in a circle as if in a dance. Some drawings show carefully outlined enclosures with round or oval spots or crosses scattered about inside them. The unvarying pattern of birds, small men and enclosures appears to represent a stereotyped form of prayer or ritual that persisted without change.

The eagle apparently was the totem bird of these cattle herders (as it was of the Yakut tribes in the Lena valley). Their mythology was evidently based on worship of the eagle as a divine relative of the sun and celestial elements and on a fertility cult of the horse. The rock drawings presumably reflected the cattle raisers' seasonal fertility celebrations.

The cult of the eagle has survived in that region up to modern times. It was worshiped among the Mongols in the time of Genghis Khan; the eagle was supposed to have made him the leader of the Mongolian tribes. It has remained a central figure in the primitive religion of shamanism up to this century. Stylized soaring eagles with wings in the shape of half-moons are still a characteristic ornament of the Buryat people of Transbaikalia; the design was commonly used as a decoration on stockings and other woolen articles as recently as 20 or 30 years ago. It is also interesting that the Turkic-speaking ancestors of the present-day Turkmenians (living near the Caspian Sea) worshiped the eagle and that to this day the ritual wedding carpets of these people bear figures resembling the birds on the rock drawings of Bronze Age Transbaikalia. Evidently this totemic tradition was brought to Turkmenia from central Asia.

Side by side with the eagle family of rock drawings, geographically speaking, stands the third variety of rock art, which originated in the Bronze Age. These drawings are found mainly on the Selenga River and along the upper Yenisei. Their central subjects are the goat and the maral, or red Siberian reindeer, an impressive creature with long, convoluted antlers stretching over its back. The best collection of rock drawings in this family was found on a small hill called "Little Porcupine" (because from a distance its rock protuberances give it the look of a bristling porcupine) on the left bank of the Selenga River in Transbaikalia not far from the Mongolian border.

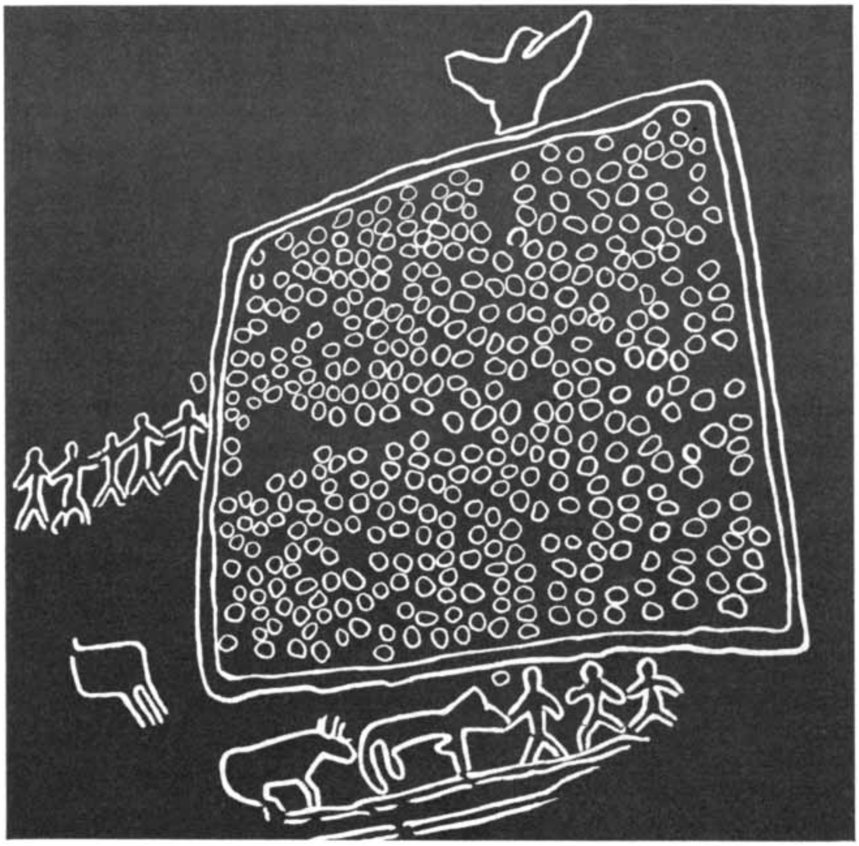
The drawings were made by an unusual method: they were produced by stamping tiny "points" on smooth basalt rocks. The stylized picture of the animal (a goat or a reindeer stag) is generally accompanied by symbols of the sun, giving the composition a cosmic theme. The subjects and style of these drawings are common to many art objects (including stones with reindeer stamped on them) that are found widely throughout central and western Asia. There are rock drawings picturing goats in Kazakhstan, Kirghizia and all over the steppes of central Asia and Mongolia. As an object of veneration the goat has a long history: it endured as a subject of rock drawings for many centuries, became the heraldic symbol of the princely "Blue Turk" families on the Orkhon River of Mongolia and was stamped, along with epitaphs, on the sepulchral shafts of lords of the Turkic tribes.

The rock-drawing episodes of the Iron Age were followed by a new style and theme created in historic times by the Turkic-speaking warriors who overran much of Europe as well as Asia. Not surprisingly, expressions of this art that are identical in spirit and style appear all over central Siberia and as far west as the Danube. The central figure, celebrated in rock drawings, runic inscriptions and epic poetry, is the mounted warrior. These drawings and inscriptions have been found on the walls of Hazara (Turkish) fortresses on the Don River in Russia and on bricks in Preslav, the ancient capital of Bulgaria. Still standing in Bulgaria is a grandiose figure chiseled in rock, called the "Madar Mounted Warrior," that is believed to commemorate Khan Krum, a ninth-century conqueror of the country. The most remarkable Turkic rock drawings were made in the Yenisei, Lena and Angara valleys between the sixth and the ninth century by the Kirghiz and Kurykan peoples. They splendidly depict the nomadic way of life of these warrior horsemen and cattle herders. There are drawings of galloping horses, warriors waving banners, horseback combat, herds of horses and camels, and hunters pursuing goats and elks with lasso or bow and arrow.

The epic poems of the Turkic tribes describe the bogatyr as a heroic figure fighting the forces of evil in the form of monsters or other diabolic creatures. The Russian historian M. P. Gryaznov has shown that in theme and spirit the Turkic way of life and mythology, as described in these sources, are reflected in the Scythian art that was found engraved on gold plates in Siberian tombs.

The long history of rock drawing and its successive periods and styles in cen-

NEOLITHIC MOOSE CALF appears on the sandstone of one of the Kamennye islands in the Angara River. Below the calf is the head of a mature moose (the European elk).



TOTEM BIRD dominates the metal-age petroglyphs of Transbaikalia and northern Mongolia. Another motif is the carefully outlined enclosure with round or oval spots scattered about inside it. The unvarying patterns apparently represent a stereotyped form of ritual.

tral Siberia, from Paleolithic times to the heyday of the Turkic tribes, is most fully depicted at Shishkino. More than 1,000 rock drawings have been found in that area, ranging in date up to fairly recent times following the coming of Russian settlers to the Lena valley. Over the past 1,500 years the rock-drawing art gradually declined in vigor and originality and eventually was abandoned.

In the Amur River region of Far Eastern Siberia we find another style and tradition of rock drawing that in some ways is the most remarkable and revealing of all. It was first reported a century ago by an explorer named R. Maak. Near the site of the present village of Sheremet'yevo on the Ussuri River, a tributary of the Amur, he found a basalt rock covered with drawings that he described in these words: "Here among the things I saw were pictures of a man on a horse, of a bird whose features were mainly those of a goose, of a human face with rays going from it in all directions; they were drawn in very coarse broken lines. However, these historic monuments have left no traditions among the local native inhabitants whom I had the opportunity to question; judging from the present state of the drawings, they were made in very remote times."

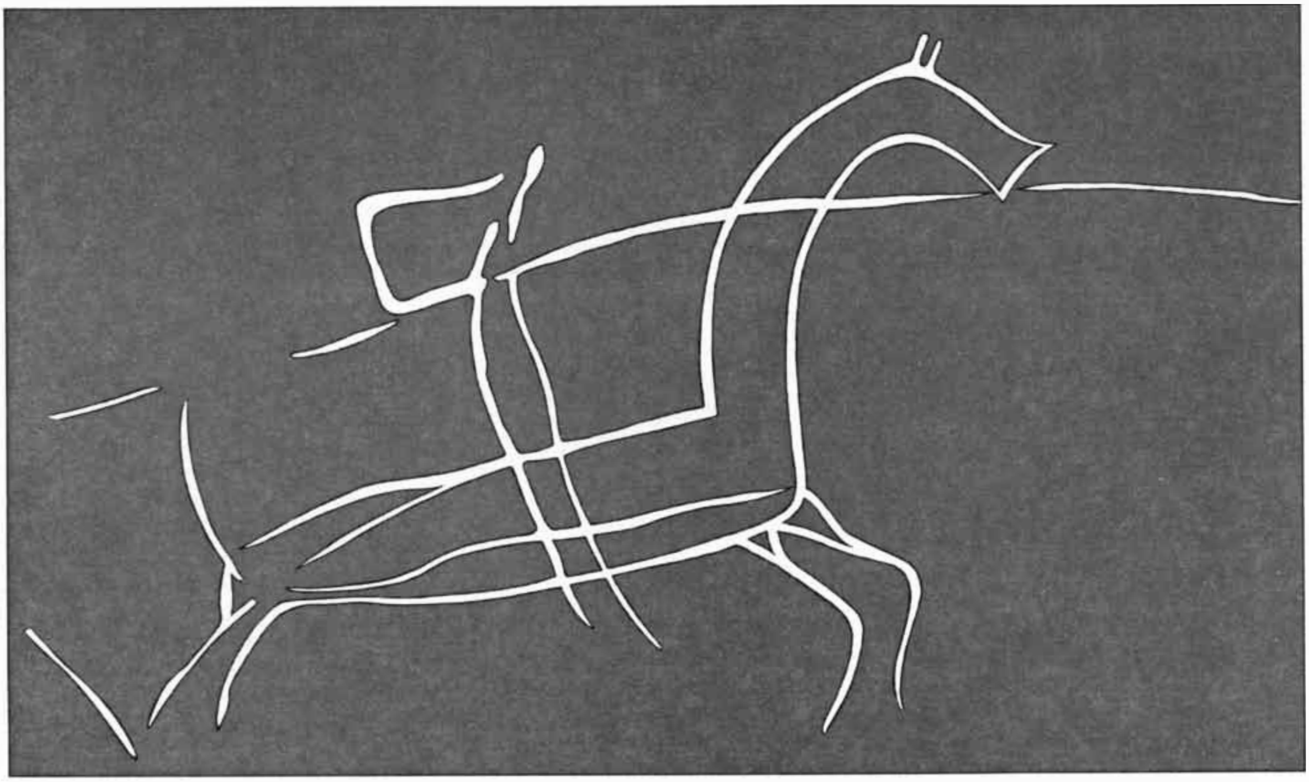
Drawings related to those at Sheremet'yevo were later discovered at several other locations in the Amur-Ussuri area, notably near the village of Sakachi-Alyan, on the Amur 90 miles from Khabarovsk. They have been studied by a number of specialists, including the well-known American orientalist Berthold Laufer, long the curator of anthropology at the Field Museum of Natural History in Chicago. The rock art created by the farmers and fishermen who inhabited this region bordering the ocean tells quite a different story from that produced by the nomadic hunting tribes of central Siberia.

In general the Amur-Ussuri drawings are characterized by a static, anatomical, symbolic quality. Whereas the art of the central Siberian hunters sought to record on the rocks fleeting visions of motion and the dynamics of life, the Far Eastern artists were analytical and given to mysticism. Their favorite motif was a mysterious kind of mask. Unlike the rock painters of central and western Siberia (or indeed of other areas within the Far Eastern region), the Amur-Ussuri artists rarely pictured any part of the human body except the face. Their drawings of masks were rigidly stylized, suggesting a tattoo or ornamental painting rather than a living face [*see top illustration on page 82*]. The masks were obviously freighted with symbolic or religious meaning.

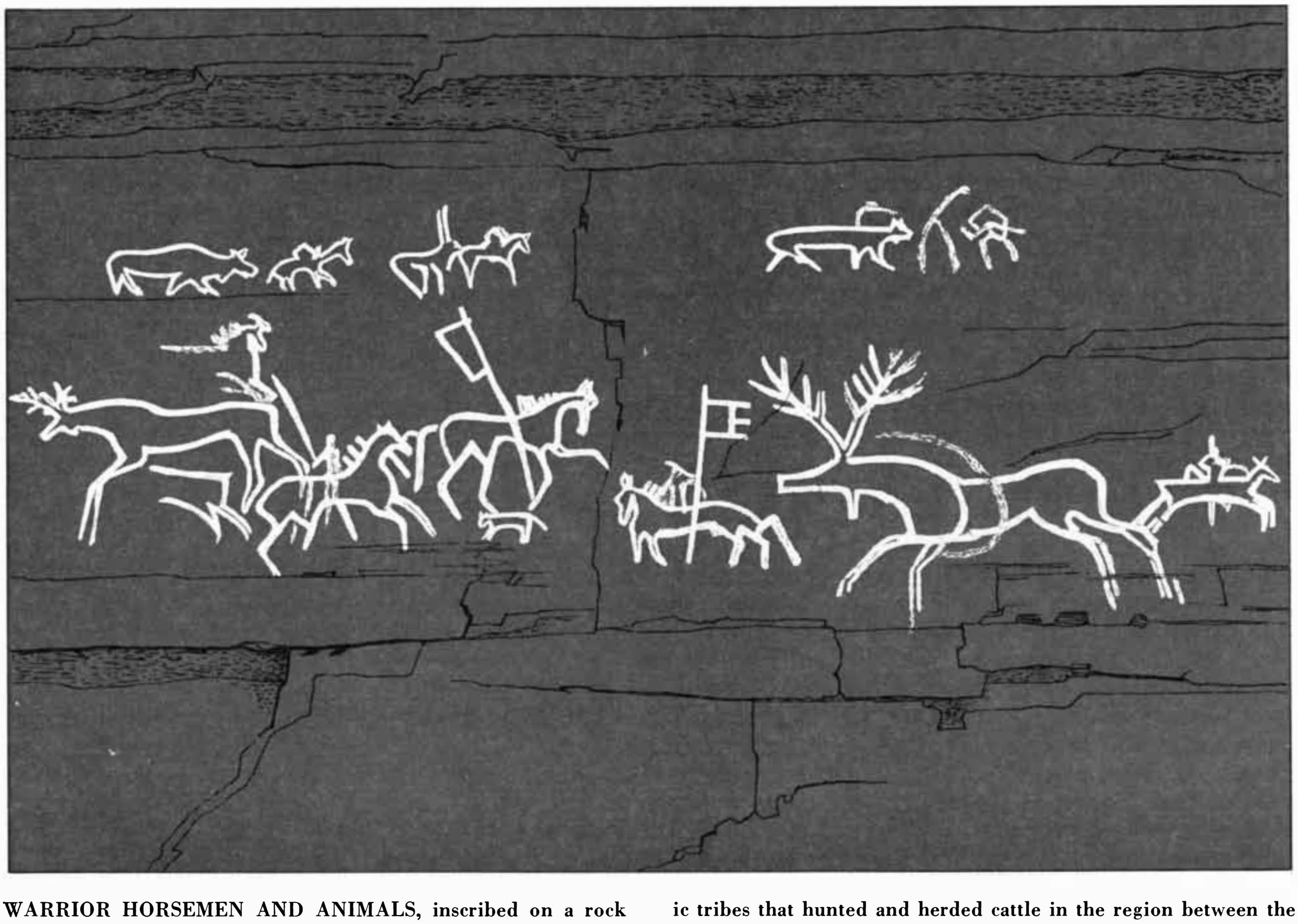
Another distinctive feature of the Amur-Ussuri rock art is that the favorite animal subjects were waterfowl (ducks, geese and swans) and snakes, sometimes entwined in a schematic tangle. (Surprisingly, although these ancient tribes lived mainly by fishing, not one picture of a fish has been found in their rock drawings.) The drawings do include pictures of elks and primitive representations of animals that look like horses. Even the elks, however, are immobile rather than running, and they are loaded with "explanatory" details—conventionalized representations of the animals' anatomic structure and organs. Like contemporary tribes elsewhere in Siberia and in northern Europe, the Amur rock painters also drew stylized pictures of rowers in boats. In at least one case the boat drawing is accompanied by pictures of masks, and in other instances the painting includes arrays of cuplike shapes that may be fertility symbols.

What is the significance of the dominance of the mask motif in the Amur rock drawings? A clue can be found in the common use of masks by ancient tribes in the Pacific islands and in the American Northwest. These masks are known to have been worn in religious rites of initiation into secret male societies. It appears, therefore, that such societies played a large role in the social life of the Amur tribes, as they did (and in many places still do) among the Pacific islanders, the Indians of North America and the tribes of Africa. Interestingly, there is no indication that any such societies existed in the taiga regions of eastern Siberia. This may well be a reflection of the considerable difference in their way of life; in contrast to the settled mode of the Amur valley farmers and fishermen, the nomadic wanderings of the forest hunters prevented the formation of tribal societies.

Masks very similar to those pictured on the Amur rocks have been found, in the form of clay statuettes, on the islands of Japan. They have the spiraled ornamentation that features the Amur masks. These statuettes and other objects with similar ornamentation in Japan date from the Neolithic period, and they made their appearance quite



HORSEMAN PETROGLYPH, near Sakachi-Alyan, exemplifies rock engravings made in historic times by Amur tribes, which are now called the Mokhe. The mounted figure, like the warrior of the Turkic tribes, is a dominant theme in these "modern" petroglyphs.



WARRIOR HORSEMEN AND ANIMALS, inscribed on a rock near Shishkino on the Lena, illustrate the way of life of the Turkic tribes that hunted and herded cattle in the region between the sixth and the ninth century A.D. Some of the horsemen carry flags.

LICHINAS, or masks, engraved on rocks near Sakachi-Alyan in the Amur River region suggest relations between their Neolithic creators and ancient tribes in the Pacific islands and the American Northwest. The faces depicted by the Amur were always rigidly stylized.

AMUR MOOSE AND HUNTER, another Neolithic rock engraving from the Sakachi-Alyan region, shows the animal with its skeletal structure and internal organs represented symbolically. Amur style endures to the present day among Nanai tribes of Far Eastern Siberia.

suddenly in the late stages of the Neolithic culture (later the Jōmōn culture). It is therefore highly probable that they were imports to the islands from the Amur region, where the curvilinear style apparently originated early in the Neolithic era. Archaeological investigations in the past few decades have shown, indeed, that the Amur region was an important cultural center in that era, generating a unique artistic tradition.

The style and the accompanying mythology have endured to this day in the Nanai tribes of Far Eastern Siberia. The art of these tribes is marked by the spiral motif, only slightly changed in form, and by the ancient Amur subjects: masks, snakes, reindeer, elks and waterfowl. Equally remarkable is the fact that the shamanist folklore of the Nanai people still echoes the ancient Amur story of the creation of the universe, which related that the waterfowl played an important role in the creation and that in the beginning three suns filled the sky. There is also a strikingly precise parallel between a present-day Nanai myth and one of the ancient rock drawings on the Amur. The drawing depicts an animal that looks like a horse with a realistic mask of a human head mounted on its back. The Nanai legend tells of a human skull, riding on a horse's back, that turns into a living man and proceeds to perform heroic exploits. The parallel is so close that it is hard to tell which may have come first—the drawing or the Nanai tale.

As time went on, Amur tribes now called the Mokhe created a new style of rock drawing and introduced new subjects, such as riders springing onto speeding mounts, snow leopards, tigers and hunting scenes. Some of these were engraved over the old Neolithic drawings on the same riverbank boulders. The new subjects resembled those we see in rock drawings by Turkic tribes in central Siberia; this is not surprising, since in historic times the Mokhe and Turkic tribes became neighbors in eastern Siberia and even fought together in Korea in wars against the Chinese.

All in all, the drawings left on the rocks of the riverbanks of Siberia give us a sweeping view of the course of human history in that part of the world over a period of at least 15,000 to 20,000 years. They shed unprecedented light on the outlooks, myths, beliefs and aesthetic standards of the peoples who inhabited that huge area. Indeed, the drawings tell us a great deal we could not have learned from any other source, even written history.

Now you can afford your own time-sharing computer: the GE starter set.

Information Systems
Serving Customer Needs - Worldwide

Watch it grow. It's inevitable. Anyone who now buys a little time-sharing, will soon be wanting far more.

General Electric makes your transition a lot easier, and a lot less expensive. With our new GE-410 system you can start out with a small computer (to serve a few people) and grow it into a very big one (to serve hundreds of people).

You can save thousands because you have no extra programming costs. Just change the heart of the computer — give it more power to handle more remote terminals — and you're back in business. All the programs you've already developed will still work — only faster.

You'll also save money because General Electric time-sharing systems do regular processing as well. Essentially, that's two computers for the price of one.

Only General Electric gives you such a wide choice in time-sharing. From a single terminal to the largest, most flexible time-sharing systems yet devised. One is right for you.

No wonder more time-sharing is done on General Electric computers than any other kind. Contact your nearest GE Information Systems Sales Representative. Or write: Section 290-38, General Electric, 1 River Road, Schenectady, N.Y. 12305.

290-38

GENERAL ELECTRIC

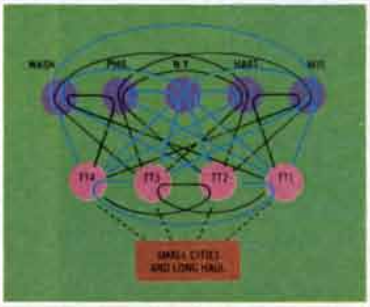
Lockheed has better ways to make better decisions.

Plan for future air travel involves new route concept • Disaster simulator prepares decision makers for crises • Computerized loading model predicts airlift flights per mission • Airline fleet economic analysis evaluates proposed routes

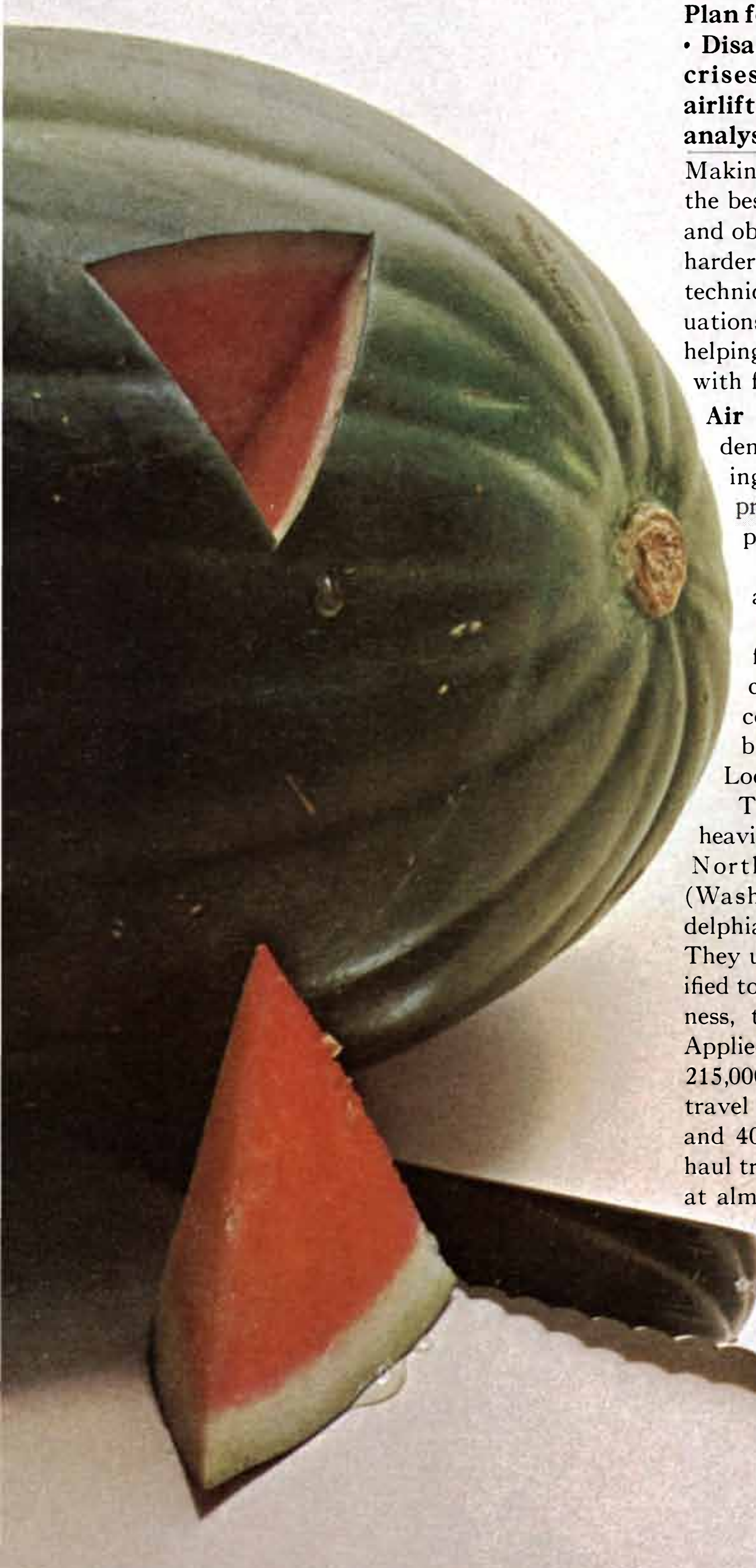
Making a decision involves making a choice. And to make the best decision, all options must be looked at logically and objectively. But the more complex the situation, the harder this becomes. Lockheed, by applying scientific techniques and analytical tools to decision-making situations, is able to approach them more systematically—helping government and business make better decisions with fewer headaches. Some brief examples follow.

Air travel plan for 1980. With air transportation demand increasing geometrically, with airlines ordering more and bigger jets, with airport expansion programs under way, how will all of tomorrow's passengers, planes, and airports best be integrated? Certainly, a more logical and systematic approach than today's is called for. And one such concept was recently developed by analysts at Lockheed.

They chose the heavily populated Northeast Corridor (Washington, Philadelphia, New York, Hartford, Boston) in 1980 for study. They used the well-known MIT demand equation, modified to include origin incomes and destination attractiveness, to determine 1980 intracorridor travel demand. Applied to 46 corridor cities, the model indicates over 215,000 intercity passengers daily. Of these, 55% will travel between major cities, 5% between minor cities and 40% between major cities and minor cities. Long-haul traffic entering and leaving the corridor is projected at almost 400,000 passengers daily, apportioned among cities by population.



Transfer terminal route concept.



LOCKHEED

LOCKHEED AIRCRAFT CORPORATION

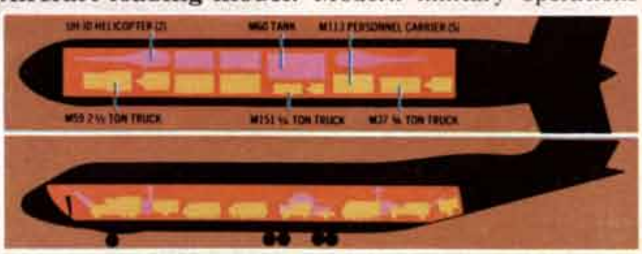
The concept would establish 4 transfer terminals, each situated on low-cost land midway between the 5 major cities. Several intracorridor airline routes (as illustrated) would connect the transfer terminals with each other and with convenient airports and vertiports in and around the major cities, and would connect the major cities with each other. All nonstop. Shuttle service by VTOL or V/STOL would be provided between the transfer terminals and their surrounding small cities. Long-haul passenger traffic would enter and leave the corridor only through the transfer terminals, while long-haul cargo shipments would use major-city airports.

The transfer terminal concept would assure good load factors for the airlines and provide 90% of all passengers—both intracorridor and long-haul—with fast, frequent, nonstop or one-transfer service.

Training for disasters. In disaster situations—floods, fires, riots, earthquakes, even military attacks—quick and often irrevocable decisions must be made. Yet the men whose job it is to make them may never have experienced such situations before.

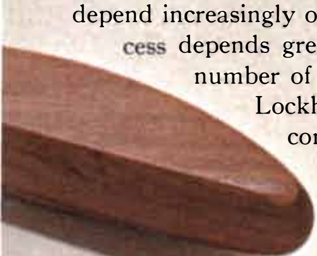
Providing experience in crises is the main purpose of Lockheed's disaster environment simulator. Developed under contract to the Texas Hospital Association, it is a computer system that dynamically simulates the medical effects of disasters. Personnel responsible for managing medical and transportation resources can use this gaming instrument to try out various disaster-control plans for different situations without the real-life consequences. By systematically evaluating the results, effective operational policies can be formulated.

Aircraft loading model. Modern military operations



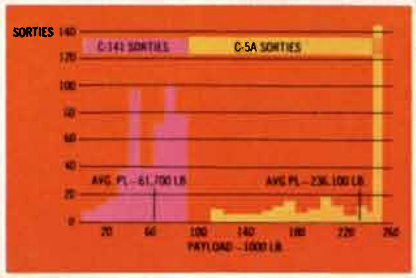
Typical C-5 vehicle load generated by Lockheed's loading model.

depend increasingly on successful airlift. And airlift success depends greatly on accurate estimates of the number of sorties (aircraft loads) needed. So Lockheed analysts have developed a computer model that simulates and thereby predicts the results of aircraft loading.



The loading model can work with any 2 aircraft types (for instance, the C-141 and the C-5) simultaneously and in any ratio. It considers the dimensional and weight limitations of the aircraft, then selects items for placement in cargo compartments by a heuristic procedure that simulates manual load planning. (See 1st illustration.)

The program output shows the vehicles, troops and palletized cargo comprising each load, plus payload and area utilization. A summary gives the payload frequency distribution for each aircraft type. (See 2nd illustration.)



Payload frequency distribution between C-141 and C-5.

In practice, this means that with a given amount and mix of cargo to be airlifted, the number of sorties needed by available transports can be predicted accurately.

Airline fleet planning. Any endeavor as complex as operating an airline, no matter how well managed, can benefit greatly from a more organized method of making decisions. With this in mind, Lockheed has developed a series of computer programs that airlines are using to help analyze their air transportation operations.

At the heart of these programs is an airline system simulation model, whose inputs are coordinated with the airline. The simulation model is then coupled with men who have airline planning experience.

The model is put into motion by balancing 2 sets of factors. One is passengers seeking to satisfy their travel needs from an array of potential services. The other is the desire of the airline to offer service, motivated by a profit objective. Because it simulates a real-world situation, the model also takes into consideration expansion of service needed to meet competition as well as compliance with government regulations.

What comes out is the number of flights required by different types of aircraft to serve an airline's total route system, the operating expenses incurred to offer this level of service, and the revenue generated. Based on this information, flight schedules, the total aircraft requirements, system load factors and utilization, earnings, and discounted cash flow return on investment are optimized.

The activities mentioned here are only a few of Lockheed's current R&D efforts in problem-solving. If you are an engineer or scientist interested in this type of work, Lockheed invites your inquiry. Write: K. R. Kiddoo, Lockheed Aircraft Corporation, Burbank, California 91503. An equal opportunity employer.

CLAW of a gecko lizard was photographed in polarized light in order to show the interference colors that result from the regular alignment of keratin molecules in the claw. Keratins make up the bulk of such epidermal structures in reptiles and mammals.

KERATINS

These proteins are the stuff of hair, feather, horn, claw, nail, quill, hoof, beak, scale and the outer layer of skin. Their intricate molecular architecture is being revealed

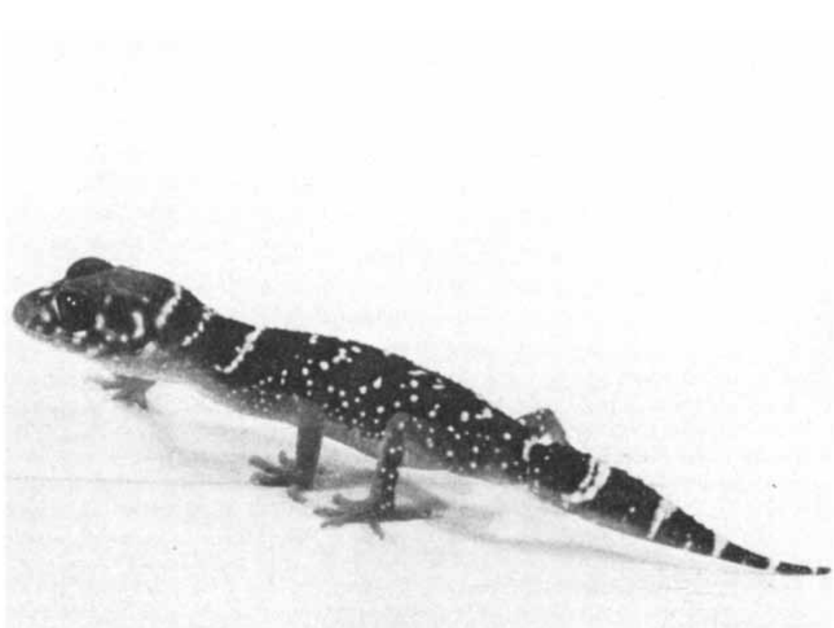
by R. D. B. Fraser

Sometime in the early history of the vertebrates, as they began to emerge from life in the water to life on land, they evolved a specialized protein now known as keratin (from the Greek word for horn). This tough, fibrous, insoluble material provided an outer coat that served to prevent the loss of body fluids. Keratins have since proliferated into a wide variety of substances performing many different functions: the claws and armor of reptiles, the feathers and beaks of birds, the hooves, horns, skin, hair and nails of mammals. The tremendous diversity of appearance among birds and mammals is due not so much to differences in the form of the body as to differences in their keratinous coverings and appendages.

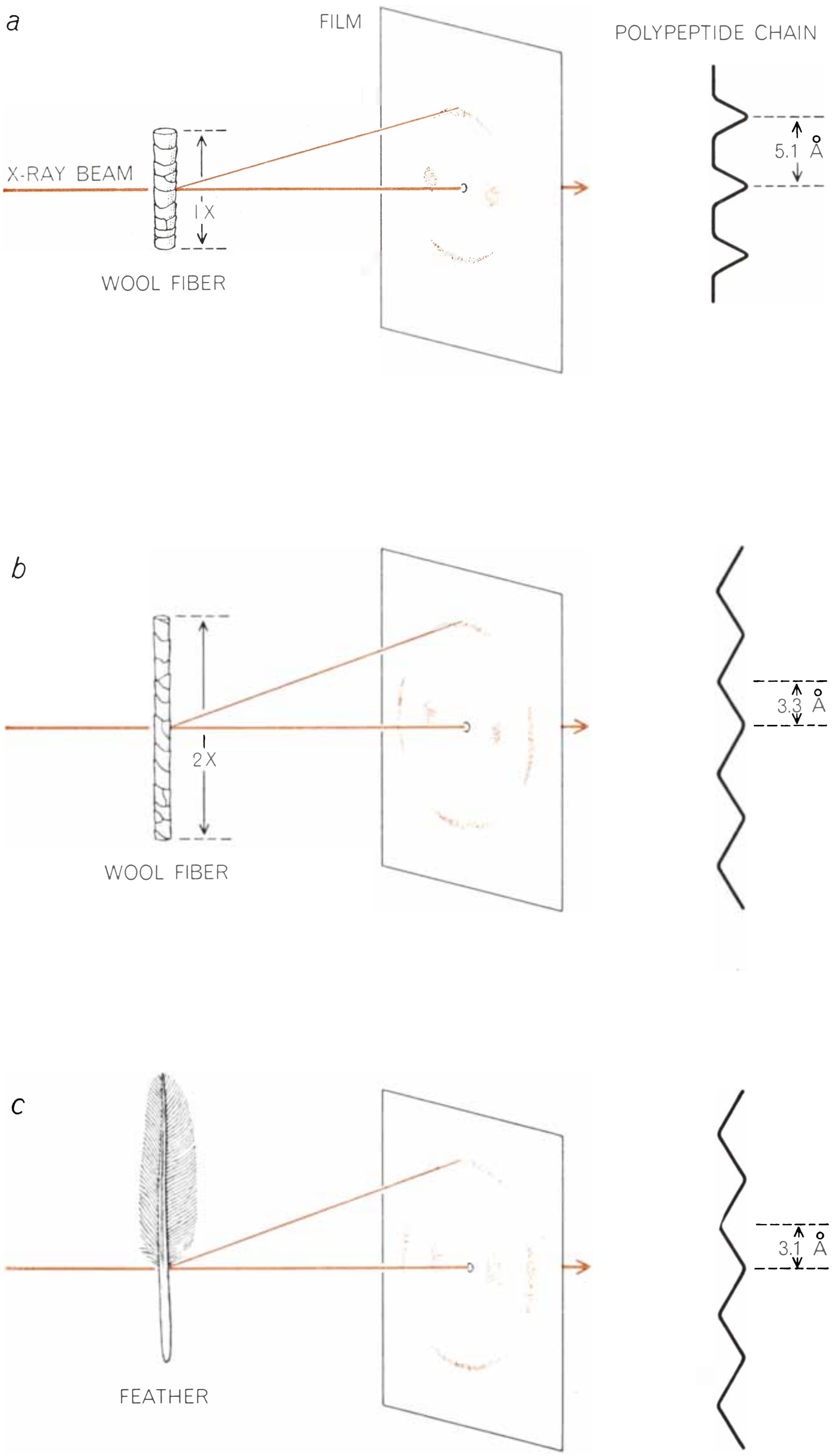
How can a single family of proteins, all built on the same basic architecture, possess such a versatility of physical properties? The keratins have fascinated many investigators, particularly biochemists and physicists. Their long-chain molecular structure lends itself to detailed examination, and keratins have been used as models for investigating the relation between protein structure and protein function. A pioneer in the study of keratins, W. T. Astbury of the University of Leeds, observed prophetically a number of years ago in this magazine: "Their structure analysis will one day demand a great tome for which the present article can hardly form even the preface to the first chapter" [see "Flagella," by W. T. Astbury; SCIENTIFIC AMERICAN, January, 1951]. Over the years some important chapters have been written, among them the historic discovery by Linus Pauling and his associates of the helical structure of protein chains [see "The Structure of Protein Molecules," by Linus Pauling, Robert B. Corey and Roger Hayward; SCIENTIFIC AMERICAN, July, 1954].

Our own investigations of the keratins stem primarily from an interest in the mechanical and chemical properties of wool. Because of its immense economic importance wool has been studied more intensively than other mammalian keratins, and Australia, as the world's major wool producer, has in recent years supported a vigorous program of research into the structure of this material in the laboratories of the Commonwealth Scientific and Industrial Research Organization (C.S.I.R.O.). Through studies of various keratins our research has been addressed to fundamental questions about keratin structure.

Like other proteins, the keratins are polypeptides formed by the linking together of amino acid units to form chains. In the early 1930's Astbury and his co-workers at Leeds made the first important progress toward understanding the special properties of the keratins in X-ray studies of wool fibers. Examining the fibers with a beam of X rays, they obtained diffraction pictures indicating reflection sites spaced at regular intervals along the fiber. The fibers of wool, like hair, are remarkably elastic, and Astbury's group examined them in both the unstretched and the stretched state. It turned out that the interval between sites in the stretched state was markedly different from the interval in the unstretched fiber. Astbury and his associate H. J. Woods deduced that in the unstretched fiber the amino acids were folded together in groups around each site, and that the stretching extended each group so that an individual amino acid occupied each site [*see illus-*



GECKO LIZARD is shown in this photograph. Keratins first evolved in such lower vertebrates to form a tough, insoluble outer coat that served to prevent the loss of body fluids.



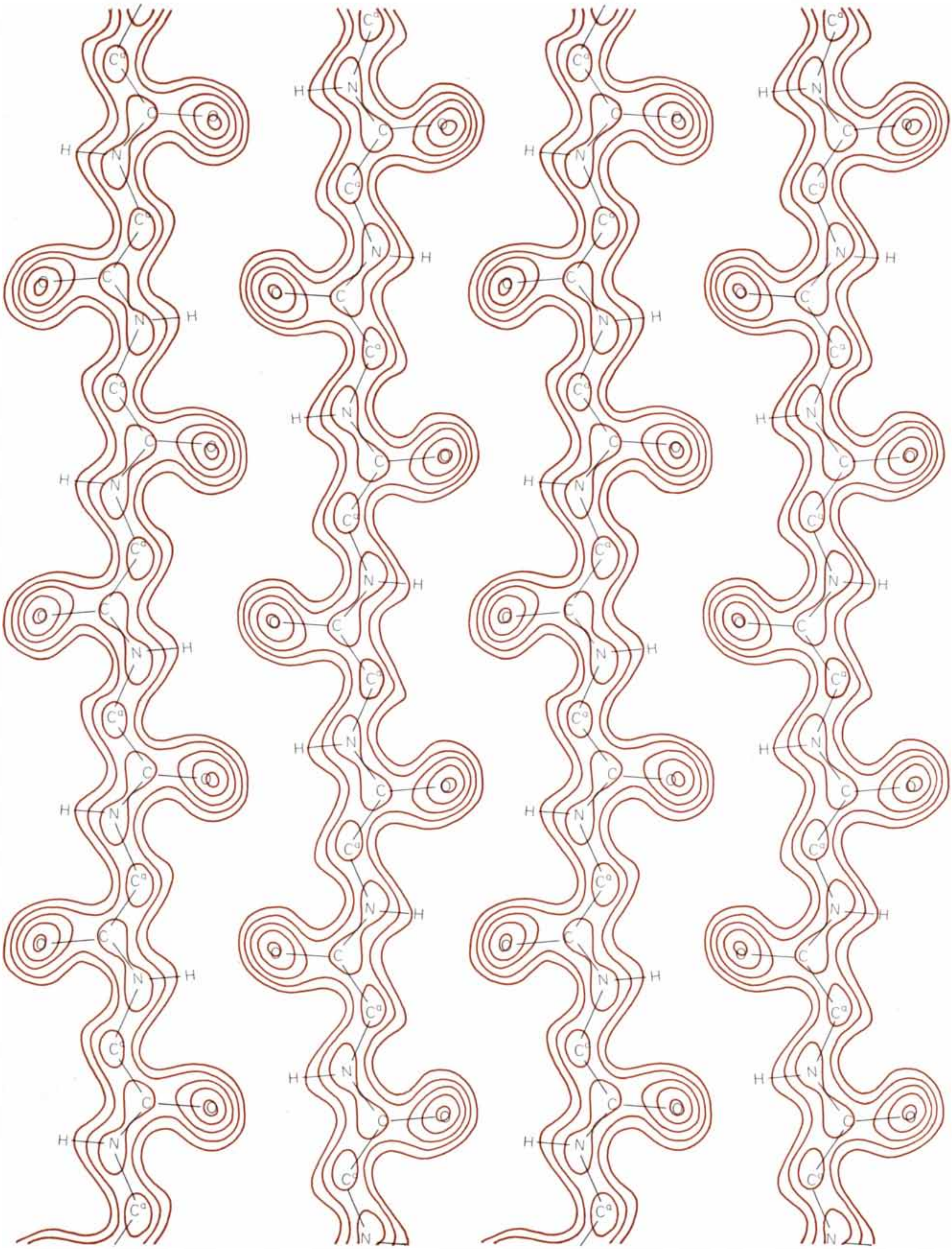
X-RAY DIFFRACTION TECHNIQUE was first used to study keratins in the 1930's by W. T. Astbury and his co-workers at the University of Leeds. The sample was irradiated with a fine monochromatic beam of X rays and the reflected beams arising from regularities in the molecular structure were recorded on film. The pattern from a mammalian keratin such as wool (*a*) was found to have a spacing of 5.1 angstroms in the direction of the fiber axis. This pattern was named the alpha pattern and was later obtained from such diverse materials as muscle and bacterial flagella. When stretched to about twice its normal length in steam, the wool gave a different pattern (*b*), called the beta pattern, which had a spacing of 3.3 angstroms in the direction of the fiber axis. Astbury and H. J. Woods reasoned that the protein chains in wool must be folded in the alpha form and extended in the beta form. Later it was found that in birds and reptiles (*c*) the chains were already extended in the native material. The alpha chain is now believed to be coiled into a slightly distorted helix.

tration at left]. The Leeds investigators named the X-ray picture yielded by unstretched wool the alpha pattern and the pictures yielded by stretched wool the beta pattern. In our laboratory the information contained in the beta pattern has recently been used to obtain an electron-density picture of the extended chains in stretched wool [*see illustration on opposite page*].

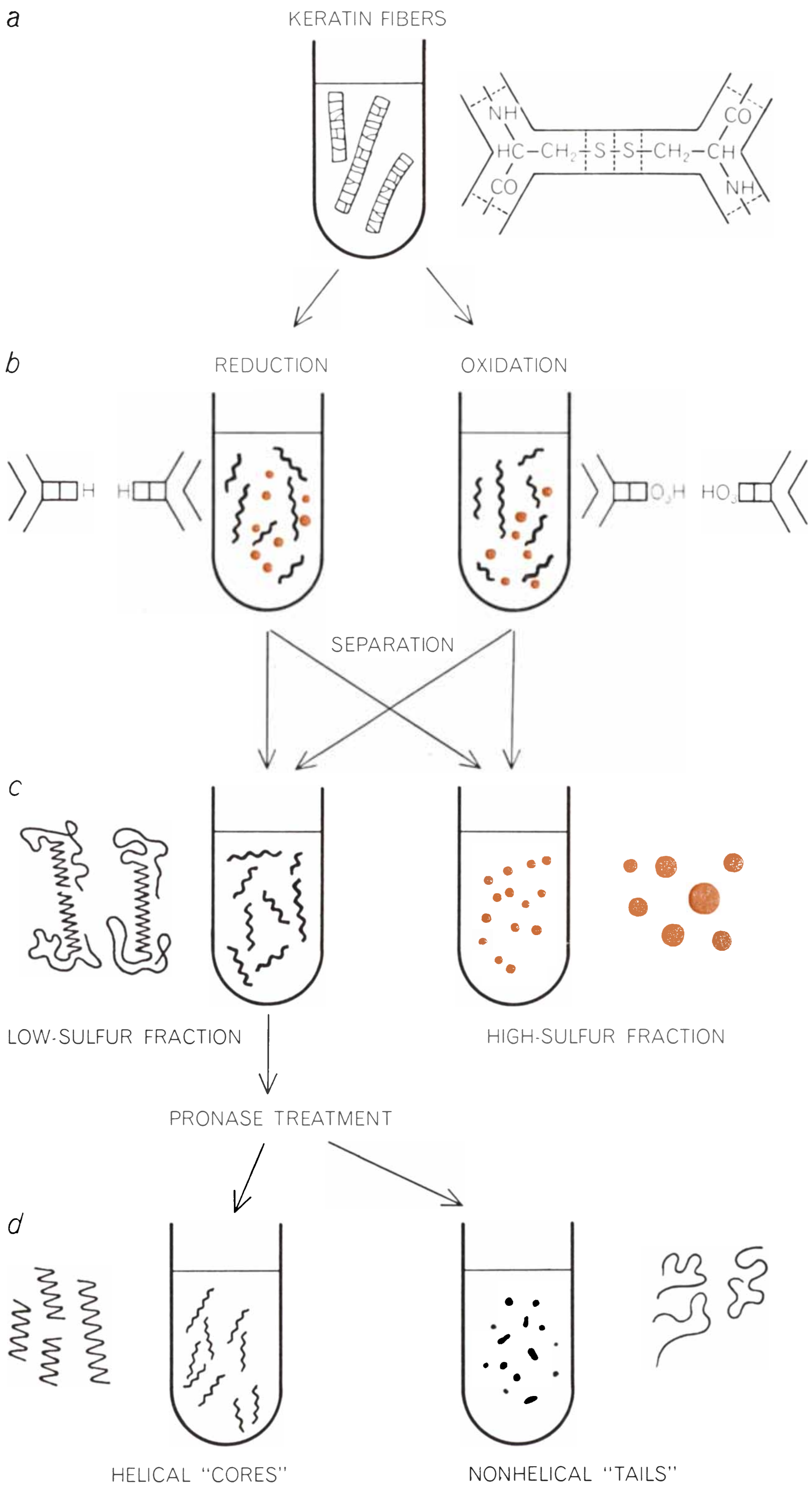
The Leeds group found that the alpha pattern was generally characteristic of mammalian keratins in the normal state (and also of other fibrous proteins, such as those of muscle). On looking into the keratins of birds and reptiles, however, they discovered that in many tissues, notably bird feathers, the keratins exhibit an X-ray pattern in their natural (unstretched) state that resembles the beta pattern of stretched wool. A general survey of the three classes of higher vertebrates made by K. M. Rudall of the Leeds group has shown that birds and reptiles can produce both the alpha and the beta type of keratin but that mammals synthesize only the alpha type.

Over the past four decades much of the investigation of the structure of proteins has centered on studies of the makeup and geometry of the alpha keratins. A big step forward came when David R. Goddard and Leonor Michaelis of the Rockefeller Institute for Medical Research found that keratin fibers owe their insolubility to the fact that they are composed of molecular chains cross-linked three-dimensionally by disulfide bonds [*see illustration on page 90*]. They were able to render the keratins of wool or feathers soluble by breaking the disulfide bonds with reducing agents (adding a hydrogen atom to each of the sulfur atoms at the bonding sites). Later Peter Alexander and his co-workers at the Wolsey Laboratories in England found that the disulfide cross-links could also be broken by oxidation. Once it became possible to study keratin molecules in solution it was learned that the keratin of wool could be separated into two fractions, one relatively rich in sulfur and the other poor in sulfur. The fraction low in sulfur proved to consist of fibrous molecules capable of regeneration into fibers, whereas the sulfur-rich fraction consisted of globular molecules.

Alexander interpreted these findings to mean that a keratin fiber was made up of two kinds of material, one forming the fibrous structure and the other providing the cement that cross-linked it three-dimensionally. By means of the electron microscope E. H. Mercer and



ELECTRON-DENSITY "MAP" (*color*) **was recently obtained from the beta X-ray pattern of stretched wool in the author's laboratory at the Commonwealth Scientific and Industrial Research Organization (C.S.I.R.O.) in Australia. The chainlike keratin molecules alternate in orientation across the page. The peaks corresponding to individual atoms are clearly resolved. This general arrangement of atoms in polypeptide chains formed by the linking of amino acid units was foreseen by Astbury and Woods almost 40 years ago.**

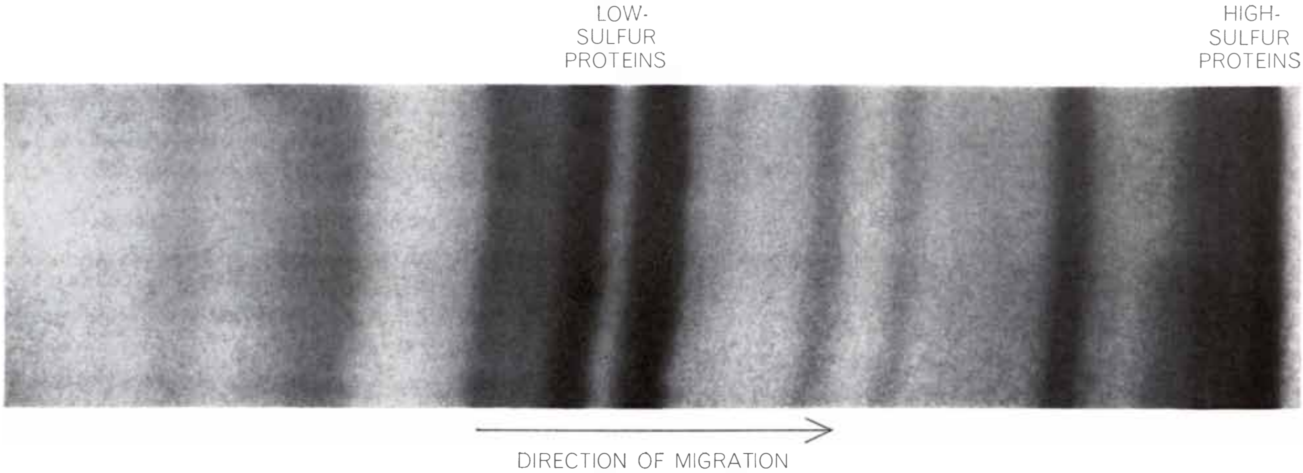


CROSS-LINKING of polypeptide chains in mammalian (alpha) keratin is effected by means of disulfide bonds (*a*). As long as these intermolecular bridges are intact the keratin is insoluble and resistant to attack by enzymes. Treatment with reducing or oxidizing agents (*b*) renders keratin soluble, and two quite different types of protein can then be isolated by a variety of techniques (*c*). The major fraction has a lower sulfur content than the starting material and contains fiber-forming proteins that can be regenerated to give an alpha-type X-ray pattern. The minor fraction is a mixture of globular proteins all containing more sulfur than the original material. The low-sulfur proteins can be further separated by means of an enzyme named Pronase into helical "cores" and nonhelical "tails" (*d*).

M. S. C. Birbeck of the Chester Beatty Research Institute eventually succeeded in finding visual evidence for this theoretical picture. They found that hair roots consist of fine, densely packed filaments (now called microfibrils) set in an amorphous matrix that apparently cross-links them together.

Much work has been done on the analysis of the two main types of keratin proteins: the low-sulfur fraction and the high-sulfur fraction. Like the molecules of many other fibrous proteins, the low-sulfur protein molecules of keratin are coiled in an alpha-helical conformation. Optical studies of these molecules in solution show, however, that their degree of coiling is only about half that of fibrous proteins such as those in muscle. W. G. Crewther and B. S. Harrap in our laboratory recently made an interesting discovery that suggests an explanation. Using a protein-splitting enzyme preparation named Pronase, they split off about half of the amino acid units from the low-sulfur protein molecule. This left a core that more closely resembled muscle proteins in chemical composition and in the high degree of helical coiling. The split-off "tails," on the other hand, were nonhelical and contained most of the sulfur. One can conjecture that these tails may provide the links joining the low-sulfur and high-sulfur fractions of the keratin molecule and perhaps may also be involved in linking the low-sulfur proteins end to end to form long chains.

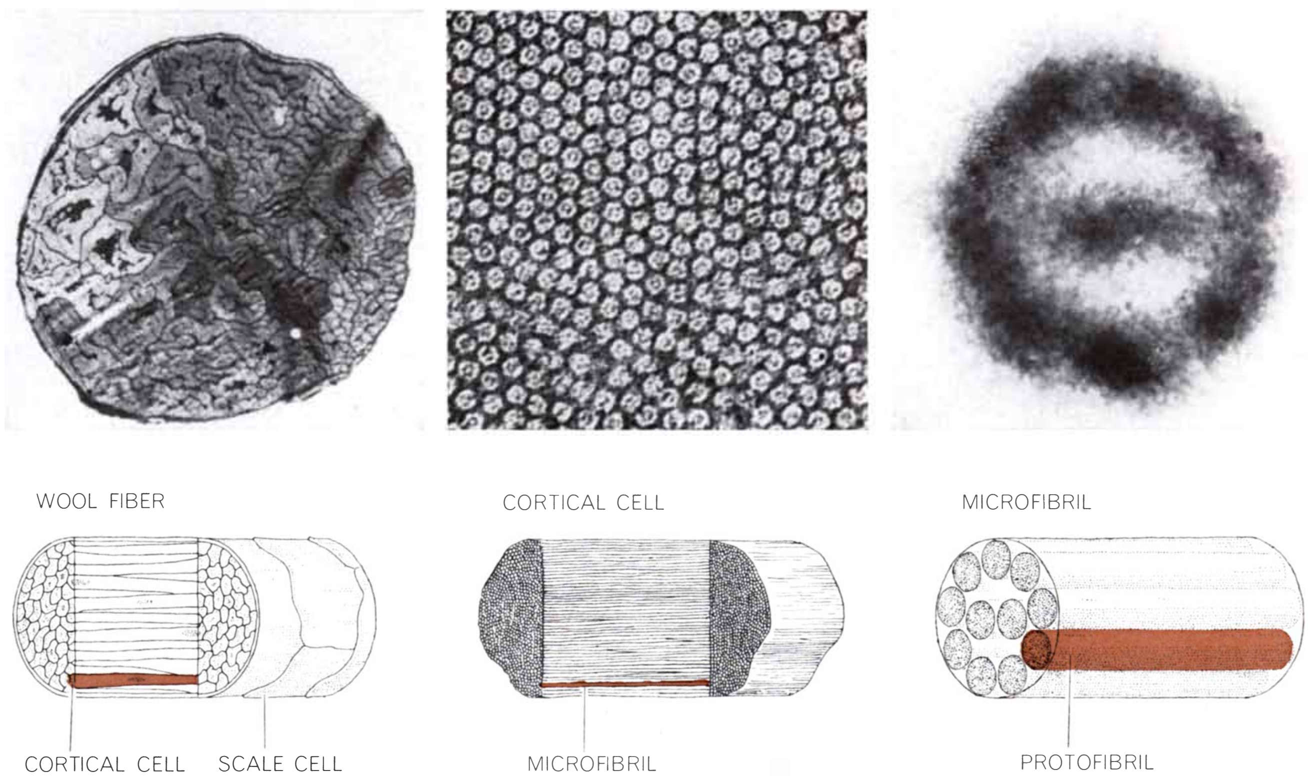
The high-sulfur fraction of wool has been studied intensively by J. M. Gillespie and his co-workers in our laboratory. They have learned that the amount of high-sulfur protein varies considerably, depending on the diet of the sheep; as the intake of nutritional sulfur increases, so does the proportion of the wool components that are rich in sulfur. When the sheep's consumption of sulfur is reduced below a certain critical level, the proportion of high-sulfur protein shows no further decrease. Instead the growth of wool is progressively retarded. The sulfur content varies widely in different keratinous tissues, even in the same animal. It is much higher in wool, for example, than in sheep horn. Among the epidermal tissues examined so far the highest content of high-sulfur proteins (45 percent) has been found in raccoon hair, the lowest content (7 percent) in rhinoceros horn. This wide variation suggests that the great differences in the various keratins may be explainable at least in part in terms of differences in the composition and amount of the ma-



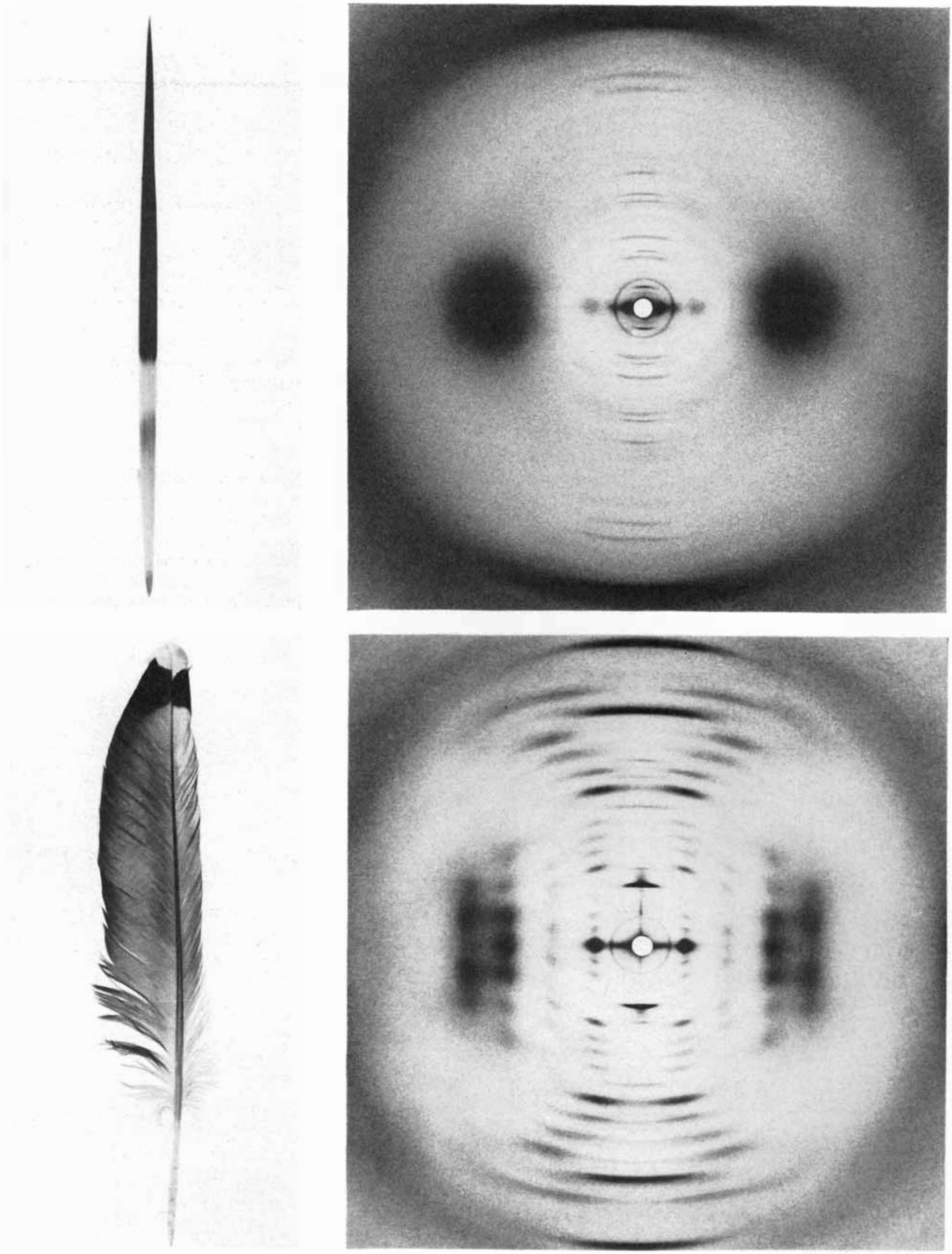
ELECTROPHORESIS TECHNIQUE is one of the techniques used to separate soluble keratin proteins into components of different chemical composition. The basis of the technique is the differing rates of migration of different fractions under the influence of an electric field. The resulting electrophoretic pattern or "spectrum" of keratin proteins varies widely among species, among animals of the same species and even in the same animal; it has also been found to be influenced in part by diet.

trix material binding the microfibrils together.

Along with the investigation of the components of keratin in solution there has been a continuing study of the architecture of keratins in their intact, native state by means of X rays. It is not possible to obtain detailed information about their atomic structure in this way, because keratin molecules do not pack together in the three-dimensionally regular arrangement characteristic of a crystal [see "The Hemoglobin Molecule," by M. F. Perutz; SCIENTIFIC AMERICAN, November, 1964]. Nevertheless, X-ray examination of hard, nearly crystalline structures such as the shaft of a seagull feather or the sharp tip of a porcupine quill has been productive [*see illustration on next page*]. In the 1930's at the Rockefeller Institute, Corey and R. W.



STRUCTURE OF A WOOL FIBER is illustrated here by means of electron micrographs and corresponding drawings. The low-power micrograph at left, made by Ruth M. Bonès of the Textile Physics Laboratory at Leeds, shows a cross section of a whole wool fiber. The cortex of the fiber is made up of spindle-shaped cells oriented parallel to the fiber axis. The high-power micrograph at middle, made by George E. Rogers and B. K. Filshie in the author's laboratory, shows that the cells in turn are made up of microfibrils embedded in an amorphous matrix. The enlargement at right, made by superposing a number of microfibril images, suggests that the microfibril itself has a substructure consisting of a "9 + 2" assembly of protofibrils, each about 20 angstroms in diameter.

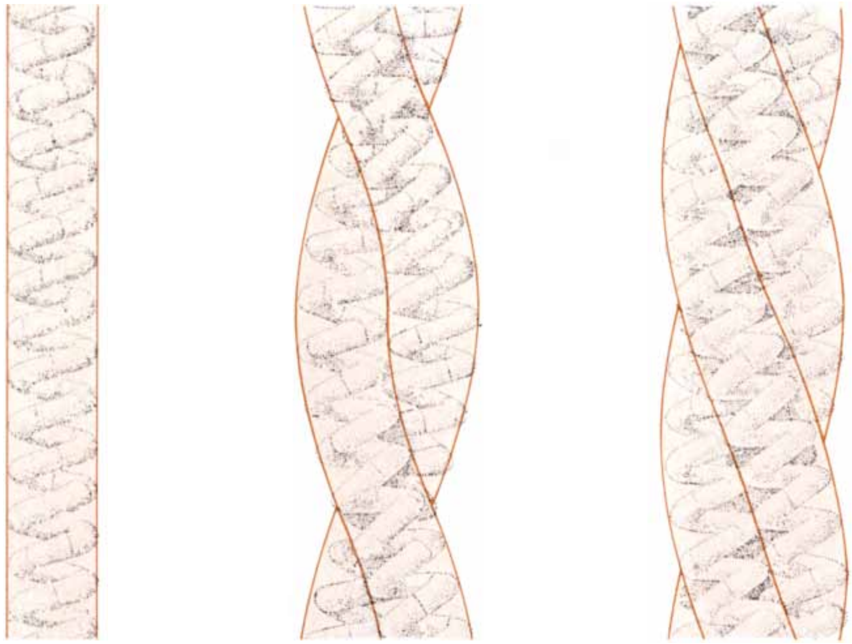


HIGH-RESOLUTION X-RAY PATTERNS obtained from the tip of a porcupine quill (*top*) and the shaft of a seagull feather (*bottom*) show a rich array of reflections owing to the fact that keratins are formed by the regular end-to-end aggregation of precisely synthesized protein molecules. The lateral spread of the reflections results from a less precise lateral organization of the molecules.

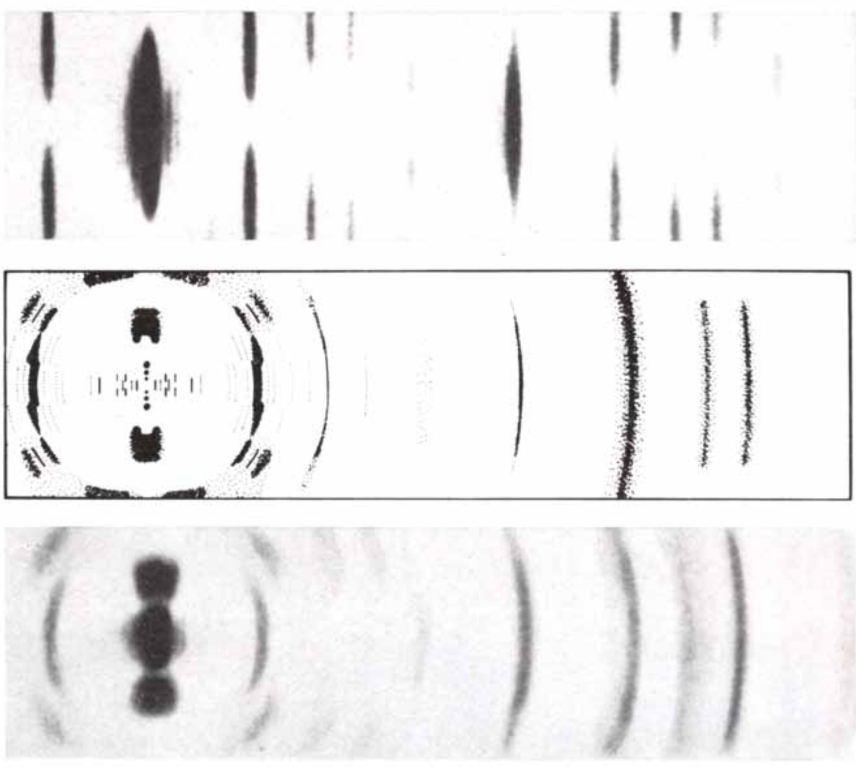
G. Wyckoff, examining these structures with X-ray beams reflected at very small angles, obtained diffraction pictures showing quite clearly that the backbone of the keratin fiber consists of long units of identical length joined end to end. Richard S. Bear of the Massachusetts Institute of Technology and Ian MacArthur of Leeds later confirmed this finding and calculated that the units were at least 200 angstroms long. These X-ray studies also indicated that a less regular lateral organization was present. The nature of this organization became clear when it was discovered, with the electron microscope, that keratin is made up of microfibrils packed together in imperfectly crystalline arrays. With this knowledge and the assumption that the microfibrils are cylindrical in shape and identical in composition T. P. MacRae and the author used the X-ray picture to show that the diameter of the microfibrils was 75 angstroms.

George E. Rogers and B. K. Filshie of our laboratory went on to discover with the electron microscope that the microfibrils in turn appeared to be made up of still finer filaments: "protofibrils" about 20 angstroms in diameter. The detailed interpretation of the electron microscope image in this way was questioned by J. Sikorski and his co-workers at Leeds, and a number of methods are being used to try to corroborate the existence of protofibrils, so far without success. The microfibril images are highly variable, and G. R. Millward and the author have used an optical-filtering technique to obtain an "average image" from the electron micrographs made by Rogers and Filshie. (We modified a technique employed to study the structure of viruses by A. Klug and his colleagues at the Medical Research Council's Laboratory of Molecular Biology in England.) In addition, MacRae and the author devised a method of obtaining an axial projection of the electron density in a microfibril directly from the X-ray picture. The results obtained by these two techniques are quite similar in showing that the microfibril consists of an annular ring surrounding a central core [*see illustrations on next page*]. Although these results provide a much more accurate idea of the inner structure of the microfibril, they have not so far revealed protofibrils.

Meanwhile progress has been made in unraveling the way in which the polypeptide chains in the microfibrils are coiled into helixes. C. H. Bamford and his co-workers at the Courtaulds Laboratory in England opened a door to



DISTRIBUTION OF AMINO ACID GROUPS (*short cylindrical sections*) along the polypeptide chain that forms the backbone of the keratin molecule is explained in terms of the helical protein structure proposed by Linus Pauling and Robert B. Corey in 1951 (*left*). F. H. C. Crick has pointed out that pairs of such helixes could approach closer and pack more neatly if they were distorted so as to twist around each other, thus producing a two-strand ropelike assembly (*center*) similar to the structure of the DNA molecule. A three-strand rope is also possible, and Pauling and Corey have designed seven-strand cables.

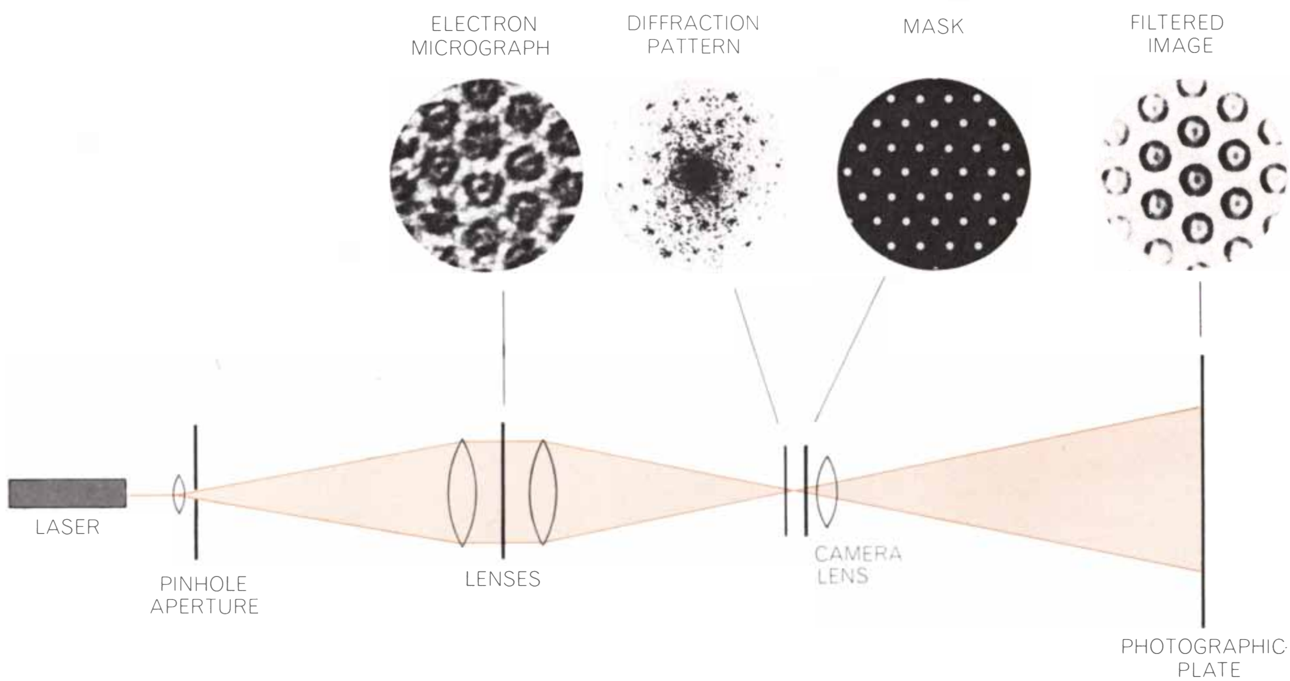


OPTICAL DIFFRACTION PATTERNS obtained from masks representing a helical polypeptide chain (*top*) and a length of three-strand helical rope (*bottom*) are compared with a diagram of the X-ray diffraction pattern given by a sample of alpha keratin (*middle*). The masks were prepared by plotting projections of the protein structures on a suitable scale and punching holes to represent the positions of the atoms. The area of each hole was made proportional to the number of electrons in the atom. The ropelike distortion of the helixes obviously results in a much improved agreement with the observed X-ray pattern.

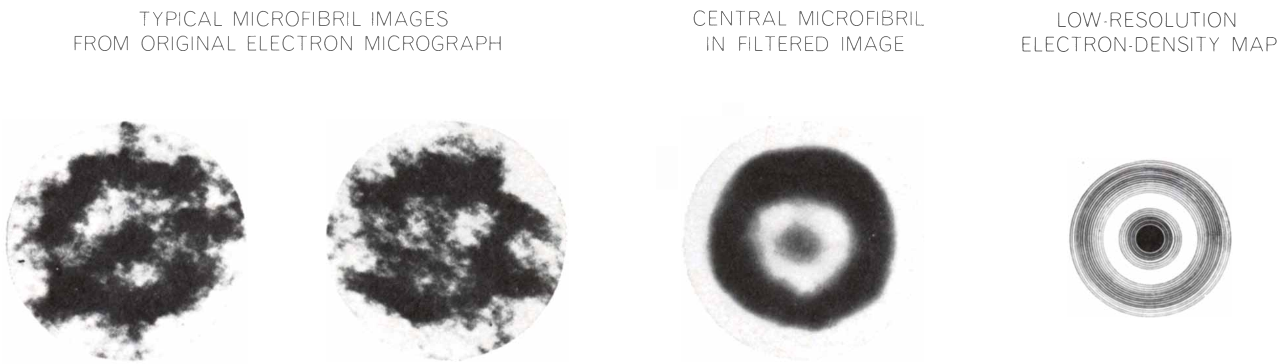
detailed investigation of this question by producing synthetic polypeptides, made up of one kind of amino acid, that yielded X-ray diffraction patterns similar to the alpha pattern of mammalian keratin. The synthetic polypeptides were found to be folded in the form of regular alpha helixes of the type proposed by Pauling and Corey. Their X-ray patterns were slightly different, however, from those of natural keratins, and this was interpreted to mean that in the keratin chain the helixes might be packed closer than they are in the synthetic chains. F. H. C. Crick, the codiscoverer of the double helix of DNA, suggested that the keratin chain might be made up of two or three helical strands, inclined to one another in such a way that each amino acid of one helix nestled in a gap between successive amino acids of another helix, thus forming a compact structure. In short, the helixes might be intertwined in a ropelike structure [*see top illustration on preceding page*].

Carolyn Cohen and K. C. Holmes of the Children's Cancer Research Foundation in Boston found that a highly helical muscle protein (paramyosin from a mollusk) does in fact give an X-ray pattern closely resembling the pattern that would be expected from a two-strand rope intertwined as Crick had suggested. The X-ray patterns from mammalian keratins and muscles are less distinct than the one from paramyosin; it seems, however, that their structure roughly approximates that of a rope composed of alpha helixes regularly distorted and intertwined according to Crick's description. So far it has not been determined with certainty whether the alpha-keratin rope is made up of two strands or three, but the balance of the X-ray evidence suggests a two-strand structure.

The overall structure of the alpha (that is, mammalian) keratins is now becoming clear. The basic fibrous unit is



OPTICAL DIFFRACTOMETER was used by G. R. Millward and the author to obtain an average image of a keratin microfibril from the variable images in an electron micrograph. The regular array of spots in the optical diffraction pattern, produced with the aid of a laser from the original electron micrograph, is isolated by a mask; the central microfibril in the filtered image then corresponds to an average microfibril. The number of microfibrils included in the average is inversely proportional to the area of the holes in the mask.



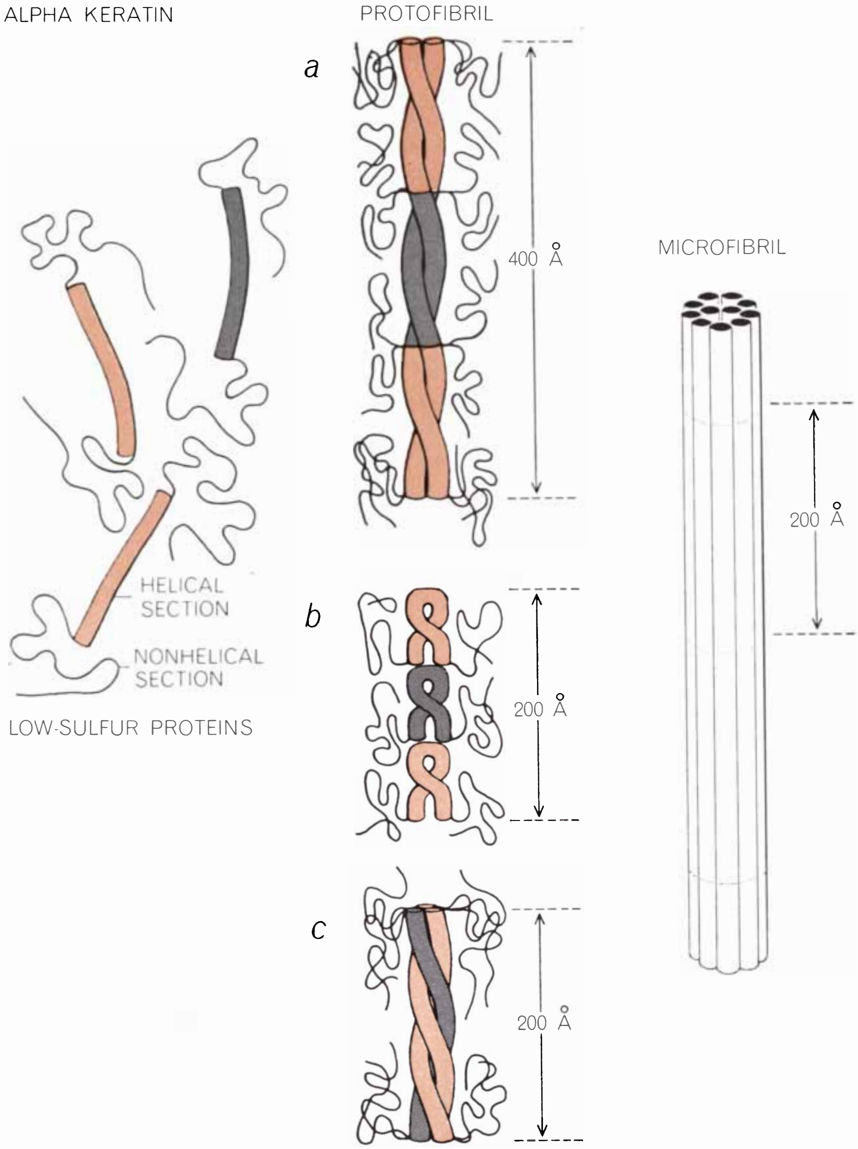
GREAT VARIABILITY in the microscopic appearance of the substructure within keratin microfibrils (*left*) is compared with the average image of a microfibril obtained by the optical-filtering technique (*middle*). The average image closely resembles the electron-density map of the end of a microfibril obtained by T. P. MacRae and the author directly from the X-ray data (*right*).

a cylindrical microfibril about 75 angstroms in diameter composed of low-sulfur proteins. The regularly folded portions of these proteins are alpha-helical, and pairs of helixes coil around each other in a ropelike manner. These ropes, or protofibrils, are in turn packed together to form microfibrils. The nonhelical component of the low-sulfur protein must play an important part in regulating this packing, and presumably it fills the interstices between the central protofibril and the outer ring of protofibrils.

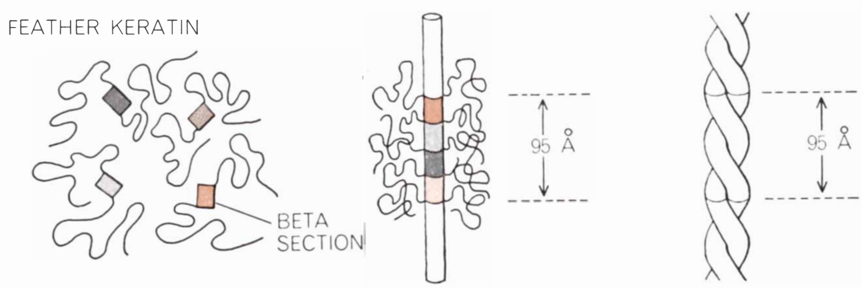
The microfibril itself seems to be a remarkably uniform structure. In electron micrographs its cross section looks much the same in all mammalian keratins that have been examined, from the quills of the spiny anteater to mammoth hair. X-ray studies show that the length of the helical units forming the protofibrils does not differ from species to species but the composition of the nonhelical portions does. The principal differences among the mammalian keratins lie in the mode of packing of the microfibrils and in the amount and constitution of the high-sulfur matrix in which they are embedded. It therefore appears that these factors are the main ones determining the specialized forms and properties of the various keratins.

The matrix of high-sulfur proteins in keratins is not as well understood as the fibrous framework. It used to be thought that this material resembled a cement or vulcanized rubber. Recent studies indicate, however, that these analogies may not be accurate. When wool or hair is loaded with water, nearly all the swelling takes place in the matrix, which is not the behavior that would be expected of a cementing substance with random three-dimensional cross-linking. The experimental observations suggest that the high-sulfur proteins have a globular structure, with many of the disulfide links serving to bond the molecule internally. It remains a fact, however, that the properties of keratins depend greatly on the amount of cross-linking by disulfide bonds. For example, there is much less linking of this kind in the soft, pliable epidermis of skin than there is in the keratin of the hard, inelastic feather.

The naturally occurring beta type of keratin, commonly called "feather keratin," has not been studied as thoroughly as mammalian keratin. Rogers and Filshie, studying feathers with the electron microscope, found that these, like mammalian keratins, consist of assemblies of microfibrils embedded in a matrix.



SPECULATIVE PICTURE of the molecular organization of alpha keratin visualizes the protofibril as a linear aggregate of low-sulfur proteins. The two major components, which are present in the ratio of about 2 : 1, may be arranged either in series (*a*, *b*) or in parallel (*c*). The helical sections of the molecules are shown twisted in the form of intermittent two-strand ropes (*a*, *b*) and an intermittent three-strand rope (*c*). The simple series model (*a*) leads to a periodicity of about 400 angstroms in each strand, and the strands must be staggered by half in either the protofibril or the microfibril to match the 200-angstrom repeat for the entire structure deduced from the X-ray data. The strong 66-angstrom reflection is readily explained both by this model and by the folded model (*b*) suggested by Hugh Lindley. Variations are obtained by staggering the strands in the protofibril longitudinally.



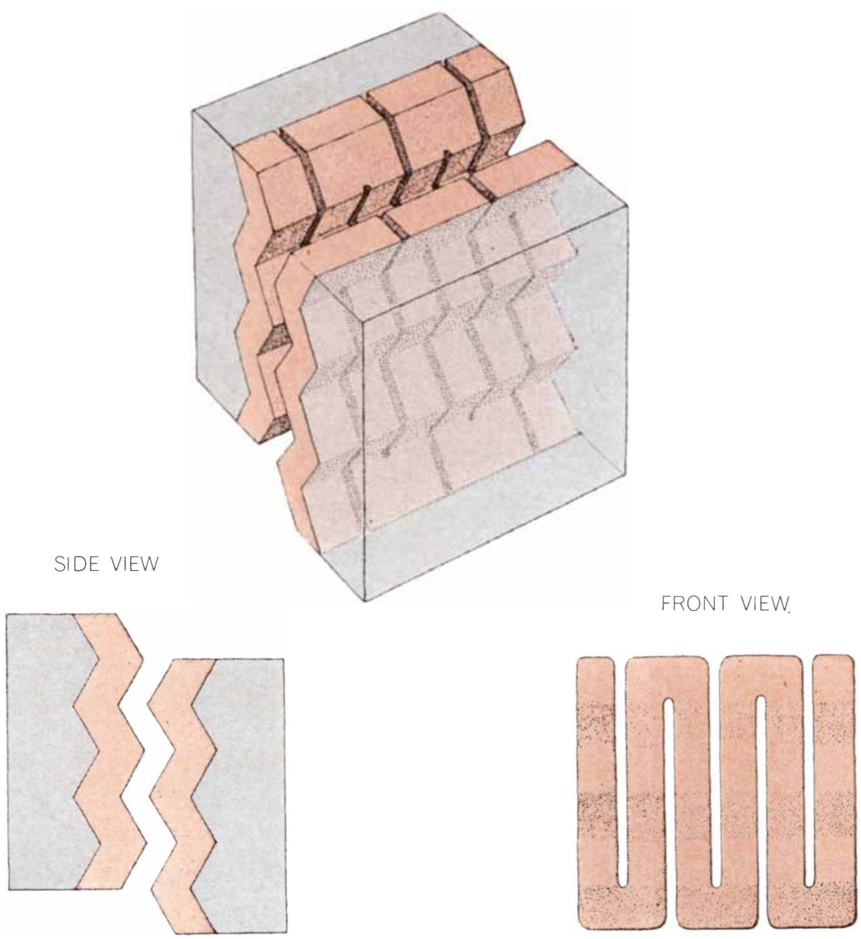
IN FEATHER KERATIN the different species of molecules (*left*) are visualized as aggregating end to end in a regular manner to give protofibrils (*center*). Pairs of protofibrils coil around each other to give microfibrils (*right*). The observed periodicity is 95 angstroms.

The feather microfibrils, however, are much finer—only about 35 angstroms in diameter.

Are the microfibrils of feather keratin embedded in a sulfur-rich matrix like that of alpha keratin? Some of our earlier studies had suggested that this might be the case, but so far it has not been found possible to separate feather keratin into low-sulfur and high-sulfur fractions. It appears more likely that the protein molecules in feather keratin are all very similar.

The mystery of the missing matrix became clear when Eikichi Suzuki, working in our laboratory, used infrared absorption spectra to show that only about half of the feather keratin molecule has a beta structure. The other half has an irregular chain structure and probably accounts for the matrix visible in electron micrographs. Combining this result with X-ray data, D. A. D. Parry, MacRae and the author have established that the microfibrils in feather keratin are formed from two intertwined strands of molecules. Each strand is helical, with four molecules in a turn 95 angstroms long. The strands are oppositely directed, and the beta portion of each "up" molecule meshes with the corresponding beta portion of a "down" one [*see illustration below*].

Work is going forward on further elucidation of the structures of both mammalian and feather keratins by various techniques. These include studies of fragments of the molecules, efforts to regenerate fibrils from extracted proteins and the use of heavy atoms to facilitate the X-ray exploration of the molecular structure and the location of specific amino acids. Ultimately it should be possible from X-ray studies to plot the exact conformation of the polypeptide in both the feather and the mammalian keratin protein molecules. Perhaps then the evolutionary relation between the two types of keratin will emerge.



RECENT X-RAY STUDIES of seagull feathers indicate that the protein molecules that form the two helical strands in the feather-keratin microfibril are partly folded into a regular pleated-sheet structure (*color*) and are partly globular in structure (*gray*). The pleated-sheet sections of the molecules in the "up" strand mesh with the pleated-sheet sections of the molecules in the "down" strand in the manner shown in this illustration.

We're fighting a

new kind of war.

Mission: A bombing run on the enemy caught creeping through a patch of forest. Bombload: 12,000 pounds of water. The enemy? Fire. Forest fire.

The plane is a unique tactical weapon that will be deployed against an enemy that burns over 6,000,000 acres of timberland a year in North America alone. On a fire alert, it will fly to a lake near the fire, skim the surface at 70 knots and scoop up 1,440 gallons of water in a matter of seconds. In an average day's mission, it can deliver more than 100,000 gallons.

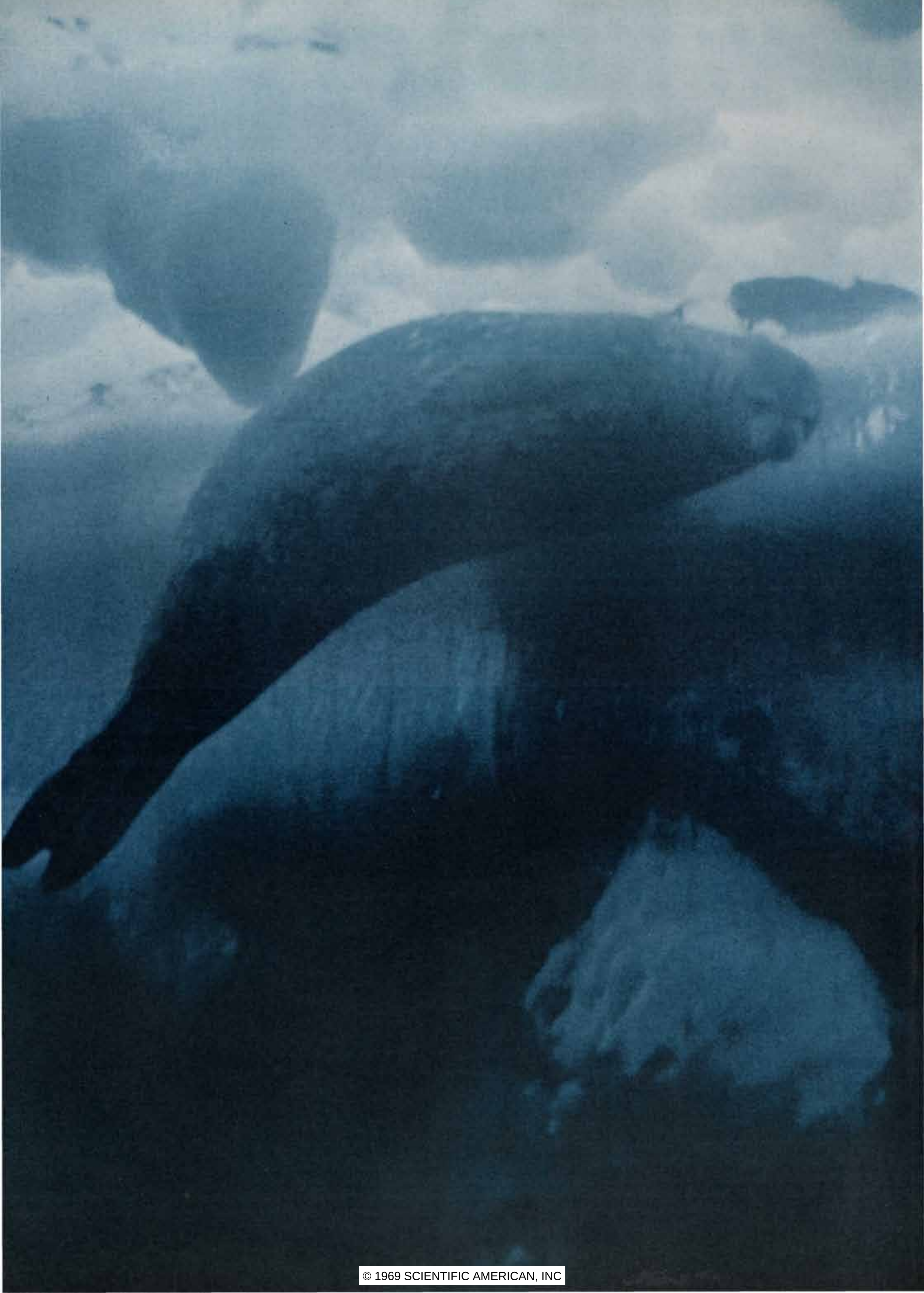
The CL-215 may not be the prettiest plane ever built. But it will do its job four times better than any airplane ever conceived before.

General Dynamics' subsidiary, Canadair Limited, is building a fleet of them for France and the Province of Quebec. And other countries are looking at it with interest. It can also be used for fertilizing and spraying, dispersing oil slicks, and can be adapted to utility transport, surveying and rescue work.

It's just one example of what technology can accomplish when it's handed a problem.

At General Dynamics, we put technology to work solving problems from the bottom of the sea to outer space...and a good bit in between.

GENERAL DYNAMICS



THE WEDDELL SEAL

This Antarctic mammal can swim under the ice for more than an hour without coming up for air. How is it able to find its way back to its breathing hole, particularly during the long Antarctic night?

by Gerald L. Kooyman

A century and a half ago the explorer James Weddell, during a record penetration into the Antarctic, discovered a remarkable mammal that spends its life in rigors of cold few other mammals are known to endure. The Weddell seal, as it has come to be known, lives in the coastal waters of the Antarctic Continent, where the air temperature drops as low as −70 degrees Fahrenheit, the water temperature is generally about 28 degrees and the sea surface is frozen to a depth of several feet for about eight months of the year. The animal has to forage for its food under the ice. Yet the Weddell seal manages to thrive in this icy fastness the year round, and during the long winter night it actually grows fat.

Weddell described the seal in his book *A Voyage Towards the South Pole* in 1825, but little more was learned about the animal until major expeditions began to visit the Antarctic early in this century. In an expedition of 1901 to 1904 the English physician Edward A. Wilson (who later lost his life in the Scott expedition to the South Pole) made a general study of the Weddell seal's habits. The animal also received the attention of investigators in later expeditions, and it has been studied intensively since the establishment of permanent bases in the Antarctic in the 1950's. Probably more is now known about the Weddell seal than is known about any other marine mammal. The reasons for the great interest in this animal are twofold: it provides a unique opportunity for study of a mammal's adaptation to extreme conditions, and both its habitat and the animal itself furnish an ideal laboratory situation for investigation of the animal's behavior.

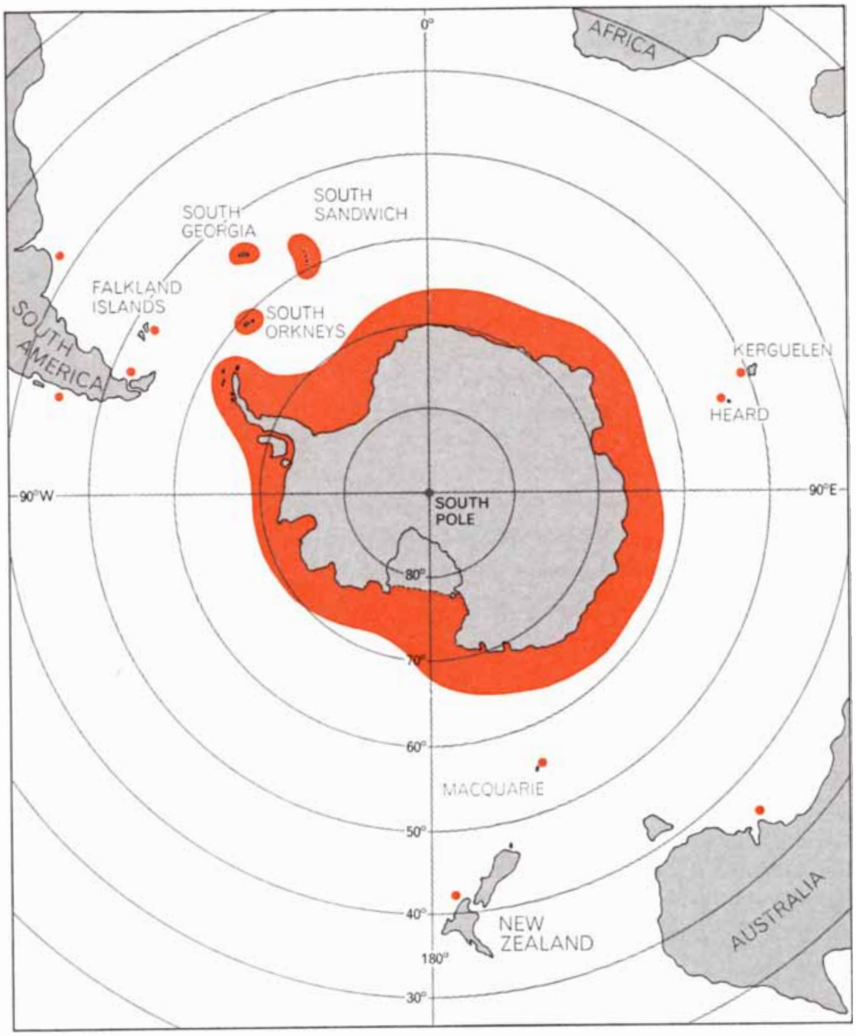
By virtue of its isolation in an environment too harsh for human invasion (except by scientific expeditions) the Weddell seal has no fear of man. Whereas seals in other parts of the world have been hunted so aggressively that they can be approached only by stealth, the Weddell seal does not flee from men either afoot or in vehicles. One can drive up to the seal in a truck and capture it easily. The ice platform over the deep sea, particularly at McMurdo Sound, where the ice remains firm into the Antarctic summer, provides a stable base for a marine laboratory (in contrast to a tossing ship). And the water under the ice is the clearest oceanic water in the world, so that the seal's movements in the water can be seen from an observation chamber under the ice. These conditions, together with the presence of large populations of Weddell seals in the sound, have drawn biologists to various specific investigations of the animal. Among others, investigators from the Woods Hole Oceanographic Institution have made a comprehensive analysis of underwater vocalization by the seals, physiologists from the Scripps Institution of Oceanography have studied their defense mechanisms against asphyxiation during dives, and zoologists from the University of Canterbury in New Zealand have analyzed the seals' population structure.

WEDDELL SEAL shown on the opposite page was photographed swimming under the ice near Cape Armitage in Antarctica. The photograph was made by the author, who was also under the ice with an underwater camera, which he used in conjunction with flash equipment. The seals are easily approached because they have no natural predators and therefore are almost fearless.

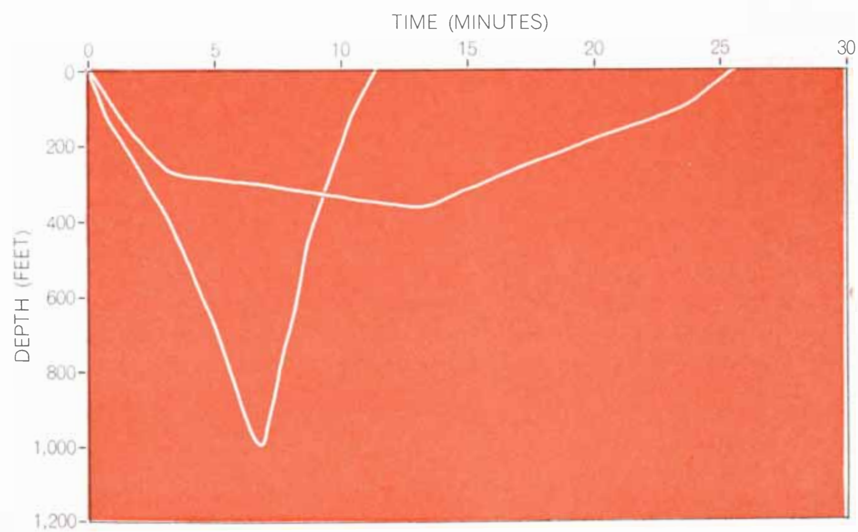
My own investigations have focused on the seals' diving behavior and their capacity for doing without oxygen and withstanding pressure in lengthy or deep dives. Thorough studies of these matters had never been made on a marine animal in its natural environment, and the Weddell seal provided an outstandingly suitable subject for experimental examination of its underwater behavior.

There are small populations of Weddell seals in various island groups of the South Atlantic and sub-Antarctic regions, but the main habitat of this animal is the icy waters off Antarctica. It is known to breed at least as close as 10 degrees from the South Pole. The Weddell seal lives on fish, and for most of the year it must hunt them under the ice. For access to the water, and to return to the air for breathing, it depends on breaks in the sea ice produced by the wind, tides and currents. The animal has unusual canine and incisor teeth that it uses to ream out breathing holes in the thin ice when the breaks freeze over. One of the principal mysteries about the Weddell seal is how it manages to find the widely dispersed breaks in the ice when it is underwater, particularly in the darkness of winter.

For our studies of the animal's underwater performances we set up a station (a heated hut) on the sea ice in a location where the water was approximately 600 meters deep. We chose an area where the ice was free from cracks and seal holes for some distance around the station. At the station itself we cut a hole in the ice through which we would release our seals into the water. Our procedure was to capture a seal on the ice several miles away, bring it to the station, attach a small pack of instruments to its back and then obtain a record of its movements after it plunged into the water through the hole we had provided. We used a variety of instruments: a manometer for measuring the maximum depth of dive, other devices for recording the depth of the dive as a function of time,



RANGE OF WEDDELL SEAL is shown on this map. The animal is primarily a coastal dweller of the Antarctic Continent; its range is shown in color. Dots indicate areas where the seal is occasionally found. Its most northerly appearances have been in sightings from the Falkland Islands and the coasts of Uruguay, Chile, New Zealand and southern Australia.



TIME AND DEPTH of exploratory and deep dives of the Weddell seal are portrayed. The patterns were obtained partly from watching seals from an underwater chamber after they were released in a new hole and partly from recording equipment strapped to the seals.

and an underwater telemetry device that supplied us with running records during the dive. With these aids we were able not only to determine the depth and duration of the seals' underwater swims but also to estimate the swimming speed.

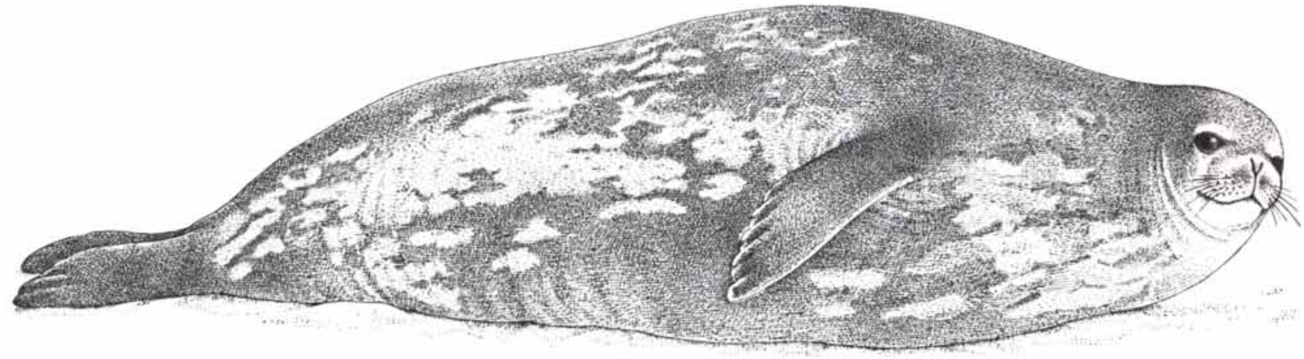
We observed more than 1,000 dives, and we found that the Weddell seal typically engages in three types of dive with different purposes. In each kind of dive the animals' performance in the water raised interesting questions for investigation.

The first type of dive is a brief foray of local exploration. When we released a seal into the water at the dive station for the first time, it would swim about at shallow depths, no more than 100 meters, and soon return to the breathing hole, usually within five minutes. In these initial, local dives, which might be repeated several times, the seal evidently occupied itself in testing the immediate surroundings and searching out other occupants of the water. If there were several seals in the water at the same time, they often engaged in competitive fighting, apparently for privileges such as being the first in line at the breathing hole.

After exploring the local scene the seal usually proceeded to a second stage of much more extensive exploration. These dives were somewhat deeper, down to 130 meters or so, and lasted from 20 minutes to an hour or more. The animal would range over considerable distances under the ice, often for miles, looking for other breathing holes. If it failed to find one, it would return to the dive station to rest. It sometimes repeated these long, exhausting swims for several hours with a few short, local dives after each prolonged dive.

When a seal did find another breathing hole, it might or might not come back to the station hole. One seal returned from its expeditions again and again and made its home at our station for more than a month. We identified it easily because its head was hairless, and we developed a fond feeling for it. When a seal did not return to us after finding another hole, we would hunt for it to recover our instruments. With the aid of our rapid means of transportation, including a helicopter, we succeeded in finding nearly all the errant seals. In one case we recaptured the animal on the ice 15 miles from the dive station.

The eminent marine physiologist P. F. Scholander, of the Scripps Institution of Oceanography, has investigated the extraordinary mechanisms that enable diving animals such as seals to spend long periods underwater without as-



ADULT WEDDELL SEAL can attain a length of about 10 feet and a weight of as much as 1,000 pounds. Most of the animal's fur is black or dark brown; the lighter spots are almost white. Seen from underwater the seal looks black and, like the normal pattern of light near the surface, rather mottled. When the seal is out of the water, the fur is seen to be predominantly brown after it dries.

phyxiating. In laboratory experiments he found that during a dive a seal slowed its heartbeat to a tenth of the normal rate, reduced its rate of metabolism and drastically restricted the circulation of the blood to all tissues except the heart and the brain [see "The Master Switch of Life," by P. F. Scholander; SCIENTIFIC AMERICAN, December, 1963]. In this way the animal conserved the supply of oxygen in its blood during its dive. Scholander showed in the laboratory that a seal could stay underwater for 20 minutes or more.

In our observations of unrestrained Weddell seals in their natural environment we found the animals to be capable of performances considerably better than this. They sometimes spent up to an hour in the water without surfacing for air, as our instruments showed. How far did they swim? We can estimate from indirect indications that they frequently covered several miles. We found that a seal normally took about 12 minutes to swim from one hole to another approximately a mile away; if we assume that it swam a straight route (and our observations indicated that the seals rarely deviate from a fairly straight line in such a journey), this meant that the animal's swimming speed was about five miles per hour. It is obvious, therefore, that in an underwater stay of half an hour or more a seal swims several miles.

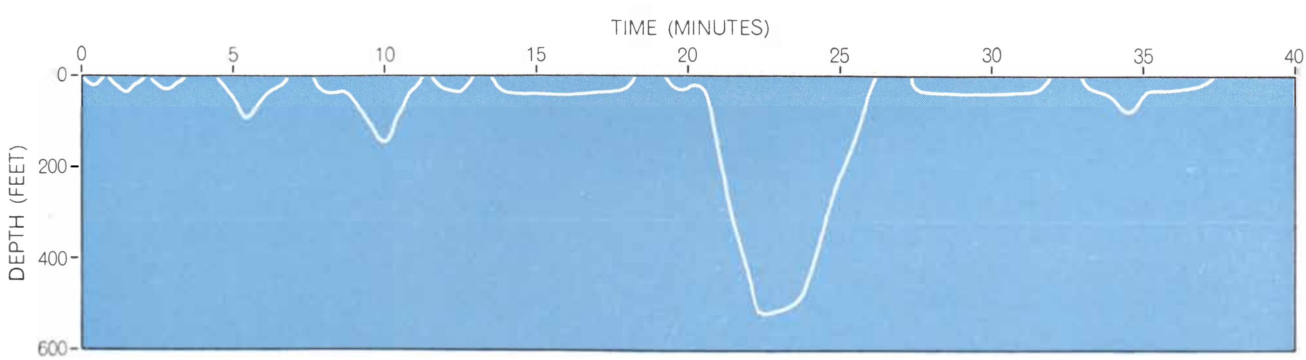
On occasion a seal would swim away from our station and return after 30 to 60 minutes without having surfaced elsewhere—an indication that it had failed to find another breathing hole. This raises a fascinating question: How does the animal know when it must turn back? What manner of internal signal induces the seal to turn back, and is this control mechanism sensitive to the half-life of the seal's submersion capacity? Research on the seal's ability to survive underwater has not yet supplied an answer to this puzzle.

There is a corollary question: What gift of farsighted perception informs the seal when a breathing hole lies ahead within reach so that it can safely pass the point of no return?

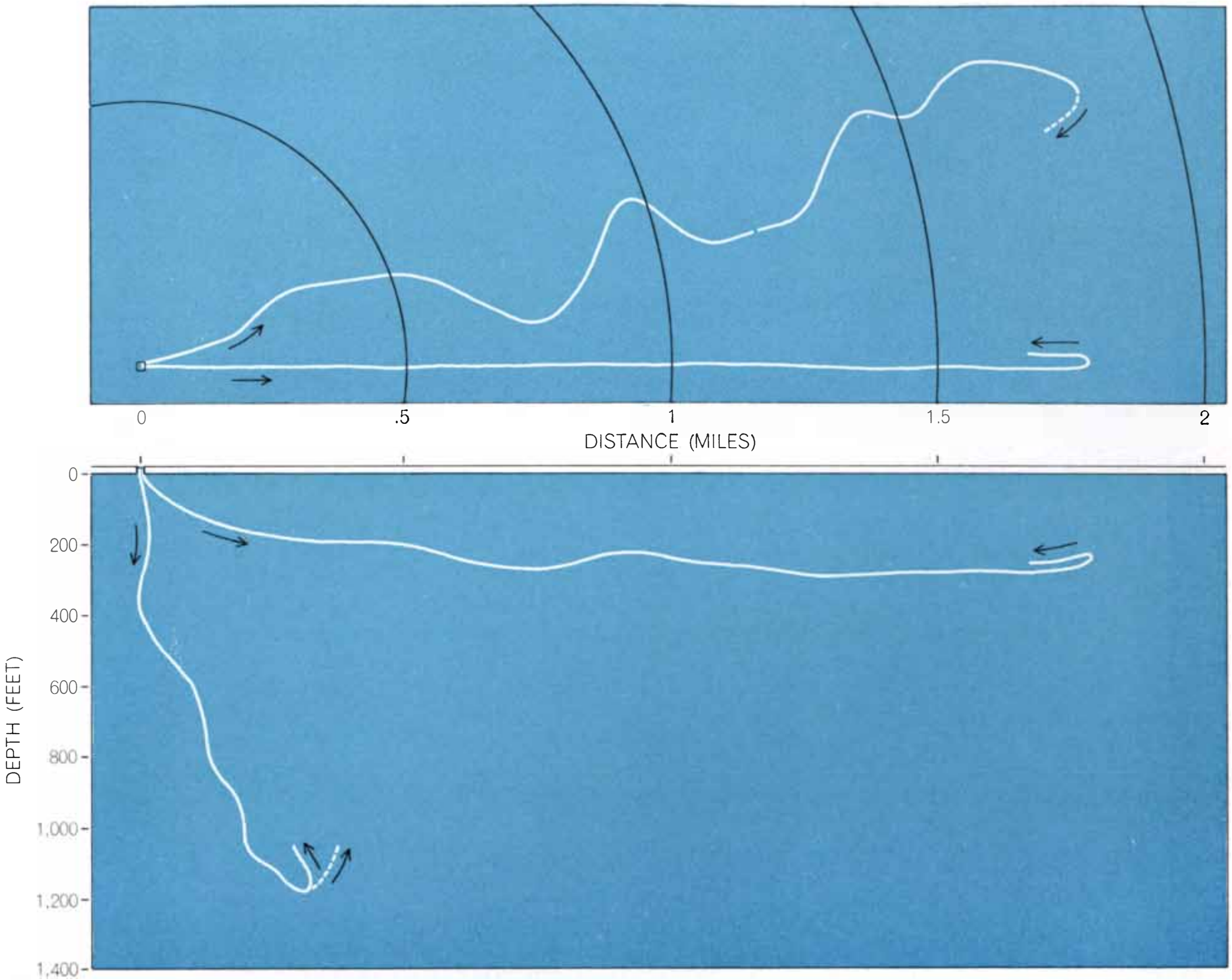
Certain observations of the seal's behavior during long trips offer grounds for speculation. On such a trip the seal seemed to maintain a constant heading: in almost 90 percent of the cases the seal would swim away from the station in a particular direction and return from the same direction. This practice makes it comparatively easy for the seal to find its way back to the breathing hole it has left. Secondly, in their long trips of exploration the animals swim at relatively shallow depths, and in the daylight conditions of the Antarctic summer the frozen sea surface (and any discontinuity in it such as a crack or hole) is readily detectable from the normal depth of their prolonged dives.

How does the seal find its way, however, in the winter darkness? During the long winter night the only light in the Antarctic seas, other than that from the moon and stars, is the bioluminescence of certain living things in the water. Does the Weddell seal use sound (perhaps echo location) or other external cues to navigate in the dark? It would seem that the animal must then depend on a navigational aid of some kind other than vision, not only for its shallow exploratory journeys but also in the third type of dive in its repertory: deep dives in pursuit of food.

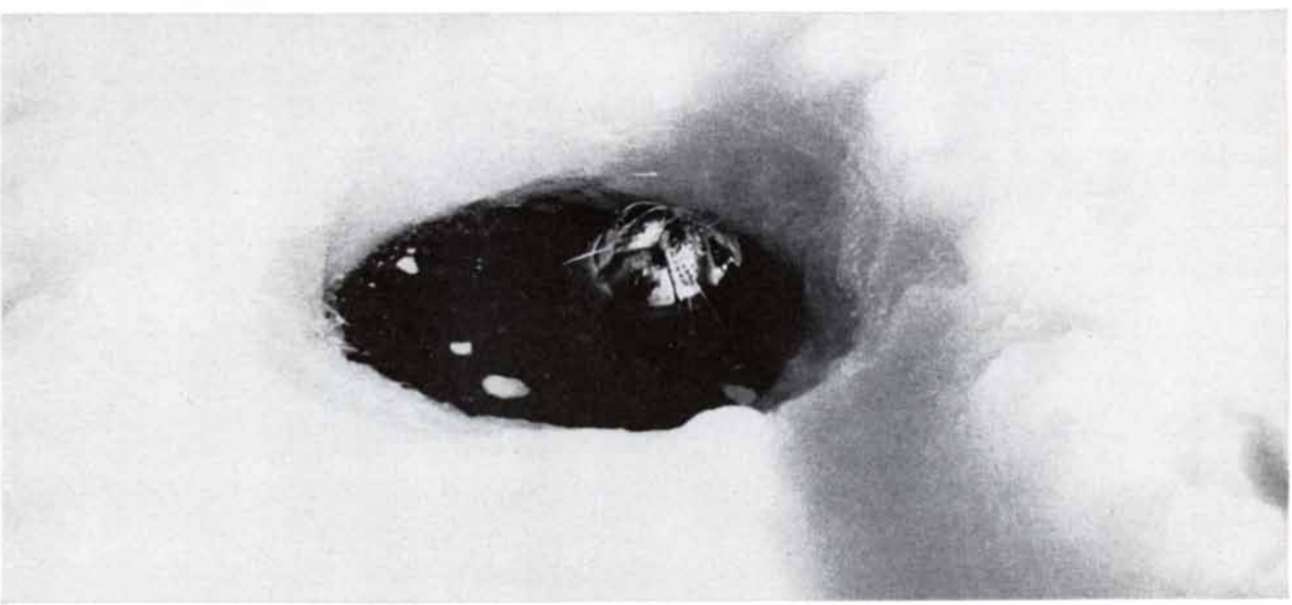
In these dives the Weddell seal commonly descends to depths of 300 to 400 meters, and we have found that it can go as deep as 600 meters. The deep dive is relatively brief, seldom lasting more than 15 minutes. With three-minute rests at



DIVE PATTERNS of the Weddell seal include a series of short, shallow explorations immediately after the seal is released in an unfamiliar hole. Spaces between dive profiles represent times when the animal returned to the breathing hole for air or rest or both.



EXPLORATORY AND DEEP DIVES are portrayed. At top are plan views of exploratory dives; headings are hypothetical. Random deviations (*upper curve*) would be hazardous because the seal might lose track of its breathing hole; the actual heading appears to be straight out and back (*lower curve*). At bottom are elevation views of two dives. In an exploratory dive the seal retains visual contact with the ice; in a deep dive random deviations are possible because the seal can find the breathing hole merely by going up.



BREATHING HOLE surrounded by a great expanse of ice is characteristic of the Weddell seal's environment. The seals have adapted to the topography by developing navigational skill at returning to familiar holes and skill at reaming new holes in thin or cracked ice.

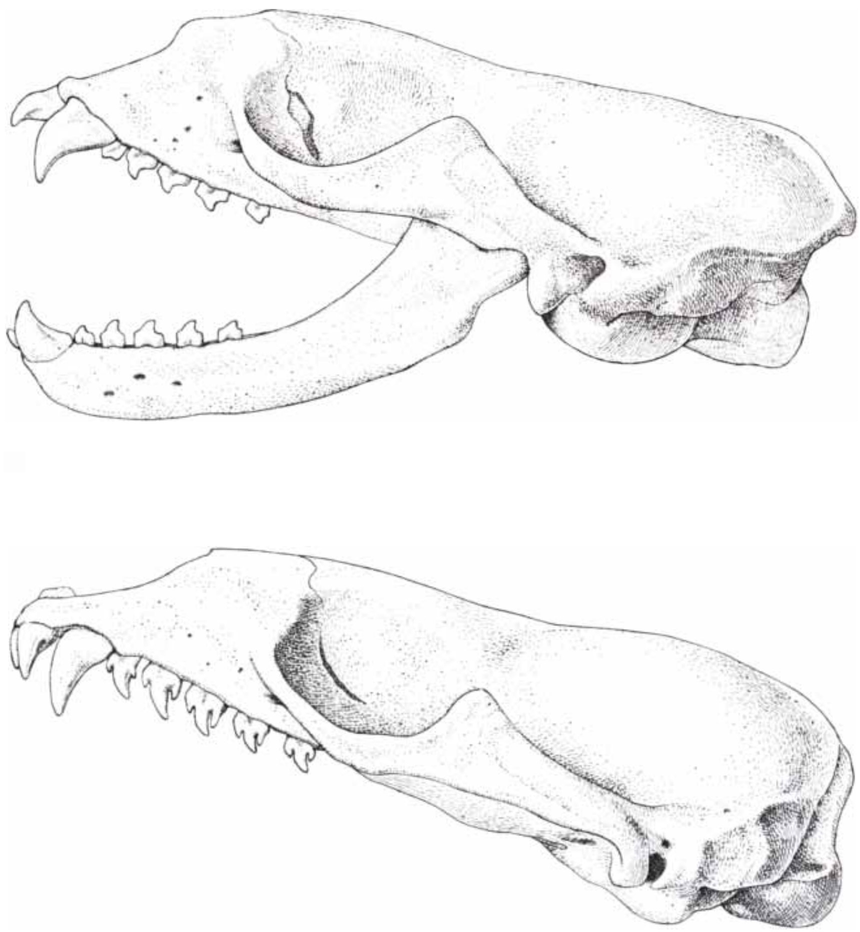
the surface between the dives, the seal may continue its feeding dives for several hours. The problem of returning to the breathing hole is much simpler than it is in an exploratory dive; in the deep dive the seal remains in the vicinity of the hole, the direction for the return (more or less straight up) is reasonably clear, and the short duration of the dive gives the animal a large margin of safety within its capacity for underwater endurance.

Our studies of the seal were conducted during the season when it has light to assist its navigation. Future investigations in the dark season may well prove highly interesting. Perhaps the animal's diving behavior changes to a different pattern during the winter. At all events, the Weddell seal's method of navigation under the ice, in summer or winter, is no doubt both effective and complex, utilizing an array of external cues and internal receptors. The analysis of its guidance system should be an inviting subject for study.

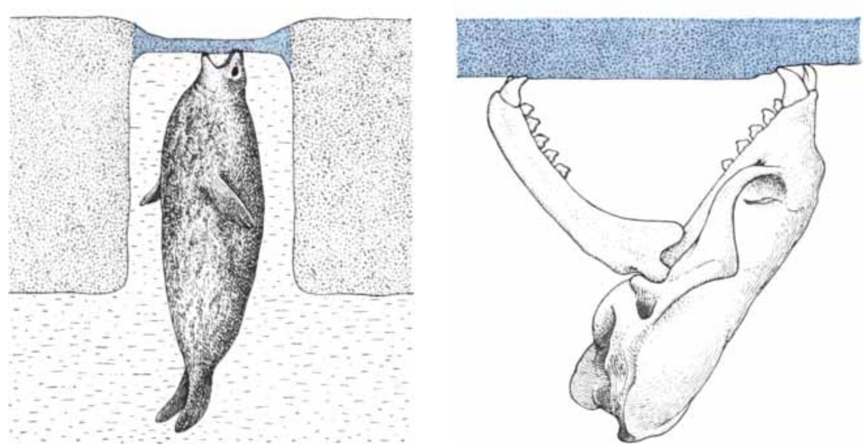
No less interesting is the seal's ability to dive to great depths without suffering ill effects from the increase in pressure or undergoing the effects of decompression ("the bends") on its return to the surface. The Weddell seal adjusts to rapid changes of pressure: it can descend to several hundred meters at the rate of 120 meters per minute without harm. How does it withstand the high pressure at such depths?

The seal's anatomy indicates that the animal copes with the pressure not by resisting it but by yielding to it. Its main body cavities are so constructed that they readily compress. The rib cage is very flexible, and the diaphragm underlying the lungs is at an oblique angle to the chest cavity. Consequently external pressure can cause complete collapse of the lungs, and the gases are pushed into the bronchi and the trachea. In the Weddell seal and other Antarctic seals the trachea has an unusual shape: the tube is not round but has the form of a flat bow. Hence the trachea in turn is easily squeezed by pressure to complete collapse. Gases may also be compressed in the cavity of the middle ear; it has been observed that in some seals, sea lions and whales this cavity is lined with a network containing expandable veins, which probably swell with blood under pressure and so reduce the volume of the cavity. These anatomical characteristics suggest that during a deep dive the external water pressure may successively drive the gases out of the lungs and bronchioles and into the bronchi and the trachea. The lungs collapse, and the trachea is compressed; thus the animal can avoid dangerous differences between external and internal pressure. If this hypothesis is correct (it has not yet been verified), the seal's tolerance of pressure in a deep dive depends on the extent to which its volume of internal gases can be compressed in the lungs, trachea and middle ear.

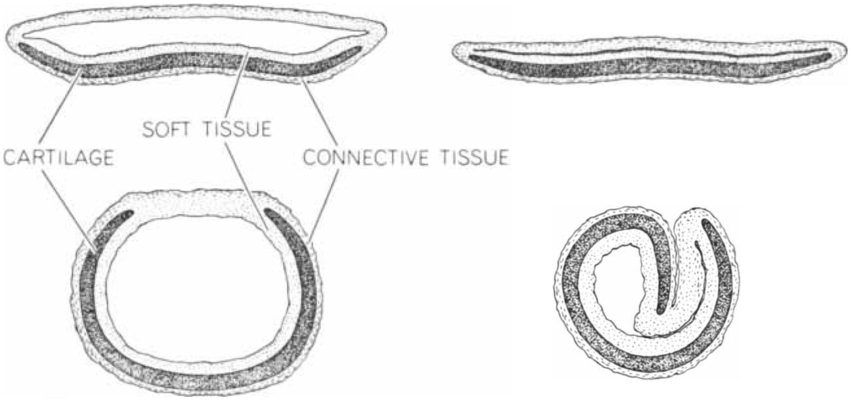
The avoidance of the bends is a more complicated matter. In mammals high pressure forces nitrogen from the air in the lungs into solution in the blood and tissues, and if the body is rapidly decompressed, the bubbling of the nitrogen



TOOTH PATTERN of the Weddell seal represents one of the animal's adaptations to its environment. At top is a Weddell seal's skull, showing how the upper teeth are shaped for reaming ice. At bottom for comparison are the upper teeth of the nonreaming leopard seal.



SEAL REOPENS HOLE that has frozen over; the task is done by a reaming action that relies mainly on the upper teeth. The smaller upper incisor teeth engage the ice first.



COMPARISON OF TRACHEAS of a Weddell seal and a human being indicate one of the seal's physiological adaptations to deep diving. The seal's trachea is normally flat (*top left*), and under pressure it collapses readily without damage to the cartilage (*top right*). A human trachea (*bottom*) is normally ringlike and cannot collapse much without damage.

DOCILE BEHAVIOR of Weddell seal when approached by humans is evident as a female calmly allows Charles M. Drabek of the University of Arizona to draw off milk with a tube.

out of the blood produces the painful embolisms known as caisson disease, or the bends. This sometimes fatal sickness, long recognized as a hazard of tunnel diggers, helmet divers and scuba divers who breathe compressed air, has recently been found to attack breath-holding divers who venture to depths of 60 feet or more in a number of dives made during a short period of time; among divers of the South Sea islands it is called *taravana* [see "The Diving Women of Korea and Japan," by Suk Ki Hong and Hermann Rahn; SCIENTIFIC AMERICAN, May, 1967]. Seals are not entirely immune; in one experiment a seal that had submerged to a considerable depth died of embolism when it rose too rapidly. We have found, however, that Weddell seals rise at the rate of 120 meters per minute without any apparent discomfort.

What mechanism protects the seals? Many years ago Scholander suggested that the lungs of marine mammals probably collapse at depth, and that the consequent absence of nitrogen in the lungs prevents absorption of the gas into the bloodstream. On the basis of lung-volume measurements he calculated that the alveoli of the lungs collapse at a depth of about 50 meters in the water. This implies that seals should be safe from attacks of the bends after dives of more than 50 meters but may be vulnerable in shallower dives. We have observed, in fact, that Weddell seals spend most of their time below 50 meters even in their comparatively shallow exploratory dives, and they usually come up slowly from a shallow dive. Moreover, they rarely make prolonged shallow dives serially. Thus their diving behavior, as well as the anatomical structure favoring collapse of the lungs, provides safeguards against the bends.

In the Weddell seal evolution has produced an animal exquisitely suited to its special environment—an environment so inhospitable that no other mammal save technologically armed man has chosen to invade it. Because of its uniqueness we are finding the Weddell seal to be a valuable animal for biological research. We must also observe, however, that for all its hardiness the species has certain disquieting vulnerabilities. One of them is its total lack of any protective fear of man or other predators. It would be an easy prey for commercial hunting. As a source of oil and protein it could provide only an ephemeral supply; on the other hand, as a source of enlightenment about life and evolution the Weddell seal has a value that is inestimable.

A portrait of a way of life as old as man

Here, safely captured for all time, is a spectacular photographic record of

one of the last Stone Age societies on earth. A society that will vanish forever by the end of this century.

Gardens of War is a complete anthropological record — in photographs and text — of the Dugum Dani, a tribe of warrior farmers in western New Guinea. Their existence revolves around constant warfare with neighboring tribes.

They do not fight for land, or for wealth, or even for victory. They fight because they believe war is sacred and necessary.

In 1961, members of the Harvard-Peabody expedition lived with the Dugum Dani for months and photographed every aspect of their daily lives. The 337 photos in the book —36 in color—vividly show the Dugum Dani at war, at work, and in their casual moments.

We see them preparing for battle and returning from it . . . treating the wounded . . . cremating the dead . . . raising pigs and sweet potatoes . . . dancing to celebrate the slaying of an enemy . . . performing rituals to placate their ghosts.

Photographs of actual battles taken from nearby heights with telescopic lenses capture the excitement and emotion of their primitive combat.

A beautiful and fascinating account of living history.

GARDENS OF WAR

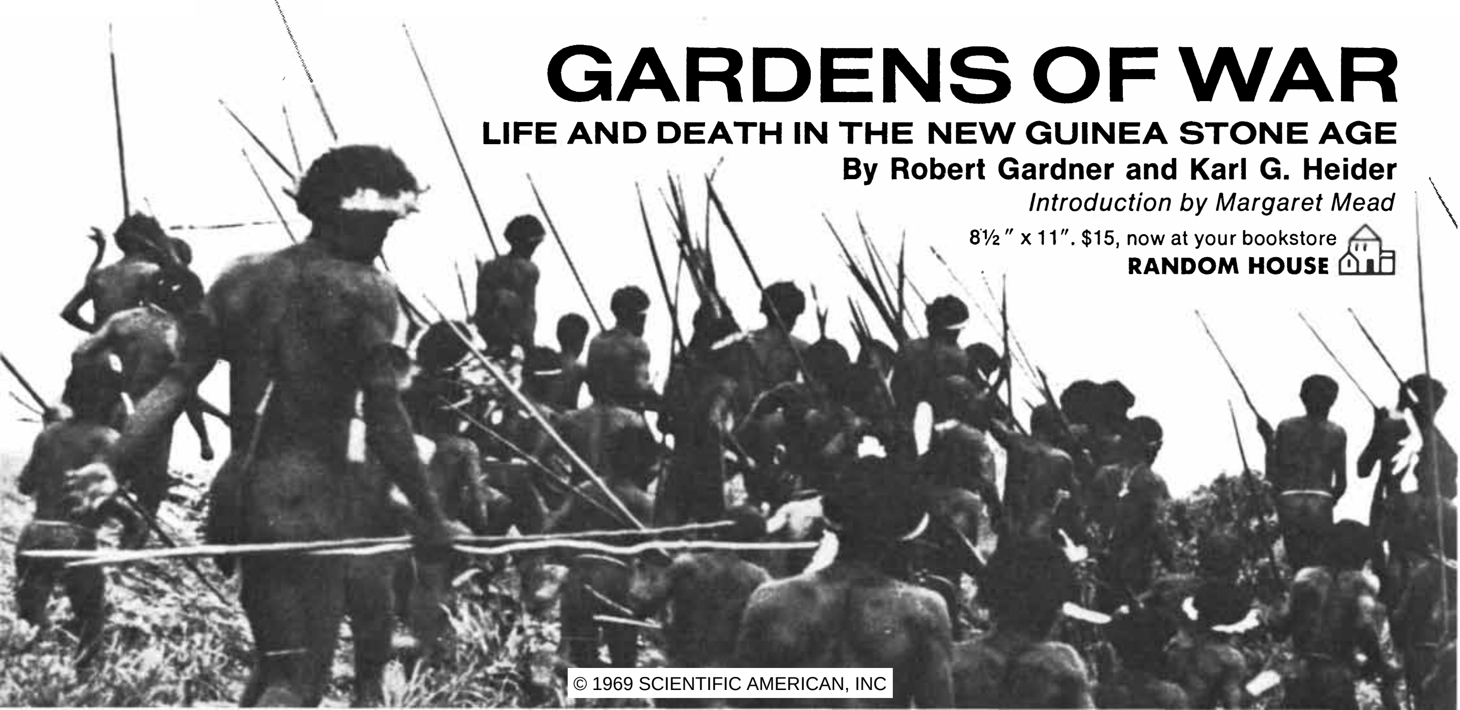
LIFE AND DEATH IN THE NEW GUINEA STONE AGE

By Robert Gardner and Karl G. Heider

Introduction by Margaret Mead

8½" x 11". $15, now at your bookstore

RANDOM HOUSE



Rudolf Diesel and His Rational Engine

The inventor of the Diesel engine hoped it would attain the ideal of the Carnot cycle for the heat engine. He failed, but today his engine provides most of the motive power for heavy transportation

by Lynwood Bryant

The Diesel engine has revolutionized heavy transportation in the past generation. The automobile, which is usually powered by an Otto engine rather than a Diesel, is responsible for vastly more passenger-miles of transportation than any other vehicle ever known, but the work of moving freight and passengers on a large scale on land and sea, which was accomplished in the time of our fathers by steam, is now done almost entirely by Diesel. The last stronghold of the old-fashioned steam piston engine—the railroad locomotive—has surrendered unconditionally in the U.S. and cannot hold out much longer in Europe. The revolution in heavy trucking, in earth-moving equipment and in water transportation is not so well known but is no less complete. The steam turbine competes with the Diesel in very large ships, say above 20,000 horsepower, but between that upper limit and the lower limit of the family car the Diesel engine is now supreme.

Rudolf Diesel fully expected his engine to cause such a revolution, but when he conceived the engine some 80 years ago, he had no idea the revolution would take so long. Diesel was a scientific engineer, one of a new type that began to appear in the 1880's, and he had a strong background in science and mathematics. In his view the engine he patented in 1892 was not just an improvement on existing heat engines but a machine of an entirely new kind. He called it a "rational engine," that is, an engine built on scientific, rational principles, and he fully expected it to displace the steam engine within a few years and to drive everything from battleships to sewing machines.

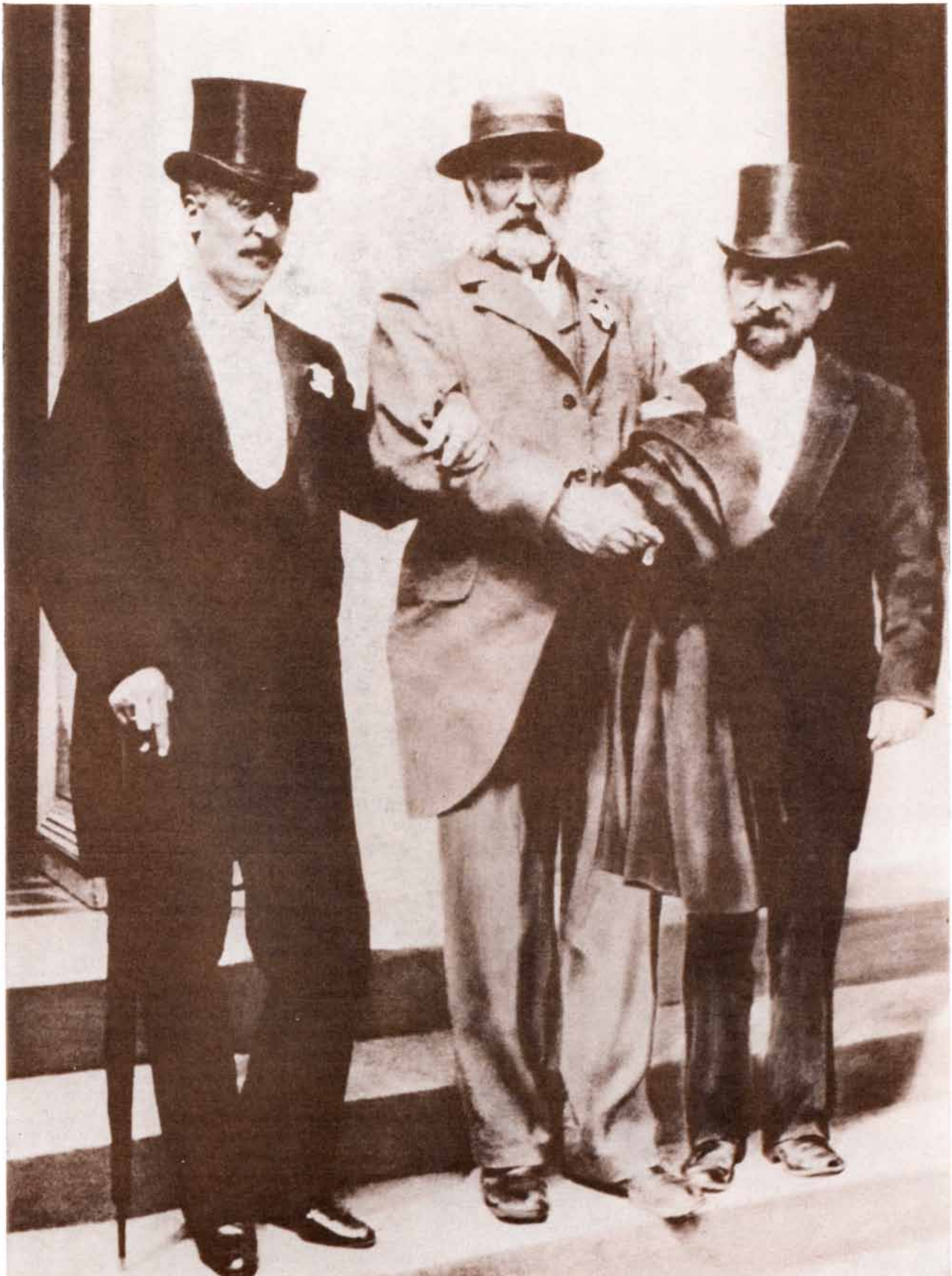
This new engine Diesel thought of as a Carnot engine. It was supposed to operate on the principles outlined more than half a century earlier by the French pioneer in thermodynamics Nicolas Léonard Sadi Carnot. Diesel's first calculations indicated that by following the ideal Carnot cycle his engine could convert 73 percent of the energy of coal—or almost any other fuel—into work. This was (and still is) an extremely high efficiency for a heat engine; in Diesel's day the best steam engines seldom showed a thermal efficiency higher than 7 percent. Even after making realistic allowances for the difficulties of handling the high pressures his engine required, Diesel was confident that it would need only a sixth as much fuel as a comparable steam engine would consume.

Like many another achievement in engineering, the Diesel engine we know today is quite different from the inventor's original ideal. It is the result of prolonged wrestling with practical problems, of painful and expensive incremental evolution, and not the quick breakthrough from theory to hardware that Diesel had envisioned. And it turned out to be not a Carnot engine at all. The story of the engine's development is nonetheless an unusual chapter in the history of technology because it began with an attempt to apply pure science to engineering. The inventor failed to realize his scientific ideal, but his failure was the first step in the evolution of an important new type of engine.

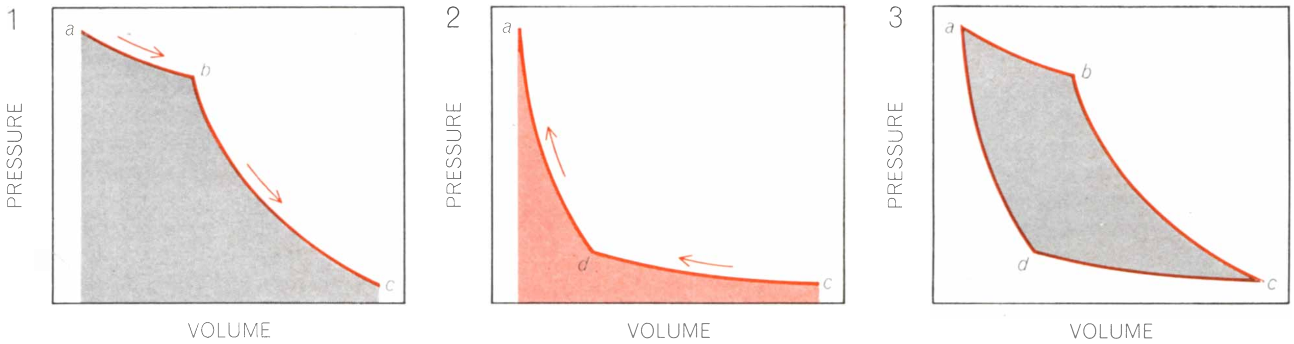
Rudolf Diesel was a German born in Paris who was displaced by the Franco-Prussian War and received his technical education in Germany. In 1880 he was graduated at the head of his class from the Technische Hochschule in Munich, and he entered the refrigeration-machinery business as the Paris representative of a firm founded by his teacher of thermodynamics, Carl Linde. Refrigeration was an exciting new field in those days, a field on the frontier of scientific engineering. Diesel's work gave him broad experience with heat engines and heat pumps, and a chance to work in his primary field of interest: thermodynamics. He was an extremely ambitious young man who fully intended to do for the approaching 20th century what James Watt had done for the 19th—to revolutionize energy conversion.

During the next 10 years Diesel devoted his spare time to a number of private energy-conversion projects, such as a sun-powered air engine and an ammonia bomb. As a refrigeration engineer he was familiar with the various working fluids used in refrigeration plants, such as ammonia, ether and carbon dioxide, and he made a special study of the possibility of using such working fluids in heat engines. A heat engine converts heat into work by adding heat to a working fluid—usually a gas—so that the fluid expands and exerts pressure on a piston or on turbine blades. In Diesel's time steam was the most common working fluid, but in theory any gas can serve as the medium for this kind of energy conversion. The efficiency of the process, as Carnot pointed out, does not depend on the choice of medium, but obviously some gases have more convenient properties than others. A steam power plant adds heat to the working fluid by external combustion, using a firebox to generate heat, a boiler to add heat to the working fluid and a complex system of pipes, perhaps including a condenser, to transfer the fluid between the boiler and the engine. All this equipment is in addition to the engine proper: the cylinder in which the steam does its work by expanding against a piston.

In the 1880's, when Diesel began experimenting with heat engines, hundreds of inventors were trying to find a substitute for steam that would not need such an elaborate and expensive system;



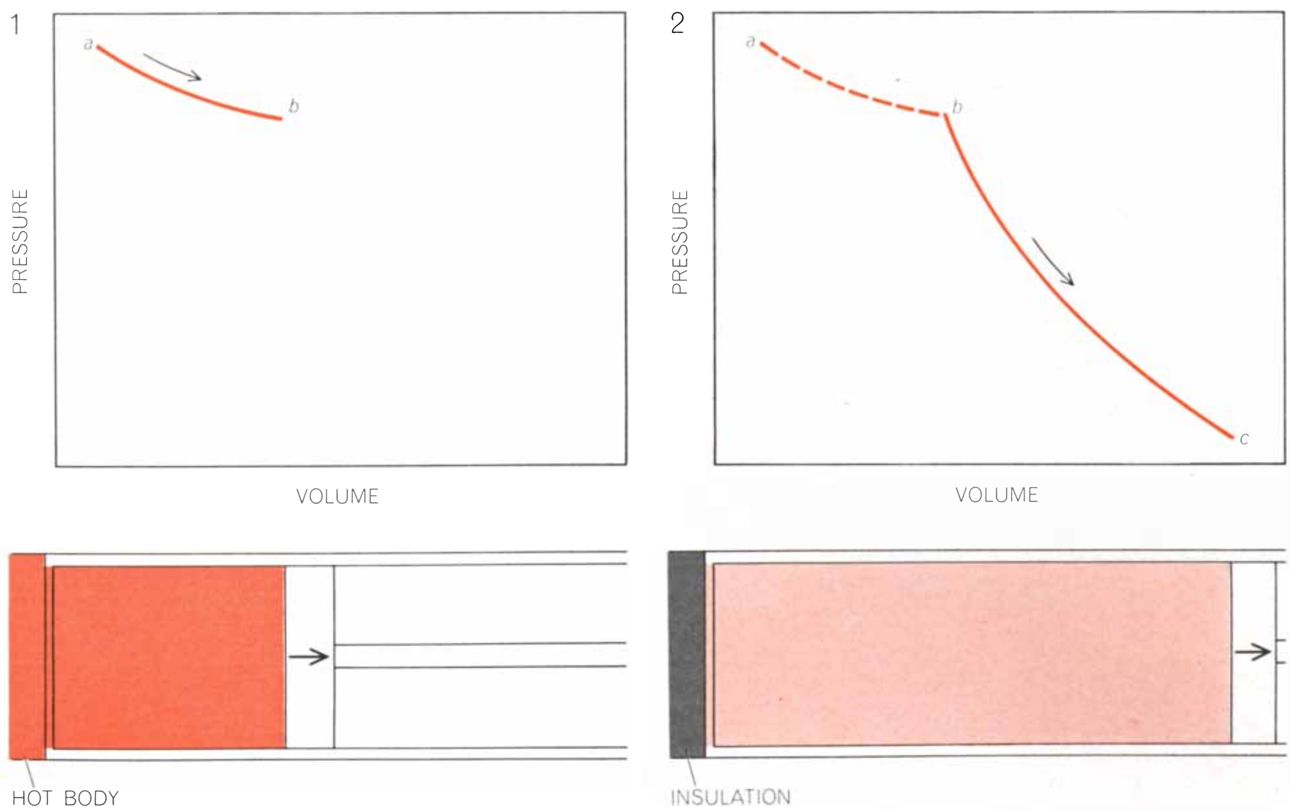
RUDOLF DIESEL (*left*) is seen with two colleagues at the time he prematurely announced successful development of his engine to the Society of German Engineers in 1897. With him are Heinrich von Buz (*center*) of the Maschinenfabrik Augsburg, one of the firms that supported Diesel, and Moritz Schröter, a professor of machine design who pronounced the engine a scientific success.



NET WORK accomplished by a heat engine is shown graphically by means of diagrams that plot pressure and volume of the "working fluid" inside the engine's cylinder as the piston first moves away from the cylinder head and then returns to its starting position. At the start (*1*), pressure is maximum and volume minimum (*a*); the curve *a–b–c* traces the decrease in pressure and rise in volume as the heat of the working fluid is converted into work during the fluid's expansion against the piston. The area below the curve (*gray*) represents "positive work." A portion of this work is needed to drive the piston back to its starting position (*2*). The increase in pressure and decrease in volume during the piston's return are traced by the curve *c–d–a*. The area in color below this curve represents "negative work." Subtraction of the negative work from the positive work (*3*) produces an area indicative of the theoretical "net work" accomplished within the cylinder during a single operating cycle. In a real engine this indicated area of net work will be reduced by various kinds of thermal and mechanical losses before it becomes available on an output shaft to drive an external load.

they were seeking a working fluid that would make possible more efficient or more economic power plants, particularly in small sizes. Air was the most popular fluid tried as a substitute for steam. At this time thousands of small external-combustion engines using air instead of steam as the working fluid were in common use. Another possibility that appealed to Diesel was ammonia, the fluid used in Linde's refrigeration machinery, which has a number of interesting properties Diesel thought he could exploit in an engine. For 10 years he pinned his hopes on a small ammonia engine that was supposed to store energy and deliver work in small quantities like a storage battery. Ammonia is so difficult to handle in an engine, however, that eventually he gave it up and turned to air.

Air has two unique advantages as a



DIESEL'S DREAM, an engine that would perform according to Carnot's ideal cycle, is shown schematically; the pressure-volume diagrams at the top trace the action of working fluid and piston in an imaginary cylinder at the bottom. At the start of the cycle (*1*), a large heat reservoir is in contact with the cylinder head, and heat flowing from it into the working fluid causes the fluid to expand (*a–b*) isothermally, that is, without an increase in temperature. In the next step (*2*) the heat source is removed and the cylinder head is insulated; the working fluid continues to expand (*b–c*). The expansion is adiabatic, that is, without the flow of heat to or working fluid for an engine. One is that it is so easily available that an engine can take in a fresh supply for each cycle. Another advantage, which Diesel said occurred to him about 1890, is that it contains oxygen. This means that in an air engine combustion can take place *inside* the working fluid, without any need for a firebox and a boiler. Air can have two functions in a heat engine: it can serve as the medium of energy conversion and at the same time it can participate in the generation of heat.

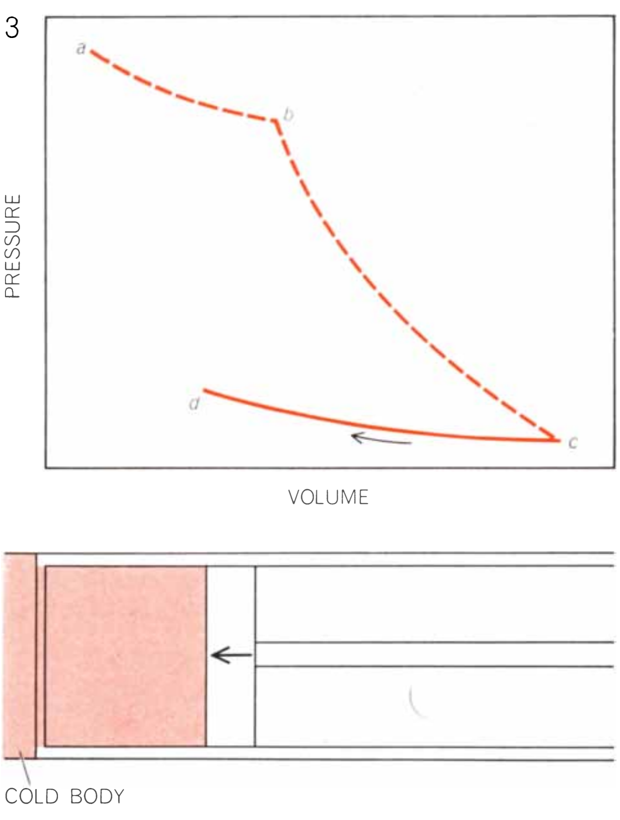
This is the essential concept of internal combustion. Diesel was by no means the first to think of it; his hero Sadi Carnot had suggested long before that air is the most promising substitute for steam because its oxygen content makes internal combustion possible, and a hundred inventors had worked on this idea before Diesel. The most successful was Nicolaus Otto, whose internal-combustion engine burning illuminating gas was well developed and widely used in the 1880's. The unique aspect of Diesel's approach was his conviction that his engine could realize the ideal Carnot cycle.

The Carnot cycle was first described in a small book titled *Reflections on the Motive Power of Heat and on Machines Fitted to Develop That Power,* published by Carnot in 1824. The book contains a number of basic insights into the nature of heat and the principles underlying the operation of heat engines. In those days the relations between heat and work were not well understood. As Count Rumford had demonstrated in his famous cannon-boring experiments, work could produce heat without limit, but a good deal of uncertainty remained about the reverse relation: the conversion of heat into work. A steam engine obviously converted heat into work, but nobody knew exactly how the work was related to the heat. In Carnot's time heat was regarded more as a kind of substance than as a form of energy. From this point of view it seemed that heat passed through a steam engine without essential change, just as water passes through a water mill. It was not generally understood that when heat is converted into work, it disappears as heat and reappears as work.

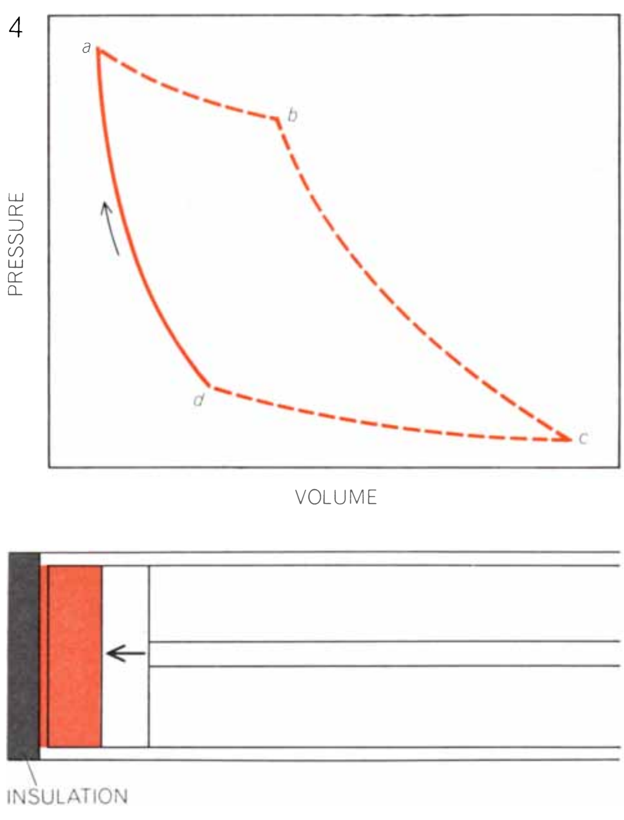
In his book Carnot showed that a heat engine can convert into work only a certain fraction of the heat supplied to it. This fraction, which is called thermal efficiency, depends on the temperature range through which the engine operates and not on the kind of working fluid used. Carnot also proved that the efficiency of a process is at a maximum when the process is *reversible.* A process, for example, that could convert heat into work and also operate in the reverse direction to convert the same amount of work into the same amount of heat would be reversible. If a water mill could extract enough energy from a certain fall of water to pump the same amount of water back up to the original level, the process would be reversible and the water mill would be perfect. Reversibility in this sense is an unattainable ideal. The concept is a way of defining the upper limit to the efficiency of energy conversion. No heat engine operating between given temperatures can possibly be more efficient than one operating on a cycle with all processes reversible.

Carnot went on to describe such an ideal cycle, with four reversible processes, and to show how it would work in an ideal engine. He imagined a cylinder containing a fixed quantity of working fluid, say air, and a piston moving back

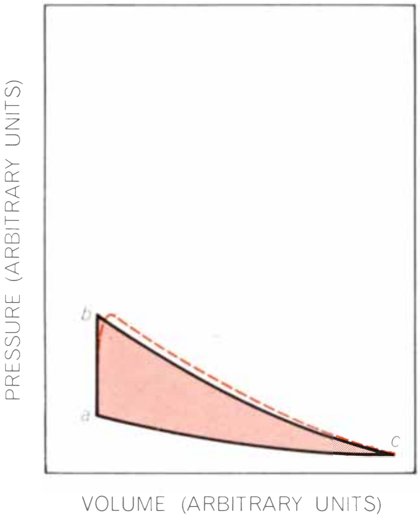
3

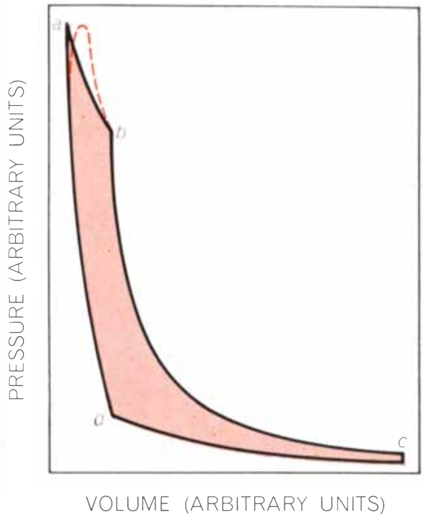


4

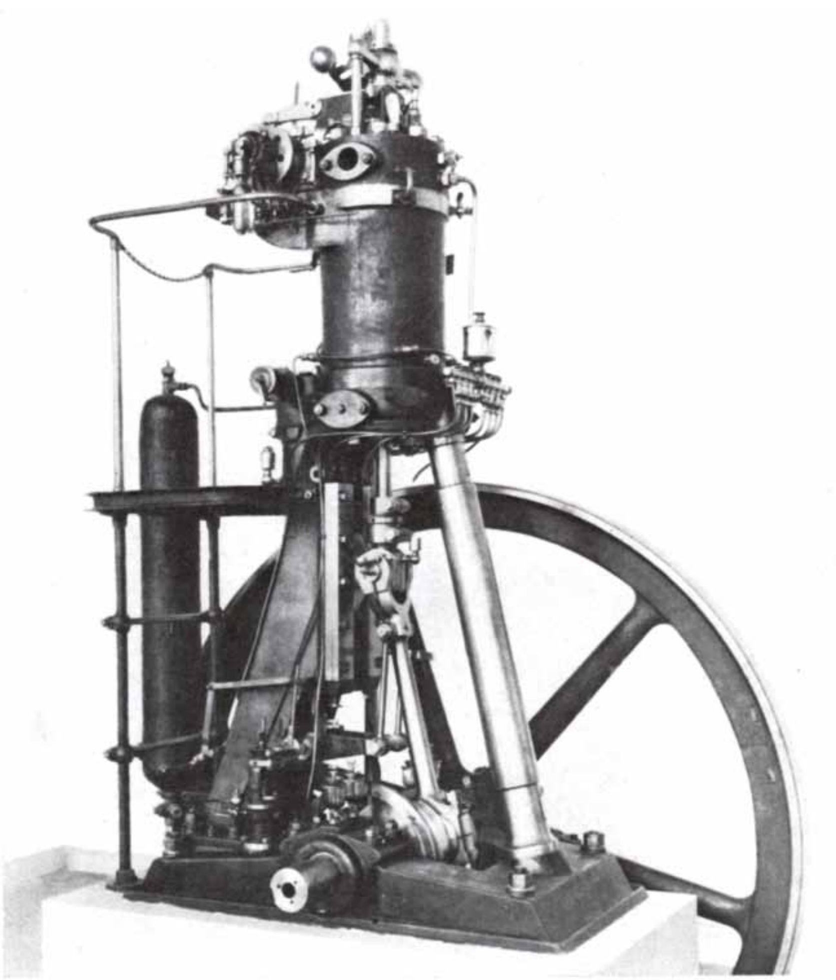


from the fluid, and the temperature of the fluid drops. Next the piston must be driven back, compressing the fluid before it. First the cylinder head is placed in contact with a cooler heat reservoir (*3*), so that the heat of compression flows from the fluid to the cold body; the temperature of the fluid does not rise and the compression is isothermal. Finally (*4*) the cold body is replaced by insulation and the piston is returned to the starting position (*d–a*). This process is adiabatic, and the heat generated by the piston's work raises the temperature of the working fluid to its original level, thereby completing the cycle. Each step of the cycle is reversible.





TWO PRESSURE-VOLUME DIAGRAMS, included in Diesel's first patent, compare the cycle of the Otto engine with the proposed Diesel cycle. In the Otto cycle (*left*) pressure rises sharply after ignition (*a*). Combustion is almost instantaneous, so that heat is added at roughly constant volume (*a–b*). The Diesel diagram (*right*) shows peak pressure reached solely through the compression of air (*c–d–a*). Ignition comes at *a*, and combustion continues from *a* to *b*, adding heat isothermally in accordance with the Carnot ideal. Solid lines in both diagrams show the ideal cycle; broken lines, the process expected in an actual engine.



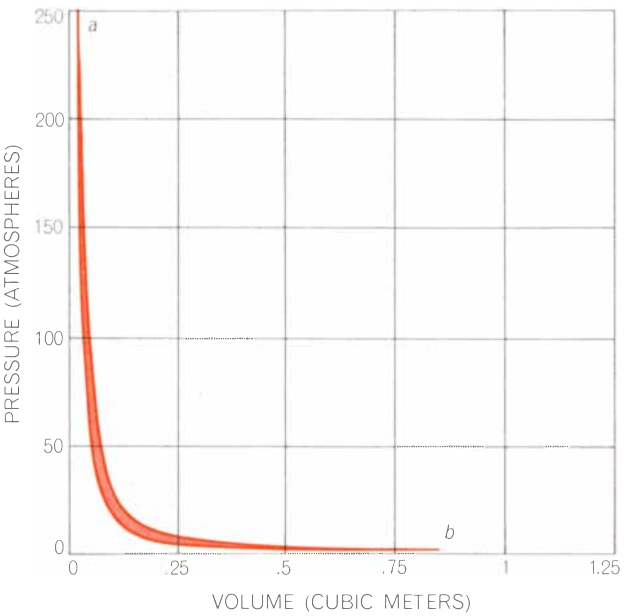
DIESEL'S FIRST SUCCESS was this one-cylinder oil-fueled engine, which in 1897 achieved a compression of about 30 atmospheres and had a thermal efficiency of 26 percent. Diesel's compound engine, which was far less efficient to operate, was junked the same year.

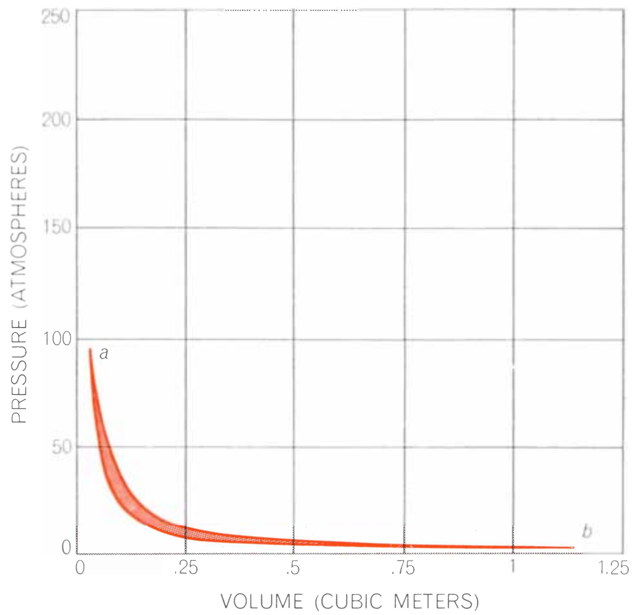
and forth in the cylinder. He also imagined two very large bodies, a hot one and a cold one, that can be placed alternately in contact with the cylinder. The hot body is a reservoir of heat from which the heat to be converted into work is taken. The cold body is a reservoir into which the engine discharges the heat it cannot use.

At the beginning of the cycle the cylinder is in contact with the hot body, and the temperature of the working fluid is assumed to be very close to the temperature of the hot body. Heat flows from the hot body into the working fluid and causes it to expand and drive the piston a certain distance along the cylinder. This expansion of the fluid against the resistance of the piston is a conversion of heat into work. The transfer of heat from the hot body to the fluid is assumed to be isothermal, that is, it takes place without changing the temperature of the fluid, and the hot body is assumed to be so large that it can supply heat without any appreciable drop in temperature. Actually there has to be some difference in temperature if heat is to flow at all, but the difference can be very slight. This first process in the expansion stroke is reversible because it is isothermal. If it were a flow of heat from a hot body to a cold one, it would not be reversible, because heat will not flow from a cold body to a hot one.

Contact with the hot body is now broken to start the second process. The fluid continues to expand, but because there is no further flow of heat into it the temperature of the fluid declines until at the end of the expansion stroke the temperature of the fluid has dropped to a point only slightly higher than the temperature of the cold body. During this process some of the energy stored in the fluid is converted into the work represented by the motion of the piston. This kind of expansion, without any heat flow into or out of the working fluid, is called adiabatic. It is also a reversible process, because it can go in the reverse direction: the piston can also compress the fluid and raise its temperature.

This is the end of the expansion stroke, or power stroke. Now if the engine is to run continuously, the piston must be brought back to its original position. This return trip, the compression stroke, is going to cost something: it will take work to drive the piston back through the cylinder, compressing the working fluid ahead of it until the piston reaches its original position and the working fluid reaches its original pressure and temperature, ready to repeat the cycle. In a





LATER PAIR OF DIAGRAMS show the problems Diesel faced when he calculated the actual pressures and volumes that would be required for a Carnot engine. The first diagram (*left*) shows the range needed to reach a temperature of 800 degrees Celsius at maximum compression (*a*) and 20 degrees C. at maximum expansion (*b*). The pressure needed (250 atmospheres, or 3,675 pounds per square inch) is much higher than was practical in Diesel's day. Moreover, the net-work area (*color*) was dangerously slender. On recalculation Diesel found that he could cut the pressure requirement to 90 atmospheres ("*a*" *at right*) at a cost of only 5 percent in theoretical efficiency. Now he saw another problem: The long thin tail at the right in the diagram (*b*) means that complete expansion would require a very large cylinder and would make little contribution to the net work. Diesel compromised again and cut off the tail.

real engine this work will come from energy stored, perhaps in a flywheel, during the power stroke. In a Carnot engine the compression must be accomplished reversibly.

At the beginning of the compression stroke the cylinder is placed in contact with the cold body, so that during the first part of the stroke the work of compression is converted into heat, all of which flows into the cold body, so that the temperature of the working fluid does not rise. The cool reservoir is assumed to be so large that it can receive the heat without a rise in temperature. This is isothermal compression, a reversible process.

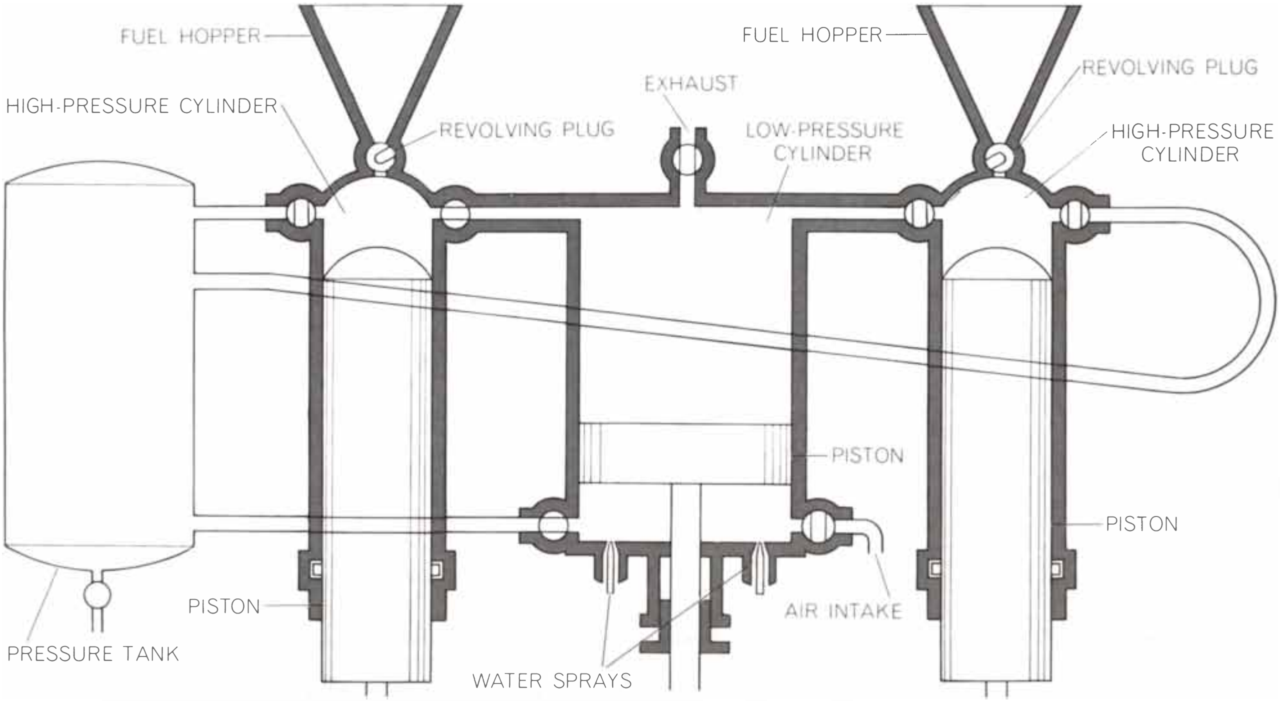
Contact with the cool body is now broken, and compression continues with no flow of heat into or out of the working fluid. In this final process of the cycle the work of compression is converted into heat, which raises the pressure and the temperature of the working fluid to its original level. This is adiabatic compression, also a reversible process.

The ideal Carnot cycle can be illustrated by pressure-volume diagrams of the kind shown in the top illustration on page 110. The vertical axis measures pressure, and the horizontal axis measures the changing volume as the piston passes back and forth in the cylinder. Engineers have used diagrams of this kind, called indicator diagrams, to study the performance of piston engines ever since James Watt invented an indicator mechanism that would record the changing volumes and pressures inside the cylinder of a real engine.

In the ideal pressure-volume diagram of the Carnot cycle the upper line indicates the declining pressure as the piston is driven out to the right by the expanding working fluid. This is the expansion curve. The area under the curve, the product of pressure and change in volume, represents the work done by the fluid on the piston during the expansion stroke. This is *positive* work. The lower curve shows the rising pressures as the piston returns to its original position. The area under the curve represents work done by the piston on the fluid in order to bring it back to its original pressure and temperature. This is the *negative* work. In a real engine one hopes that the positive work is larger than the negative work, that is, that the expansion stroke produces at least enough power to drive the piston back. The difference between positive and negative work is the area between the two curves, which represents the *net* work. The larger this area is, the more work the engine can do.

This is the Carnot cycle, which Diesel studied in school, as all engineers still do in their thermodynamics courses, and these are the kinds of diagrams Diesel used in his study. The Carnot cycle, however, is a theory, not an engine. The question for an idealistic, ambitious engineer such as Diesel was how the ideal could be approached in a practical engine. It would not be hard, he might have said to himself, to approximate adiabatic expansion and compression in a piston engine. The expansion and compression of air in a cylinder as a piston moves back and forth would normally be more or less adiabatic, if the piston moves fast enough, and the flow of heat through the cylinder walls could be retarded by insulation. Isothermal compression could also be achieved by known techniques. For example, water injected into a gas as it is being compressed could remove heat so that the temperature does not rise. It is not easy, however, to see how in a real engine heat can be added at the highest temperature of the cycle without raising the temperature further, as the Carnot cycle requires, particularly if the heat comes from internal combustion. Who ever heard of combustion that does not raise the temperature of the air in which it takes place?

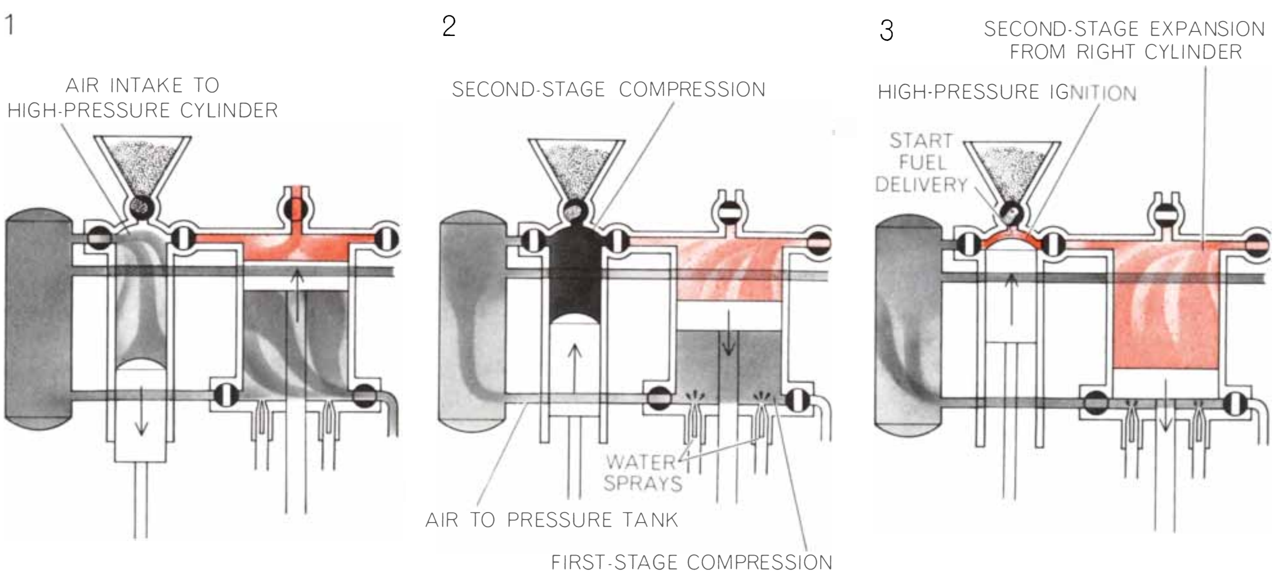
This is precisely what Diesel thought he had discovered, a way to achieve isothermal combustion. The key idea was this: His engine would reach the highest temperature of its cycle exclusively as a result of compressing the air. Having reached that point, he would add to the



COMPOUND ENGINE, splitting compression and expansion into separate low-pressure and high-pressure processes, was Diesel's first proposal for approximating the Carnot cycle. Low-pressure processes occurred in the large cylinder and high-pressure processes in the smaller cylinders on each side. The operating cycle of the engine is shown in the illustration at the bottom of these two pages.

very hot air a small quantity of fuel at a controlled rate, so that the natural tendency of the air temperature to rise as the fuel burned would be exactly counterbalanced by the tendency of the temperature to fall as the air expanded with the movement of the piston. The amount of fuel would have to be very small and the amount of air very large (nearly 10 times the minimum required to support combustion), but if the idea succeeded, *all* the heat developed from the fuel would be transformed into work during the isothermal part of the expansion stroke. If this process could be achieved, an approximation of a Carnot engine was possible.

In 1892 Diesel applied for a patent on his engine, giving isothermal combustion as the essence of the invention. The patent was granted. To show the differ-



CYCLE OF COMPOUND ENGINE requires the two small high-pressure pistons (one not shown) to work 180 degrees out of phase with the double-acting, large low-pressure piston. The small pistons' cycle begins with a downstroke (*1*), drawing a charge of air that has already gone through first-stage compression from a pressure tank (*far left*). The following stroke (*2*) compresses this air further. Near the top of this stroke (*3*), at maximum compression, fuel begins to spill into the very hot air and ignites. Combustion continues during the first part of the following downstroke (*4*). This stroke effects first-stage expansion. On the next stroke (*5*) the expansion enters a second stage as the small piston rises and the large piston falls. The next upstroke of the large piston (*6*) ence between his engine and the common Otto type, the patent has two pressure-volume diagrams. In the Otto engine combustion is practically instantaneous, so that the pressure of the working fluid rises steeply after ignition and then declines slowly as the volume of the fluid increases. In Diesel's engine, however, combustion begins at the peak pressure, and it continues, Diesel says in the patent, essentially without increase in temperature or pressure. The diagrams in the patent give no indication of scale, but Diesel's engine seems to have a much higher maximum pressure than Otto's, and the area of net work in the Diesel diagram is broad enough to suggest a good power output.

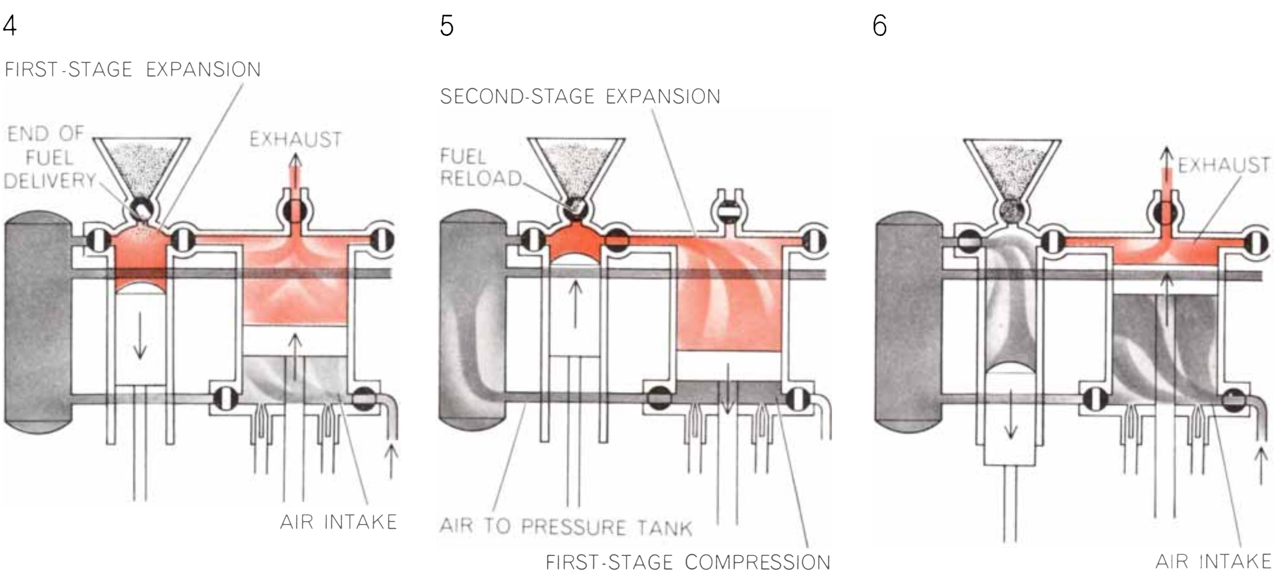
In the same year that he applied for a patent Diesel also described his ideas at greater length in a manuscript that included detailed calculations and supporting drawings. He sent copies of this study to several experts for criticism. Most of the critics were much interested in the theory, but they foresaw great practical difficulties. Diesel's original proposal was for an engine that would compress air to a pressure of 250 atmospheres and a temperature of 800 degrees Celsius. If the expansion could be carried down to a pressure of one atmosphere and a temperature of 20 degrees, as Diesel hoped, then the theoretical efficiency, on Carnot's formula, would be 73 percent. Carnot had shown that the maximum efficiency possible in a heat engine operating in a given temperature range is the difference between the highest and the lowest temperatures of the cycle divided by the highest temperature (all the temperatures being given on an absolute scale such as degrees Kelvin). In Diesel's first proposal the highest temperature would be 1,073 degrees K. (800 degrees C.) and the lowest 293 degrees K. (20 degrees C.). The difference is 780, which divided by 1,073 gives a figure for maximum possible efficiency of 72.7 percent.

The maximum pressure of 250 atmospheres that Diesel was proposing was far beyond the state of the art at the time. In a piston engine this would require a compression ratio of 60 to one. (Modern gasoline-engine compression ratios are about eight to one and modern Diesel ratios about 16 to one.) The prospect of dealing with such unprecedented pressures was naturally disturbing to technical critics and potential backers.

When Diesel plotted the pressure-volume diagram for his first proposed engine, his calculations produced an extremely narrow diagram [*see illustration on page 113*]. This might easily have alarmed a practical man. Not only was the pressure extreme but also the negative-work curve lay dangerously close to the positive-work curve. This meant that a large fraction of the work developed during the expansion stroke would have to be used to bring the piston back through the compression stroke.

In response to criticism along these lines, Diesel calculated the curve for an engine with more modest specifications, and he found he could reduce the peak pressure at the start of the cycle to 90 atmospheres—a reduction of almost two-thirds—at a cost of a mere 5 percent reduction in thermal efficiency. He then added a section to his manuscript describing a compromise engine with a thermal efficiency of only 68 percent, and he published the work early in 1893 under the sweeping title *Theory and Construction of a Rational Heat Engine to Take the Place of the Steam Engine and of All Presently Known Combustion Engines*. He sent advance copies to engineers and industrialists all over Europe.

From the book's detailed drawings and faultless mathematics the casual reader might easily have gained the impression that Diesel's engine already existed, but Diesel had no hardware at all, only an idea. He had to publish the book first so that he could attract the financial support he needed to develop a real engine. In order to convince manufacturers of heavy machinery that his project should be underwritten he first had to gain the endorsement of the scientific establishment. The book therefore had a strong promotional flavor, including speculation on the social and economic effects of a perfected-rational engine. The engine's efficiency, Diesel declared, would make it economic even in small sizes, which the steam engine was

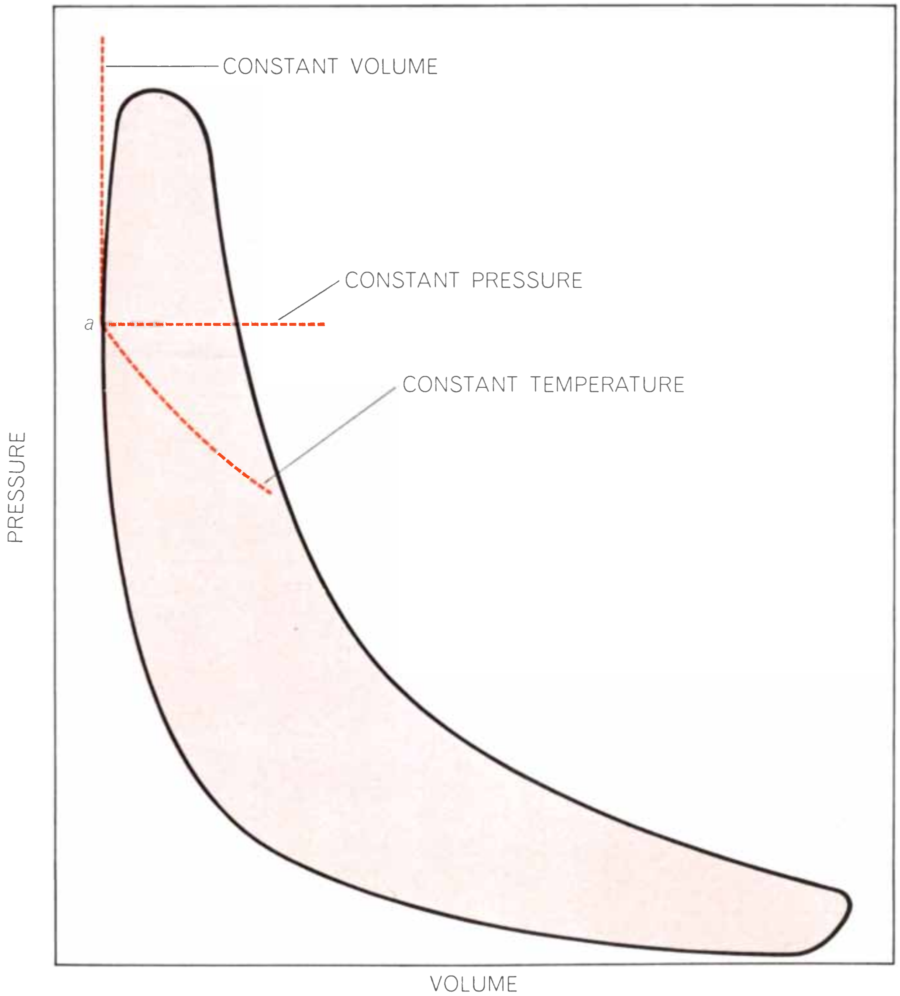


drives out the exhaust at atmospheric pressure and at the same time starts the next cycle by drawing in a new charge of air. The large central cylinder serves both high-pressure cylinders; each upstroke of its piston (*1, 4 and 6*) draws fresh air into the lower part of the cylinder and drives exhaust from the upper part. Each downstroke (*2 and 5*) performs first-stage compression in the lower part of the cylinder, forcing the compressed air through jets of water into the pressure tank, and at the same time provides room in the upper part of the cylinder for second-stage expansion of the gases from one or the other of the small cylinders. For the large cylinder this is a two-stroke cycle, for the small cylinders a four-stroke cycle and for the whole system a seven-stroke one.

not. Such an economic small power plant would enable industry to decentralize, he thought, and it would perhaps restore the small craftsman to the position of significance he had lost with the advent of steam. Diesel also outlined some of the effects the new type of power could be expected to have on railroads and ships.

Diesel's book made a good case for the Carnot engine, good enough to convince nearly all the leading European authorities on thermodynamics that his theory was sound, although there might be difficulties putting it into practice. Even the great Lord Kelvin gave Diesel's project his blessing. These endorsements, together with Diesel's eloquence, induced two major German firms—Krupp in Essen and the Maschinenfabrik Augsburg—to back the project. Both were major producers of steam engines; their support may have been a prudent hedge against the chance that Diesel's scheme would work.

The form Diesel had originally chosen for his engine was a curious three-cylinder arrangement. He realized that if he tried to accomplish all the expansion and compression that theory demanded in a single cylinder, the cylinder would have to be impossibly large. Therefore he planned a compound engine with both compression and expansion accomplished in two stages in separate cylinders. The low-pressure processes were to be accomplished in a large central cylinder and the high-pressure ones in two smaller cylinders that flanked it. Conical hoppers on top of the small cylinders would be filled with powdered coal; the fuel would be added to the air near its point of maximum compression by means of ingenious rotating plugs that would supply just enough coal at the proper rate to produce isothermal combustion.



PRESSURE-VOLUME DIAGRAM of the 1897 Diesel engine (*color*) has the same shape as diagrams of modern Diesel engines. The point of ignition (*a*) indicates the start of the combustion process that adds heat to the working fluid. If Diesel's hope of adding heat isothermally could have been achieved, as it was not, the lowest of the three broken lines shown in color would have bounded the top of the pressure-volume diagram. In actuality addition of heat in a Diesel engine is a compromise between a constant-pressure process (*middle curve*) and the constant-volume process (*top curve*) characteristic of Otto engines.

Diesel's plan made no provision for igniting the fuel because it was assumed that almost any combustible substance would burst into flame the moment it came in contact with the very hot air. People did not understand combustion very well in those days, and Diesel found he had been much too optimistic about the possibility of using almost any fuel—solid, liquid or gaseous. He never succeeded with either coal or illuminating gas. He had a great deal of trouble with ignition in general; at one point he despaired of compression ignition altogether and began to work on various kinds of ignition devices.

Today compression ignition is considered a major defining characteristic of the Diesel engine, but Diesel always insisted that it was only an incidental feature. The essence of the invention, he held, was isothermal combustion; it made no difference how the combustion was started. One measure of his conviction is that Diesel's first plans made no provision for cooling the engine. Combustion would not raise the temperature of the working fluid, as it does in other combustion engines, and no heat would be wasted, Diesel thought, so that no cooling would be necessary. In fact, Diesel intended to insulate his cylinders to cut down heat losses.

What has been described so far is the idea for an engine and not the engine that became a reality. Diesel knew he would have to compromise with his ideal at first, and he was ready with plans for a simplified engine as soon as his backers agreed to support the development. Experimental work began in 1893 with a one-cylinder oil-burning engine that was supposed to develop 25 horsepower with a maximum compression of 70 atmospheres. To keep the size of the cylinder within reason the expansion stroke was cut short long before the pressure and temperature of the working fluid fell to the point required by theory, and isothermal compression was not attempted. Diesel was still hoping for isothermal combustion when he planned this engine, and at first he made no provision for cooling.

Early in his experimental work Diesel found that in order to get his engine to run at all he would have to use more fuel than his mathematics called for. It turns out that an internal-combustion engine cannot realize the Carnot cycle. The isothermal combustion process requires large quantities of air, much more than is necessary for combustion. The work of handling the extra air and compressing it to the high pressures required cannot be done by amounts of fuel small enough to keep the combustion isothermal. The rational engine might have a very high internal efficiency, but it would not develop enough power to turn itself over. The pressure-volume diagram is simply too thin. The slightest displacement of the compression curve, such as might be caused by internal friction, would bring it right up to the expansion curve, which means that the engine could do no net work at all. The work done on the piston during the expansion stroke would scarcely be enough to bring the piston back to its starting position.

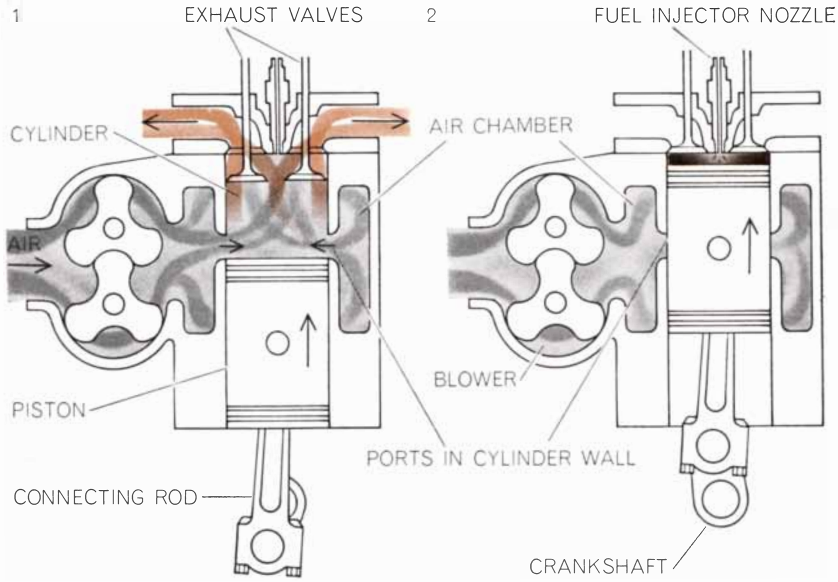
Diesel returned to his theoretical studies and wrote a long explanation of the need for making certain adjustments that would increase his engine's power. He saved as much of his theory as he could, but he acknowledged that isothermal combustion was out of the question. He saw that instead of adding heat at a constant temperature during the expansion stroke he would have to add it at approximately constant pressure.

Diesel intended to publish his explanation in a second edition of his book but there never was one. He did, however, apply for a supplementary patent that covered various devices for regulating fuel flow in his engine so that the diagram could be broadened.

The patent implies that this broadening can be achieved by an adjustment in the rate of fuel injection, but it actually requires larger amounts of fuel. This means abandoning the constant-temperature process, with isothermal combustion, and moving in the direction of a constant-pressure process, although Diesel did not say so. He clung to the ideal of the Carnot cycle as a symbol, and on ceremonial occasions he continued to refer to his engine as an approximation of the ideal Carnot engine, although he knew that the temperature of the working fluid must have risen 1,000 degrees C. after ignition. In those days uncertainties about temperatures inside engines were understandable; they could not be directly measured, and the pressure-volume diagrams used (like the ones in this article) say nothing directly about temperature. Temperatures would have to be calculated from the rough data taken from indicator diagrams, and even in a slow-moving engine the temperature would fluctuate through a range of 1,000 degrees in less than a second.

The development of the real Diesel engine is another story. It was a long and costly struggle to find ways of coping with high pressures. The problem of injecting fuel into highly compressed air proved to be extremely difficult. The first one-cylinder experimental engine never ran at all under its own power, but it gave the inventor sobering experience. It was rebuilt several times. After four years of hard work Diesel formally (and quite prematurely) announced the successful development of his engine at a meeting of German engineers in 1897. The engine he exhibited had a thermal efficiency of 26 percent with a compression of 30 atmospheres. It was a very respectable achievement, but the engine was by no means reliable or economic.

After the first year or two of practical work the Carnot cycle ceased to play any real part in the development of the Diesel engine. Diesel was nonetheless committed to it by his patent and his professional pride. As the simplified one-cylinder engine moved closer to practical reality the inventor continued to work on his scheme for the three-cylinder model. This engine, he hoped, would come much closer to the Carnot ideal because it would incorporate both the high initial pressures and the full expansion of the working fluid he had originally intended. A large compound engine was finally built, more or less according to Diesel's original plan, except that it burned oil. Its performance was far from ideal. Its fuel consumption per horsepower-hour was twice the consumption of the one-cylinder engine, and it developed only half the anticipated horsepower. Its chief problem was heat losses. The essential advantage of an internal-combustion engine is that the working fluid need not be shunted from place to place; the entire cycle is confined to the combustion chamber. In Diesel's compound engine, however, the working fluid was first pumped from the large cylinder to the pressure tank, then on to the small cylinders and finally back to the large cylinder through pipes and valves that lost heat. In 1897, the same year Diesel announced the successful one-cylinder engine, the compound engine was junked.



MODERN DIESEL ENGINE is seen here in a two-cycle model. A blower (*left*) is the source of high-pressure air that also helps to drive out exhaust gases after the power stroke (*1*). Near the end of the compression stroke (*2*) fuel is injected hydraulically and ignites.

The Diesel evolved slowly over the next 25 years. By 1910 a number of ships were being propelled by Diesel engines, and the submarines of World War I were mostly Diesel-powered. These first engines were heavy and spent an appreciable fraction of their output driving the air compressors then used for fuel injection. The modern Diesel, a mature and economic engine, did not appear until the 1920's, when fuel-injection devices were perfected. The Diesel of today bears little resemblance to the original rational engine. Nonetheless, three of its essential features are the same as those Diesel envisioned in 1892: it is a high-compression engine with air as the working fluid, the fuel is injected near the end of the compression stroke and the fuel is ignited by the heat of compression. The engine is economic both because the heavy oil it burns is cheap and because its thermal efficiency is high. The efficiency is up to 40 percent for the best Diesels—about the same as the figure for the most efficient large steam turbines. But the combustion is not isothermal.

MATHEMATICAL GAMES

Simplicity as a scientific concept: Does nature keep her accounts on a thumbnail?

by Martin Gardner

Simplicity, simplicity, simplicity! I say, let your affairs be as two or three, and not a hundred or a thousand; instead of a million count half a dozen, and keep your accounts on your thumb nail.

—Henry David Thoreau, *Walden*

Thoreau's advice raises deep and difficult questions regarding the universe. Does nature keep her accounts on a thumbnail? Are the basic laws of science few in number or, as the mathematician Stanislaw Ulam believes, infinite? Are they, as the physicist Richard Feynman has suggested, perhaps finite but increasingly hard to discover, so that there will always be a tiny but forever shrinking margin of mystery about the universe?

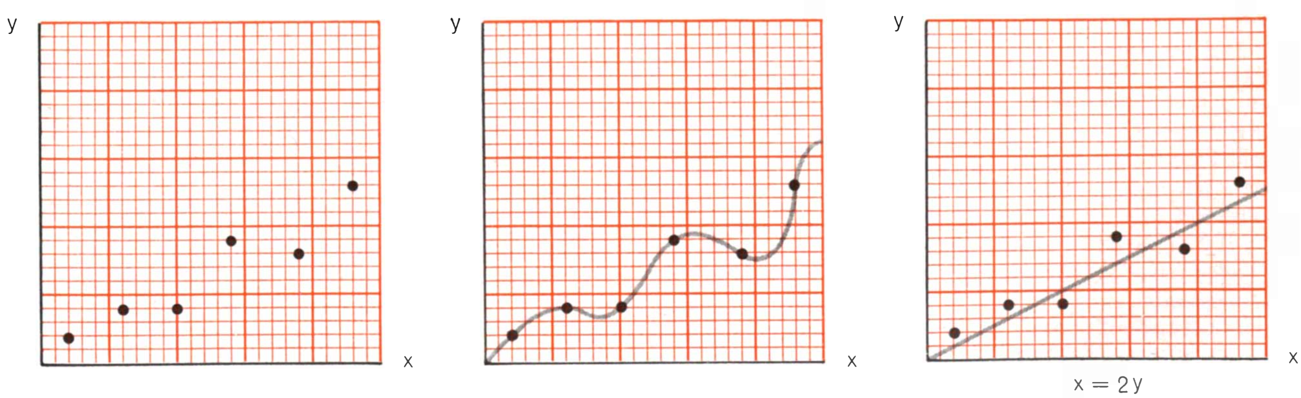
A closely related question is whether the basic laws themselves are simple or complicated. Most biologists, particularly those working with the brain and the nervous system, are impressed by the complexity of life. In contrast, although quantum theory has become enormously more complicated with the discovery of weird new particles and interactions, most physicists retain a strong faith in the ultimate simplicity of basic laws. This was especially true of Albert Einstein. "Our experience," he wrote, "justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas." When he chose the tensor equations for his theory of gravitation, he picked the simplest set that would do the job, then published them with complete confidence that (as he once said to the mathematician John Kemeny) "God would not have passed up an opportunity to make nature that simple." It has even been argued that Einstein's great achievements were tied up with a psychological compulsion, like Thoreau's, to simplify his personal life.

"Einstein's bedroom was monkish," Peter Michelmore writes in *Einstein, Profile of the Man*. "There were no pictures on the wall, no carpet on the floor.... He shaved roughly with bar soap. He often went barefoot around the house. Only once every few months he would allow Elsa [his wife] to lop off swatches of his hair.... Most days he did not find underwear necessary. He also dispensed with pajamas and, later, with socks. 'What use are socks?' he asked. 'They only produce holes.' Elsa put her foot down when she saw him chopping off the sleeves of a new shirt from the elbow down. He explained that cuffs had to be buttoned or studded and washed frequently—all a waste of time."

"Every possession," Einstein said, "is a stone around the leg." The statement could have come straight out of *Walden*.

Yet nature seems to have a great many stones around her legs. Basic laws are simple only in first approximations; they become increasingly complex as they are refined to explain new observations. The guiding motto of the scientist, Alfred North Whitehead wrote, should be: "Seek simplicity and distrust it." Galileo picked the simplest workable equation for falling bodies, but it did not take into account the altitude of the body and had to be modified by the slightly more complicated equations of Newton. Newton too had great faith in simplicity. "Nature is pleased with simplicity," he wrote, echoing a passage in Aristotle, "and affects not the pomp of superfluous causes." Yet Newton's equations in turn were modified by Einstein, and today there are physicists, such as Robert Dicke, who believe that Einstein's gravitational equations must be modified by still more complicated formulas.

It is dangerous to argue that because many basic laws are simple the undiscovered laws also will be simple. Simple first approximations are obviously the easiest to discover first. The most one can say is that science sometimes increases simplicity by producing theories that reduce to the same laws phenomena previously considered unrelated (for example the equivalence of inertia and gravity in general relativity), but that science equally often discovers that behind apparently simple phenomena, such as the structure of matter, there lurks an unsuspected complexity. Johannes Kepler struggled for years to defend the circular orbits of planets because the circle was the simplest closed curve. When he finally convinced himself that the orbits were ellipses, he wrote of the ellipse as "dung" he had to introduce to rid astronomy of vaster



***Observed data* (left), *a possible function curve* (middle) *and the most likely function* (right)**

amounts of dung. It is a perceptive statement because it suggests that the introduction of more complexity on one level of a theory can introduce greater overall simplicity.

Nevertheless, at each step along the road simplicity seems to enter into a scientist's work in some mysterious way that makes the simplest workable hypothesis the best bet. "Simplest" is used here in an objective sense. Naturally there are many pragmatic aspects, but they are irrelevant to the big question. If two theories are identical in all respects except the way they are expressed (for example one in English measures, the other in the metric system), a scientist would be foolish not to pick the formulation that is easier to handle. When two theories are not equivalent—when they lead to different predictions—he also will prefer to test first whichever theory is easier to test. This depends, however, on what apparatus he has available, what kind of mathematics he understands best and so on. The same theory may seem simple to one physicist and complicated to another.

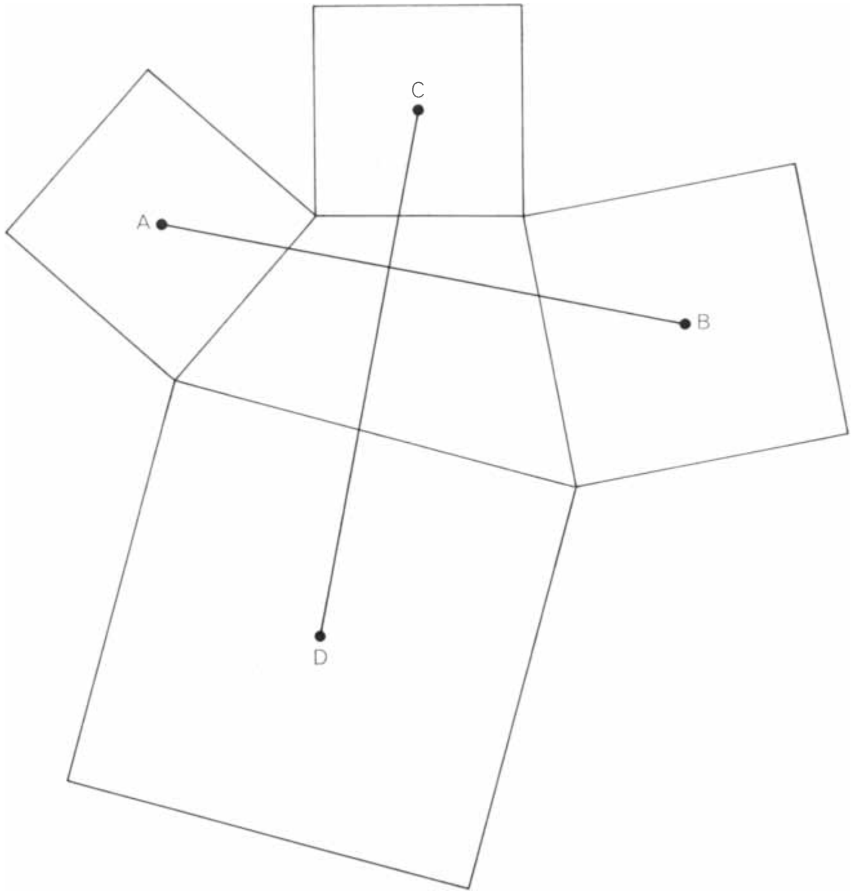
Although such subjective matters play their role, they fail to touch the heart of the mystery. The important question is: Why, other things being equal, is the simplest hypothesis usually the most likely to be on the right track? The classic instance is the graphing of a relation between two variables. The physicist records his observations as dots on a graph, then draws the simplest curve that comes close to those dots. Simplicity even overrules the data. If the spots fall near a straight line, he will not draw a wavy curve that passes through every spot. He will assume that his observations are probably a bit off, pick a straight line that misses every spot and guess that the function is a simple linear equation such as $x = 2y$ [*see illustration on opposite page*]. If this fails to predict new observations, he will try a curve of next-highest degree, say a hyperbola or a parabola. The point is that, other things being equal, the simpler curve has the higher probability of being right. A truly astonishing number of basic laws are expressed by low-order equations. Nature's preference for extrema (maxima and minima) is another familiar example of simplicity because in both cases they are the values when the function's derivative equals zero.

This raises some of the most perplexing questions in the philosophy of science. How can this particular kind of simplicity—the kind that adds to the probability that a law or theory is true—be defined? If it can be defined, can it be measured? Scientists tend to scorn both questions. They make intuitive judgments of simplicity without worrying about exactly what it is. Yet it is conceivable that someday a way to measure simplicity may have great practical value. Consider two theories that explain all known facts about fundamental particles. They are equal in their power to predict new observations, although the predictions are different. Both theories cannot be true. Both may be false. Each demands a different test and each test costs $1 million. If simplicity enters into the probability of a theory's being right, there is an obvious advantage in being able to measure simplicity so that the simplest theory can be tested first.

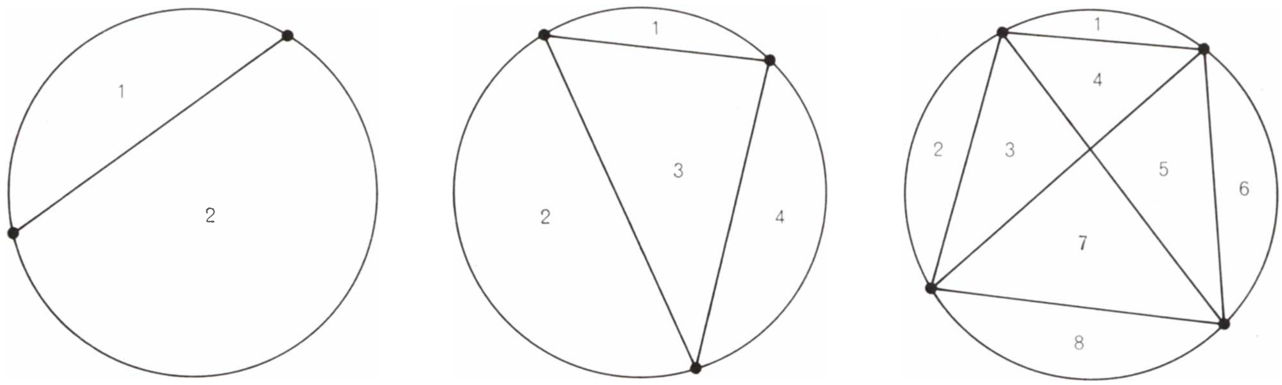
No one today knows how to measure this kind of simplicity or even how to define it. *Something* in the situation must be minimized, but what? It is no good to count the terms in a theory's mathematical formulation, because the number depends on the notation. The same formula may have 10 terms in one notation and three in another. Einstein's famous $E = mc^2$ looks simple only because each term is a shorthand symbol for concepts that can be written with formulas involving other concepts. This happens also in pure mathematics. The only way to express pi with integers is as the limit of an infinite series, but by writing π the entire series is squeezed into one symbol.

Minimizing the powers of terms also is misleading. For one thing, a linear equation such as $x = 2y$ graphs as a straight line only when the coordinates are Cartesian. With polar coordinates it graphs as a spiral. For another thing, minimizing powers is no help when equations are not polynomials.

Even in comparing the simplest geometric figures the notion of simplicity is annoyingly vague. In one of Johnny Hart's *B.C.* comic strips a caveman invents a square wagon wheel. Because it has too many corners and therefore too many bumps, he goes back to his drawing board and invents a "simpler" wheel in the shape of a triangle. Corners and bumps have been minimized, but the inventor is still further from the simplest wheel, the circle, which has no corners. Or should the circle be called the most



A "simple" geometrical theorem

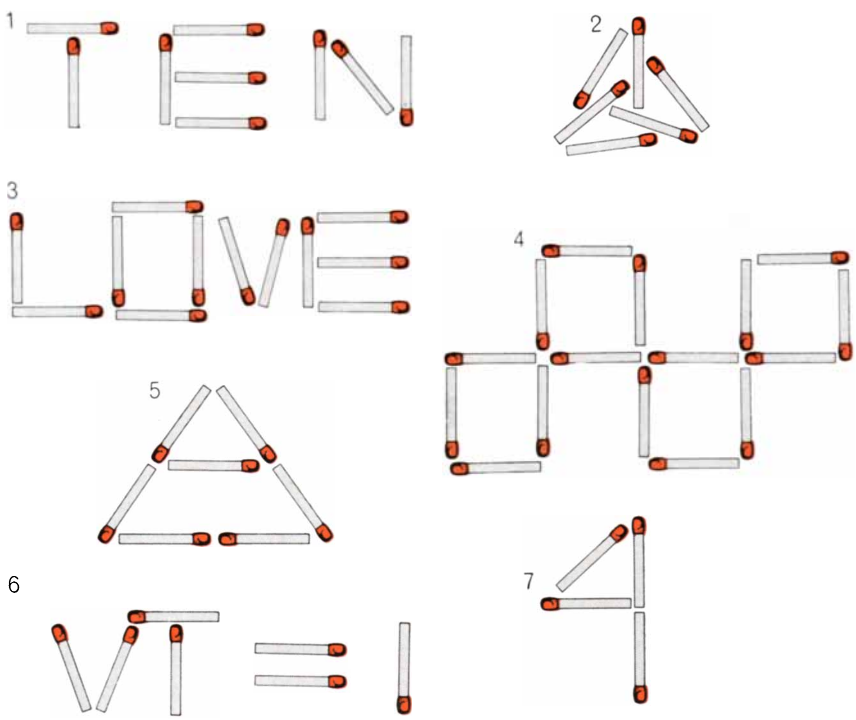


A combinatorial problem

complicated wheel because it is a "polygon" with an infinity of corners? The truth is that an equilateral triangle is simpler than a square in that it has fewer sides and corners. On the other hand, the square is simpler in that the formula for its area as a function of its side has fewer terms.

One of the most tempting of many proposed ways to measure the simplicity of a hypothesis stated in words is to count its number of primitive concepts. This, alas, is another blind alley. One can artificially reduce concepts by combining them. The philosopher Nelson Goodman brings this out clearly in his famous "grue" paradox about which dozens of technical papers have recently been written. Consider a simple law: All emeralds are green. We now introduce the concept "grue." It is the property of being green if observed, say, before January 1, 1970, and being blue if observed thereafter. We state a second law: All emeralds are grue.

Both laws have the same number of concepts. Both have the same "empirical content" (they explain all observations). Both have equal predictive power. A single instance of a wrong color, when an emerald is examined at any future time, can falsify either hypothesis. Everyone prefers the first law because "green" is simpler than "grue"—it does not demand new theories to explain the sudden change of color of emeralds on January 1, 1970. Although Goodman has done more work than anyone on this narrow aspect of simplicity, he is still far from final results, to say nothing of the more difficult problem of measuring the overall simplicity of a law or theory. The concept of simplicity in science is enormously complex! It may turn out that there is no single measure of simplicity but many different kinds, all of which enter into the final evaluation of a law or theory.



Answers to last month's match puzzles

Surprisingly, even in pure mathematics similar difficulties arise. Mathematicians usually search for theorems in a manner not much different from the way physicists search for laws. They make empirical tests. In pencil doodling with convex quadrilaterals—a way of experimenting with physical models—a geometer may find that when he draws squares outwardly on a quadrilateral's sides and joins the centers of opposite squares, the two lines are equal and intersect at 90 degrees [*see illustration on preceding page*]. He tries it with quadrilaterals of different shapes, always getting the same results. Now he sniffs a theorem. Like a physicist, he picks the simplest hypothesis. He does not, for example, test first the theorem that the two lines have a ratio of one to 1.00007 and intersect with angles of 89 and 91 degrees, even though this theorem may equally well fit his crude measurements. He tests first the simpler guess that the lines are always perpendicular and equal. His "test," unlike the physicist's, is a search for a deductive proof that will establish the hypothesis with certainty. (A proof and generalization of this theorem was given by Paul J. Kelly in *Mathematics Magazine*, January, 1966, pages 35–37. A different proof, using symmetry operations, is in *Geometric Transformations*, by I. M. Yaglom, Random House, 1962, pages 95–96.)

Combinatorial theory is rich in similar instances where the simplest guess is usually the best bet. As in the physical world, however, there are surprises. Consider the following problem presented by Leo Moser. Two or more spots are placed anywhere on a circle's circumference. Every pair is joined by a straight line. Given n spots, what is the maximum number of regions into which the circle can be divided? The top illustration on the opposite page gives the answers for two, three and four spots. The reader is asked to search for the answers for five and six spots and, if possible, find the general formula before it is disclosed next month.

David Silverman's puzzle of last month is answered by observing that any player who wins the game of Connecto obviously must have two matches forming the letter L in the boundary of his region. The second player can prevent the first player from winning, on a board of any size, simply by playing to prevent his opponent from forming an L. If the first player plays the vertical bar of a possible L, the second player plays the horizontal bar. If the first player forms the horizontal part of a possible L, the second player forms the vertical part. This guarantees at least a draw for the second player. Answers to the seven match puzzles are given in the bottom illustration on the opposite page.

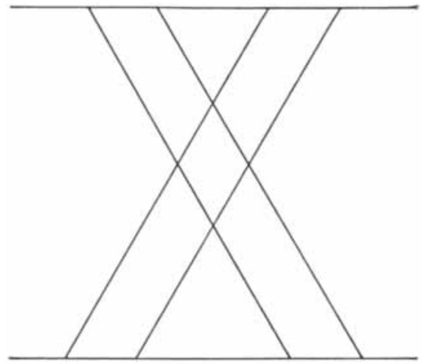
An ambiguity in the use of "simplicity" was involved in the first of the 10 "quickies" in April's collection of short problems. I asked for the "simplest" way to time the boiling of an egg for 15 minutes with a seven-minute hourglass and an 11-minute one. Had "simplest" been defined as the procedure taking the least amount of time, the answer in May would have been best. As scores of readers have pointed out (too many for me to list their names), "simplest" can also mean a procedure requiring the fewest turns of the two timers. The answer given required four turns and 15 minutes, but there is a 22-minute procedure requiring only three turns. Start both timers simultaneously. When the seven-minute hourglass stops, put the egg into the boiling water. (The first seven minutes can be used for bringing the water to a boil.) Reverse the 11-minute timer when its sand has stopped running. When it stops again, 15 minutes have elapsed.

In answering the problem of the rotating table I said I knew of no way a full proof of the theorem could be "compressed to a reasonable length." Again, so many readers sent proofs that I cannot list their names. Many professional mathematicians sent elegantly worded proofs of just a few lines, but the price of such compression was the use of technical symbols that few readers would have understood. Others sent less technical proofs that required several typewritten pages. All the proofs were essentially the same. The following, from physicist George Rybicki of the Smithsonian Institution, is typical of many that are good compromises between the too technical short proof and the too lengthy nontechnical one.

The theorem to be proved is that if any even number of people take seats at random around a circular table bearing place cards with their names, it is always possible to rotate the table until at least two people are opposite their cards. Assume the contrary. Let n be the even number of persons, and let their names be replaced by the integers 0 to $n - 1$ "in such a way that the place cards are numbered in sequence around the table. If a delegate d originally sits down to a place card p, then the table must be rotated r steps before he is correctly seated, where $r = p - d$, unless this is negative, in which case $r = p - d + n$. The collection of values of d (and of p) for all delegates is clearly the integers 0 to $n - 1$, each taken once, but so also is the collection of values of r, or else two delegates would be correctly seated at the same time. Summing the above equations, one for each delegate, gives $S - S + nk$, where k is an integer and $S = n(n - 1)/2$, the sum of the integers from 0 to $n - 1$. It follows that $n = 2k + 1$, an odd number." This contradicts the original assumption.

"I actually solved this problem some years ago," Rybicki writes, "for a different but completely equivalent problem, a generalization of the nonattacking 'eight queens' problem for a cylindrical chessboard where diagonal attack is restricted to diagonals slanting in one direction only. I proved that this was insoluble for any board of even order. The above is the translation of my proof into the language of the table problem. Incidentally, the proof is somewhat easier if one is allowed to use congruences modulo n."

Donald E. Knuth also called attention to the equivalence of the round table and the queens problem, and cited an early solution by George Polya. Several readers pointed out that when the number of persons is odd, a simple arrangement that prevents more than one person from being seated correctly, regardless of how the table is rotated, is to seat them counterclockwise to their place-card order.



New solution to an April problem

Many readers informed me of an error in the working of quickie No. 4: "Find a number base other than 2 or 10 in which 121 is a perfect square." I should have said "3 or 10" since there is no 2 in the binary system. A quick proof that 121 is a square in any system with a base greater than 2 is to observe that 11 times 11, in any system, has a product (in the same system) of 121. Craige Schensted showed that, with suitable definitions of "perfect square," 121 is a square even in systems based on negative numbers, fractions, irrational numbers and complex numbers. "Although the bases may not be exhausted, I am and I assume you are, so I will stop here," he concluded.

John Appel and Daniel Rosenblum were the first to tell me that quickie No. 7 was a version of a joke attributed to Abraham Lincoln. He once asked a man who had been arguing that slavery was not slavery but a form of protection how many legs a dog would have if you called its tail a leg. The answer, said Lincoln, is four because calling a tail a leg does not make it a leg.

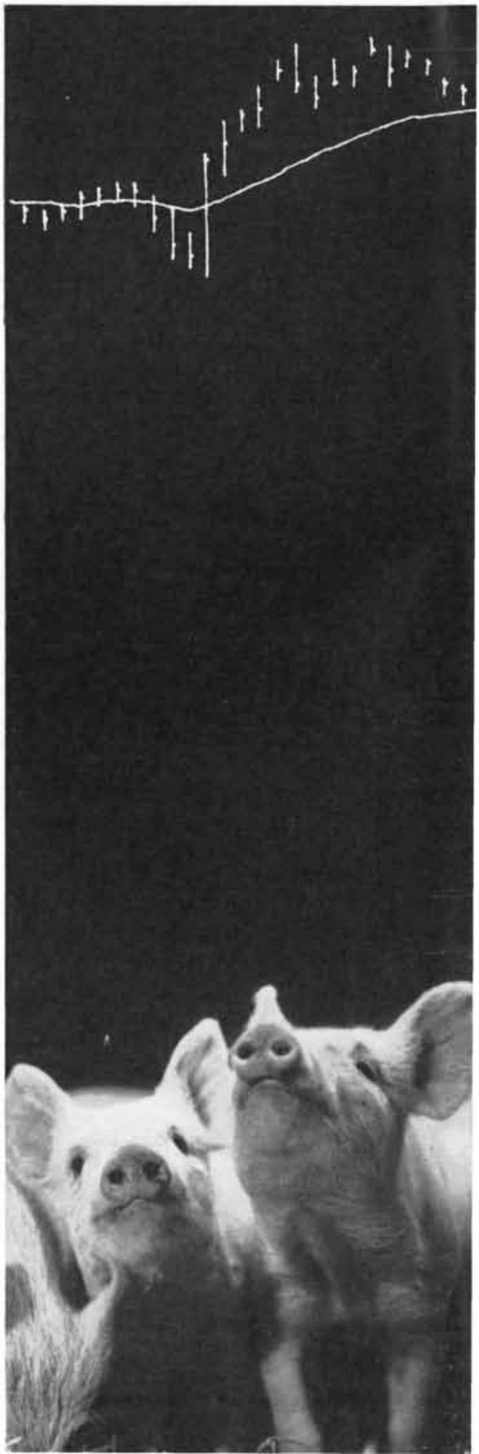
My translation of quickie No. 8 ("He spoke from 2222222222222 people") prompted numerous other interpretations. David B. Eisendrath, Jr., wrote that if the speaker had been called a colonel, it would have implied military time and the translation "He spoke from 22 to 22 to 22:22 to 22 people." Gary Rieveschl and Harry Kemmerer each sent an alternate solution to quickie No. 5 [*see illustration above*]—one that would have been ruled out if I had added "No loose ends." Many readers wrote to say that quickie No. 10 is solved by asking twice any question the answer to which is known with certainty. Others, also too numerous to mention, sent different proofs for the problem about the three beer rings.

Measurement display spots trends in the commodity market. Stops bullets from ruining the 6 o'clock news. Takes the pain out of learning impedance.

Charting futures in pork.
Consider the problems of a broker in futures – precious metals, wheat or pork bellies. Imagine the seasonal shifts, or changing daily demands, which could radically alter his view on cocoa in the morning.

A large securities house may have the time, staff and possibly a computer to follow the action of slower trading. The commodities man needs an answer now. Fortunately, he can produce his own charts overnight with the new Hewlett-Packard 9100A Computing Calculator and its 9125A X-Y Recorder.

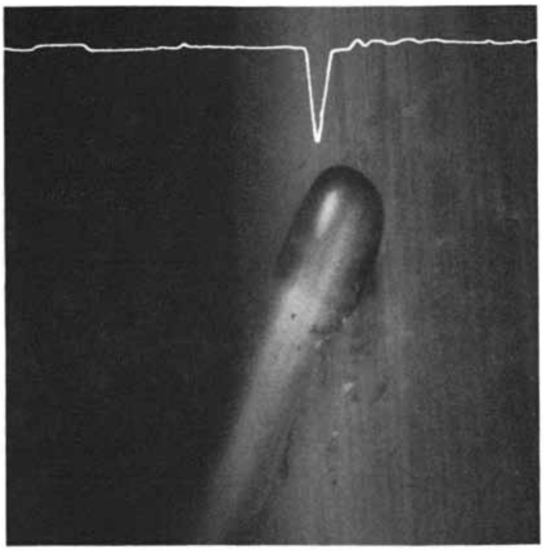
He inserts a wallet-sized magnetic card already programmed for high,



low, close. Percentage loss or gain. Extended logarithmatic average. Possibly a "Nervousness Indicator" developed out of his own experience.

He punches in today's reports. In seconds the plotter starts drawing curves. He can even continue a chart started yesterday. And draw his own conclusions. Privately. Smoothly.

The 9100A Calculator can also handle such problems as coordinate geometry, network analysis, statistical distribution, strain loadings or gear design; the 9125 X-Y Recorder plots the results. Cost: $4400 for the calculator, $2475 for the recorder. Write today for our illustrated brochure on the 9100A Computing Calculator and its accessories.



Bouncing echoes off bullet holes.

The nation's highest TV transmitter, 1500-feet tall, was set up to improve TV pictures over a vast area of Texas. It had scarcely been put into use when there were complaints about more bad pictures. When engineers inspected the tower, they found that marksmen were using it for target-practice. The only answer was to find each hole or dent in the cable.

They could have inspected the cable inch by inch. Instead they used HP's new time domain reflectometer that plugs into our 180A Portable Oscilloscope. This device acts much like a miniature radar to pinpoint breaks in transmission lines within a quarter inch. The reflectometer marks the transmission of a pulse with a step in the sweep on an oscilloscope. Any disturbance in the path reflects some energy which is sampled by the reflectometer and easily recognized on the scope. The distance can be read off in nanoseconds or feet of travel. The first hole is a big one, causing a major reflection 35 feet $3\frac{3}{4}$ inches up, the second a dent in the outer conductor, causing a capacitive change 70 feet $4\frac{1}{2}$ inches up, and so on.

The HP 1815A Time Domain Reflectometer obviously has other applications than on the Texas range. It can locate faults in the maze of cables on aircraft, in buried cables, in cables on production lines, or in cable connectors. It operates with frequencies up to 12.4 gigahertz (12.4×10^9 Hz). Reflectometer, sampler and step generator cost $3150. If you need only inches of resolution rather than millimeters, a less expensive system sells for $2125. Of course, you'll need the HP 180A Portable Oscilloscope, at $825.

If you have an interest in fault finding—TV or otherwise—simply ask for Data Sheets 1815A and 180A.

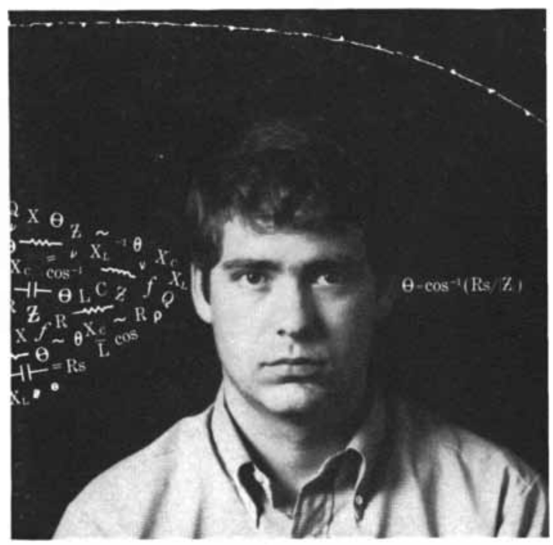
Demonstrating the dynamics of Z.

To undergraduate students of physics, some theories are absolutely Greek. For example, take the behavior of a simple electrical circuit faced with alternating currents of different frequencies. No longer is resistance simply a matter of E=IR; now the student is involved with the Zeta and Theta of impedance, inductance and capacitance.

With all the calculations required, demonstrating the effect of frequency in the laboratory is next to impossible. Spot measurements with bridge, ac generator, ac voltmeter or oscilloscope can hide rapid changes of impedance with changes in frequency. They can also obscure the construction errors of the student, even if all his calculations are correct.

A happy way out of this hang-up is the use of a Hewlett-Packard Model 4800A Vector Impedance Meter. Two meters on the face indicate directly the magnitude and phase angle of complex impedance (Zeta and Theta), regardless of how often the frequency is changed. The meter is just as responsive when the frequency is electronically swept over a range from 5 hertz to 500 kilohertz. And the student can readily plot curves which demonstrate that inductive reactance and capacitive reactance are, as a matter of fact, directly and inversely proportional to frequency. (The HP 4815A Vector Impedance Meter will do the same job for frequencies between 500 kilohertz and 108 megahertz.)

The cost of the meters may seem expensive for the undergraduate institution ($1650 for the HP 4800A and $2650 for the HP 4815A), but they compare favorably with the significant advantage of greater student confidence in the utility of basic concepts.



A complete discussion of the teaching technique is found in AMERICAN JOURNAL OF PHYSICS, Volume 37, Number 1, January 1969. Detailed descriptions of the instruments are available from Hewlett-Packard for the asking.

If there's something we can do to help you solve a problem with a new approach to measurement display, give us a call. Hewlett-Packard, 1506 Page Mill Road, Palo Alto, California 94304; Europe: 1217 Meyrin-Geneva, Switzerland.

HEWLETT PACKARD

Measurement, Analysis and Computation

00911

The Milky Way as photographed with J. R. Bruman's apparatus

The Crab Nebula

The Great Nebula in Andromeda

The Ring Nebula

The Dumbbell Nebula

THE AMATEUR SCIENTIST

Color photographs of the night sky are made by refrigerating the film

Conducted by C. L. Stong

It is unlikely that many observers would regard the night sky as a natural spectacle in color. On clear moonless nights two colors predominate: the purplish black of the background sky and the silver of the stars. The patient observer can spot a few reddish stars such as Antares, and a number of bluish ones such as Rigel, but most laymen agree that the celestial vault owes its splendor more to the profusion of objects than to its colors. Yet the night sky is actually filled with great splashes of unseen color bathing the earth from every direction in space: deep reds from clouds of interstellar gas, intense blues from the center of exploding stars and greenish hues from the arms of our galaxy.

These colors first came to public attention 10 years ago when astronomers at the Mount Wilson and Palomar Observatories made a series of photographs with color film. Amateur astronomers enthusiastically responded to the pictures and promptly tried to duplicate them with homemade equipment. The results were disappointing: the film came out either blank or foggy. The Mount Wilson and Palomar pictures had been made by exposing high-speed film for periods up to four hours in the 200-inch telescope at a focal length of $f/3.3$ and in the 48-inch Schmidt telescope at $f/2.5$. The combination of large aperture and long focal length is of course beyond the reach of amateurs.

Not all amateurs were discouraged. One who kept trying is J. R. Bruman of 3527 Cody Road in Sherman Oaks, Calif. Last year Bruman, adapting a technique devised by Arthur A. Hoag, an astronomer at the Kitt Peak National Observatory, worked out a simple method of making good color photographs of the night sky with ordinary roll film in a camera equipped with an inexpensive refrigeration unit that he made at home. Several of his photographs appear on the opposite page. Bruman writes:

"If we received more light from space, the problem of photographing deep-sky objects in color would be simple. The light that enters the telescope ranges from a ten-thousandth to a millionth of the brightness of an average daylight scene. When we make photographs in daylight, we compensate for dim light by increasing the exposure, perhaps even switching from snapshots to time exposures. The reciprocal relation between available light and exposure time is known to everyone who owns a camera.

"When the illumination falls to the point where the required exposure exceeds a minute or so, strange things begin to happen in the film. The reciprocity rule breaks down, and eventually, as one attempts to make pictures with less and less light, a time comes when the film is either blank or seriously underexposed, no matter how long the shutter remains open. Deep-sky objects are in this range of illumination. They cannot be photographed with ordinary color film employed in the ordinary way. Other complications also arise. Even if one manages to obtain an image by extended exposure, the colors come out wrong, because the color sensitivity of each of the several layers of color emulsion in the film changes uniquely with the intensity of the illumination.

"What accounts for the failure of the reciprocity rule? Specialists explain that the latent image created when photons dissociate silver halide molecules in the film is not necessarily permanent. If the dissociation is weak as the result of a scant supply of photons, the ions tend to drift back together at about the same rate at which they are formed. Continued exposure at this level of illumination results in no net gain in ions.

"As one expert has explained it, forming a latent image on film is something like filling a leaky bucket with hot molasses. Pour in the thin molasses quickly and you can fill the bucket, but if you pour slowly, the thin fluid refuses to accumulate. If you switch to cold molasses, the thick fluid leaks at a much lower rate. You can fill the bucket with cold molasses by pouring at a much lower rate. The analogy applies to photographic emulsions. Some years ago it was discovered that if an emulsion is refrigerated during exposure, it will accumulate a latent image even under very dim light.

"At about the time the Mount Wilson and Palomar pictures appeared Hoag was experimenting with Ektachrome film in this way and getting excellent results. Hoag's refrigeration device consisted of a metal plate, just behind the film, that was cooled by various means, including dry ice. There was a difficulty, however, because fog and frost developed inside the camera as the moisture-laden air came in contact with the cold emulsion. To prevent this Hoag and others enclosed the film in a sealed compartment that had a glass window; they evacuated the compartment with a laboratory vacuum pump. The scheme worked but was awkward: the film had to be loaded one sheet at a time, and the need for a vacuum pump ruled out the technique for use in the field, at least for amateurs.

"A couple of years ago the notion occurred to me that if I could cool the film from the front by a flow of cold dry gas, I might not only overcome the problem of frosting without the use of a vacuum pump but also could substitute roll film, in a magazine, for the single sheets of cut film. To try the idea I made a well-insulated box for the film magazine and let the vapor from dry ice into the chamber. The resulting fog of carbon dioxide obscured the image. I added a fan to draw air across the film, but the film did not get cold enough. Therefore I substituted liquid nitrogen in the dry-ice container. The cold gas chilled the film enough, but the arrangement consumed nitrogen at a rate that required an inconveniently large reservoir of liquid nitrogen. Next I tried using the boil-off of liquid nitrogen from a Dewar vacuum flask that was connected to the film chamber by a long, flexible tube. I was unable to insulate the hose well enough to provide adequate cooling in the film chamber without consuming liquid nitrogen at an unreasonable rate.

"Thus I arrived at the present configuration: the container of liquid nitrogen is attached to the camera [*see top illustration on opposite page*]. A 15-watt heater immersed in the reservoir boils off liquid nitrogen at a rate of about 500 milliliters per hour. The gas flows from the reservoir through a short tube, swirls across the film and escapes through the telescope. I feared that the resulting currents and variations in the temperature of the gas would distort the image, but no trouble developed. The only difficulty arose when I tried to advance the film. At the low temperature it became brittle and snapped. Now I use two magazines. After an exposure I remove the cold magazine and quickly wrap it in a plastic bag for 15 minutes to prevent frost from forming. After the film warms up it can be handled normally. On the whole the technique appears to be adequate. I am getting about the same improvement in film performance that Hoag reported.

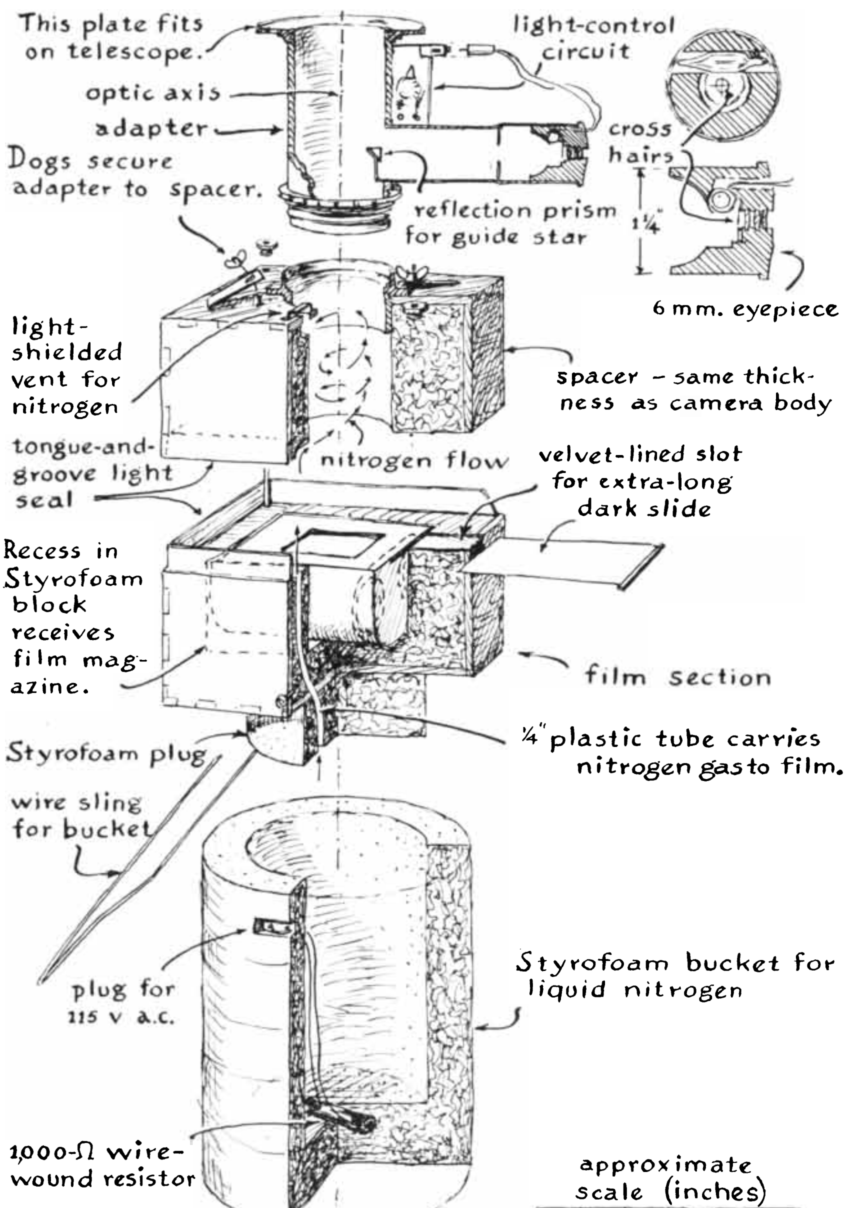
"Liquid nitrogen is not particularly hazardous, but several precautions should be observed. Never put the liquid, even briefly, in a tightly stoppered container. Avoid the use of containers that become brittle at cryogenic temperatures. Styrofoam (or its equivalent) is a satisfactory material for use at low temperature. Make sure that the liberated gas does not displace the oxygen you require for breathing. This possibility is remote in the case of telescopes operated outside the observatory. All cryogenic liquids can cause serious injury if they are splashed in the eyes or on the skin. Do not handle them carelessly.

"Liquid nitrogen costs only a few cents per liter when it is bought in large quantities. Generally it is more widely available than dry ice. In most communities with a population of more than 10,000 one can find suppliers who stock liquid nitrogen to meet the routine demand of machine shops, numerous small industries and laboratories of all kinds. The technique of making color photographs of the sky with cooled emulsion can be undertaken by any resourceful person, certainly by amateur astronomers who have done guided celestial photography with black-and-white film or plates.

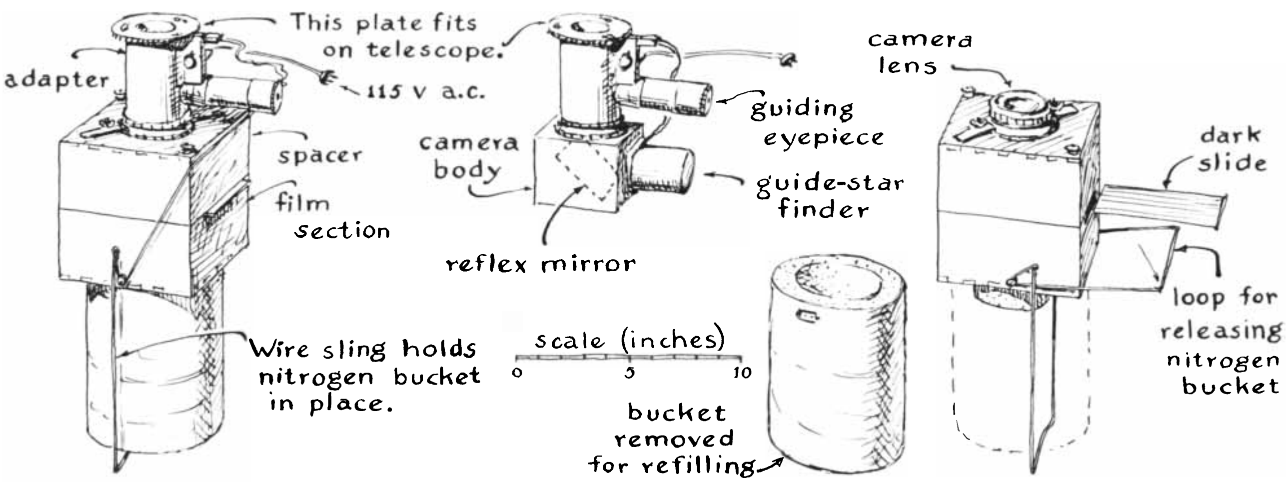
"At about the time I became interested in color photography I was a guest at the Table Mountain Observatory of the Jet Propulsion Laboratory near Los Angeles. The observatory is devoted to planetary studies in connection with the program of the National Aeronautics and Space Administration, but the main telescope is a general-purpose instrument of the Cassegrain type with a 24-inch aperture and a focal ratio of $f/16$. The color photographs that accompany this discussion [*page 124*] were made either with this instrument, with a war-surplus aerial camera lens of 36-inch focal length or with the 80-millimeter lens of my reflex camera.

"Specialists will have no difficulty pointing out technical flaws, but the pictures demonstrate what can be accomplished with film bought in a retail camera shop. For example, the photograph of the Crab Nebula was made on Ektachrome EH, which has an ASA rating of 160. The exposure was for one hour at a film temperature of −60 degrees Celsius. Development, a few hours later, was normal except for an increase of first developer time from seven minutes to 15 minutes. Forcing development in this way raises the effective ASA rating of Ektachrome EH to about 600, as I determined by making exposures of Jupiter at .5 second at $f/112$.

"It is also possible to enhance the results slightly by baking the film before use. I baked the Ektachrome used for the Crab Nebula for 24 hours at 50 degrees C. immediately before making the exposure. The instantaneous speed of the film appears to be unaffected, but its response to long exposure at low light intensity is somewhat improved, extending the range of subjects that can be



Exploded view of Bruman's refrigerated camera



Alternative configurations of the assembled camera

tackled with forced development alone. Baking film beforehand is much less effective than cooling during the exposure, so that I have given it up.

"Incidentally, amateurs who have never made photographs of the sky in black and white are urged to acquire a bit of experience before attempting to work with color, so that they will have at least a casual acquaintance with the location of various celestial objects, the principles of an equatorial clock drive and so on. To make successful photographs the telescope must automatically track the field where the desired object appears; in addition a selected star within the field must be kept centered in a reticle manually by means of auxiliary drives. Amateurs usually use a small separate telescope for observing the selected guide star, but professionals prefer to guide on part of the image that is being photographed, a method that eliminates possible error arising from mechanical flexure between the principal objective and the small guiding telescope. A prism at the edge of the field diverts a small part of the image into the eyepiece used for guiding.

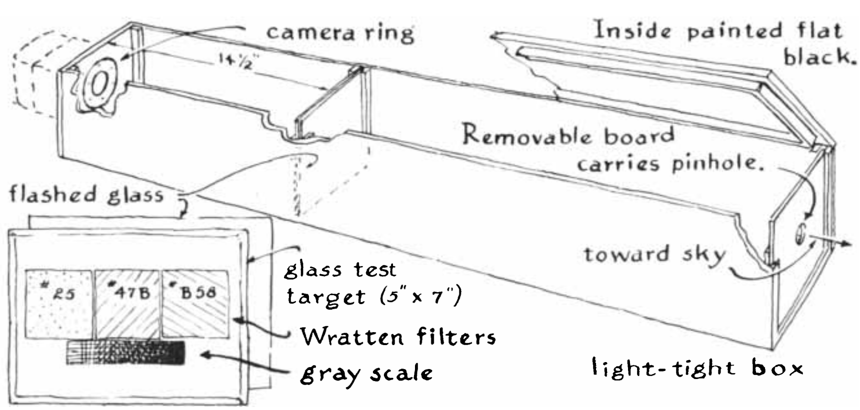
"For this purpose I use a cheap six-millimeter orthoscopic eyepiece obtained from the Edmund Scientific Co., Barrington, N.J. 08007, to which I added a cross hair that is illuminated by a miniature neon lamp. The cross hair consists of two strands from a piece of glass cloth; they appear to be finer than the strands of a spider web and are much stronger. The six-millimeter eyepiece magnifies the image of the guiding star and its motion 40 diameters, which is about the optimum. The brightness of the neon lamp can be adjusted by a one-megohm radio volume-control unit that acts as a voltage divider. Commercial eyepieces of this type employ midget incandescent lamps, which are easier to control but require special power supplies. The light from lamps of either type must be kept dim so that it will not mask the image of the guide star.

"The controls that regulate the movement of the telescope may be either electrical or mechanical, but they must be designed for operation without removing the eye from the eyepiece. It is not easy to find and center a guide star in the small field of view. I first locate the guide star by examining on the ground glass the entire field to be photographed. This is done with the aid of a large magnifier. A dim light source, consisting of a piece of 1/16-inch brass tubing illuminated at one end by a second neon lamp, indicates on the ground glass the position of the small prism that diverts part of the image into the guiding eyepiece. It is easy to shift the selected guide star to this illuminated area, where it vanishes and immediately appears in the guiding eyepiece. Next I detach the reflex camera and finder, replacing it with the cooled-emulsion assembly.

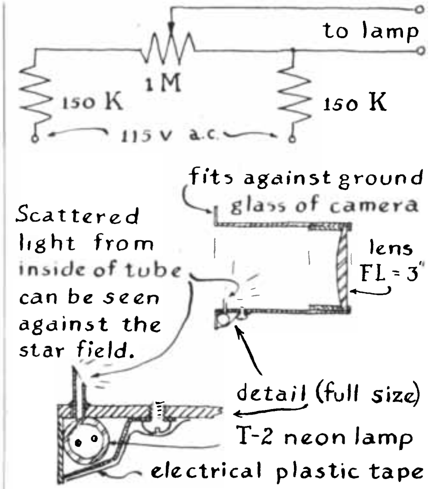
"To avoid sky fog I limit exposures (in minutes) to approximately the square of the *f* ratio, for example 10 minutes at *f*/2.8, 60 minutes at *f*/8 and so on. The published magnitudes of deep-sky objects indicate the total amount of light, not the brightness, so that the appropriate time of exposure must be determined by trial and error. Generally I start by making the longest possible exposure, as determined by the *f*-ratio rule. Then I work downward and use forced development.

"The structural details of the apparatus were designed to be compatible with my 2¼-by-2¼-inch Hasselblad reflex camera, but the dimensions can be changed easily for cameras of other types. Commercially available parts, such as the closeup extension tube that joins the spacer and the camera to the telescope (and also functions as a guiding microscope), were modified with hand tools for accepting the neon lamps and the small prism.

"Having made an exposure with color film at low temperature and developed the image, how certain can one be that



Details of the sensitometer



Lighting the finder hood

the resulting colors are truly representative of the color emitted by the object? At the light levels involved in deep-sky photography one cannot see much color. Moreover, the sensation of color is subjective, so who is to say what an invisible color should be if it is intensified? The color values of astronomical subjects are not well known, and some specialists are dubious about the color rendition of the cooled-emulsion technique. Little quantitative information is available because exposure times far exceed those of ordinary photography and each film has its own distinct properties.

"To check the verity of the resulting colors I constructed a simple three-color sensitometer that enables me to make a series of photographs of a colored target under illumination that can be varied in intensity without change in color. From the resulting photographs the reciprocity factor of the film can be judged in order to determine its exposure requirements. The photographs also provide a basis for the selection of corrective filters. The illumination of common deep-sky objects about a millionth of ordinary daylight, so that the sensitometer was designed for this broad range.

"Essentially the sensitometer consists of a light-tight box, which supports a test pattern of three color filters in front of a camera lens, and a variable aperture through which the test pattern is illuminated. The camera is at one end of the box, the test target is in the middle and the variable aperture is at the other end [*see bottom illustration on preceding page*]. The largest aperture is simply the open end of the box, the smallest a hole .01 inch in diameter made in a piece of aluminum foil. The foil can be supported by a board that slides into a set of grooves in the box. The ratio of the area of the two openings is 360,000 to one, hence an exposure of .01 second through the large opening corresponds to an exposure of one hour through the small hole. An exposure of .01 second is well within the reciprocal range of commercial films, and an exposure of one hour is typical of time exposures made through the telescope.

"The back of the sensitometer is adapted to the front ring of the Hasselblad lens, which in turn fits either the cooled camera or the standard camera, so that long exposures can be made with refrigerated film or short ones with film at room temperature. Fractional-second exposures are made with the existing camera shutter, since they do not need to be cooled. The test pattern consists of three gelatin filters and a gray scale sandwiched between sheets of glass. To ensure uniform, diffused illumination across the pattern I use white paper over the variable aperture and opal glass as the front layer of the sandwich. The filters consist of a Wratten No. 47, blue; 58, green, and 25, red. I recommend this combination for making tests with Ektachrome because each filter transmits one of the three color ranges of the film without overlapping the other two. Therefore a test exposure made with the sensitometer will identify any layer of the film that may respond abnormally.

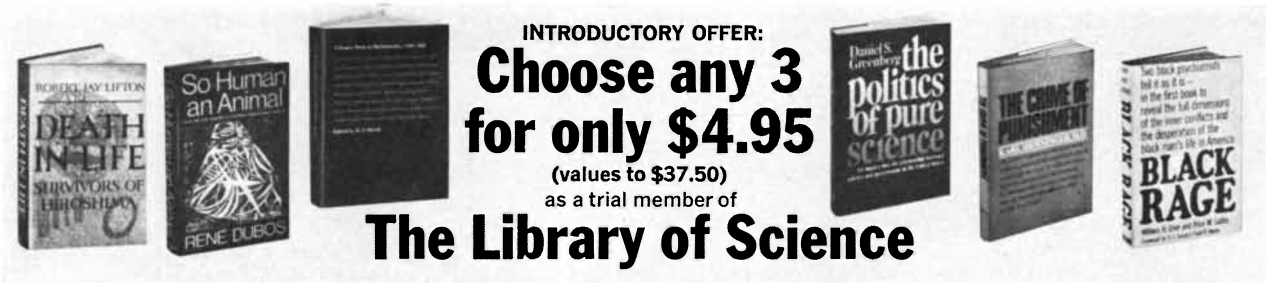
"Exposures have been made with the sensitometer using both sunlight and an incandescent lamp as the source. Both yield comparable results. One-hour cooled exposures made with the small aperture correspond closely, in terms of the results, to those made with the large aperture at .01 second. In every test using cooling the best match resulted when the illumination was exactly reciprocal to the exposure time; for the short exposure it was 360,000 times as intense as for the long exposure. In contrast, exposures made through the small aperture for one hour with uncooled film resulted in severe underexposure, equivalent to at least two stops, and the color balance could not be properly evaluated.

"The use of liquid-nitrogen cooling should open many new experimental opportunities to amateurs because the technique is both simple and inexpensive. The exposed films can be developed by local processing laboratories if they are equipped to do forced development on a custom basis. Although further experimentation is needed to determine quantitatively the response of films other than Ektachrome, I feel that the principle of cooling films on the emulsion side by nitrogen is now established."

Why do so many books chosen for you by The Library of Science turn out to be prize winners?

Because our editors spare no effort to search out the very best new books in all scientific disciplines and the interesting by-paths to which they lead. Generally, these are submitted to us in manuscript form; out of a hundred books submitted, we may choose only one. If a book later wins a prize, or some special citation, it is merely confirmation of good judgment and discerning choice.

A surprisingly large number of selections have been so honored. It's all very flattering, but we feel that *you* are the real prize winner. You get the best of the new books, on publication, at savings up to 30% (and not infrequently, more). And for every four books you buy, you earn a free bonus book (which might be a prize winner, too). Add this money-saving offer—and can you think of a good reason for not joining now?



1969 National Book Award for Science

41890. DEATH IN LIFE. *Robert Jay Lifton.* First study in depth of the "psychohistorical" effects of the Hiroshima holocaust. Based on 75 interviews with survivors. Vivid, moving and meaningful. Also cited by the American Library Association as one of its "44 Notable Books" of the past year. $10.00

1969 Pulitzer Prize

78730. SO HUMAN AN ANIMAL. *René Dubos.* The famous biologist warns that a runaway technology is destroying the environment man needs to survive as a human being. $6.95

American Institute of Graphic Arts "Fifty Books of the Year"

79470. A SOURCE BOOK IN MATHEMATICS. *D. J. Struik.* Big, 472-page volume of original selections from the mathematical thought and writings of authors from the 13th to the 19th centuries. A rich history of mathematical thought. $11.95

American Library Association—"44 Notable Books"

69390. THE POLITICS OF PURE SCIENCE. *Daniel S. Greenberg.* An important and valuable contribution to the growing body of literature on the political aspects of science. Includes detailed accounts of such major political involvements as Project Mohole and the MURA accelerator. $7.95

40840. THE CRIME OF PUNISHMENT. *Karl Menninger, M.D.* A startling book which shows why our penal system aggravates crime, and how we can avoid criminality. The author, long recognized as one of our most gifted psychiatrists sheds a "refreshing aura of excitement and hope."—*The New York Times Book Review.* $8.95

36440. BLACK RAGE. *William H. Grier, M.D., and Price M. Cobbs, M.D.* The first book to reveal the full dimension of the inner conflicts and the desperation of the black man's life in America. Written by two black psychiatrists, this eloquent book depicts the depth of the fury in America's black world. $5.95

...and some additional notable books to choose from:

(retail price shown)

66060. OPTICAL PHYSICS. *Stephen G. Lipson and Henry Lipson.* Principles of optics, and their influence in other branches of physical theory. A brilliant new analysis, written with clarity, insight and wit. $12.50

32530. ALGEBRA. *Saunders MacLane and Garrett Birkhoff.* A completely new standard work by the eminent authors who revolutionized the study of algebra. Here are the new approaches, the new vocabulary indispensable for today's students. $11.95

56330. INTRODUCTION TO SOLID STATE PHYSICS. *Charles Kittel.* Complétely revised and rewritten 3rd Edition of the authoritative work on the subject. $12.50

70710. PROGRAMMING: An Introduction to Computer Languages and Techniques. *Ward Douglas Maurer.* A remarkably lucid presentation of the basic techniques of programming in an algebraic language and in machine language. Concepts are defined in terms of their application to specific computers in current use. $11.50

44190. ELEMENTARY PARTICLES AND THEIR CURRENTS. *Jeremy Bernstein.* A highly acclaimed new presentation which concentrates on the interaction of currents. The author is Professor of Physics at Stevens Institute of Technology. $12.00

74170. THE RIDDLE OF GRAVITATION. *Peter G. Bergmann.* Masterful exposition of the general theory of relativity by a physicist who worked closely with Einstein. $7.95

70350. PROBABILITY THEORY AND ITS APPLICATIONS. Vol. 1, Third Edition. *William Feller.* Illustrates a broad variety of practical problems and the modern techniques for solving them. $10.95

81870. SYSTEMS ANALYSIS: A Diagnostic Approach. *Van Court Hare, Jr.* A major work on a new discipline. Contains an abundance of graded problems and exercises. $12.50

53030. HISTORICAL ARCHAEOLOGY. *Ivor Nöel Hume.* The first handbook that tells how to search for Colonial and post-Colonial relics that may lie under a few inches of earth just outside your door. $10.00

73900. RESEARCH AND DEVELOPMENT MANAGEMENT: The Economics and Adminstration of Technology. *Daniel D. Roman.* The first book to cover comprehensively the logistics of managing research and development. $10.00

TRIAL MEMBERSHIP APPLICATION

THE LIBRARY OF SCIENCE Front and Brown Streets, Riverside, N. J. 08075 2-958

Please enroll me as a trial member, and send the three books whose numbers I've indicated on this coupon. Bill me only $4.95 for all three books, plus a small shipping charge. If I am not delighted, I will return them within 10 days, and my membership will be cancelled.
As a trial member, I need accept as few as 3 more selections during the next 12 months, *always* at reduced member's prices, plus shipping. I understand that savings range up to 30% and occasionally even more. Each month I will receive advance reviews describing forthcoming selections, along with convenient form for requesting alternate selections or no book at all. For every 4 selections purchased, I may choose a *free* bonus book. (This introductory offer counts as the first selection).

3 books for only $4.95

_______ _______ _______
(write in numbers)

NAME..

ADDRESS..

CITY.................. STATE.................. ZIP CODE..........

Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.

BOOKS

The modern migration of European intellectuals, among other matters

by Philip Morrison

THE INTELLECTUAL MIGRATION: EUROPE AND AMERICA, 1930–1960, edited by Donald Fleming and Bernard Bailyn. Harvard University Press ($12.95). The gifted amateur writing history out of direct experience is one of the intellectual successes of our times. This is a many-author work that embodies not only that fine tradition but also the approach of the professional historian: comparing sources and presenting the document together with its gloss. The result is magnificent; it is unlikely that a better volume on this important topic will soon arrive.

Here is a passage from the taped reminiscences of that most original and consequential of emigrants to America, Leo Szilard. In 1932 Szilard read H. G. Wells's *The World Set Free*. It made a great impression on him as a fictional account of atomic war leading to the "shall I say utopian" establishment of world government. Szilard left Germany (he had foreseen the catastrophe and was well prepared) a few days after the Reichstag fire. He made his way to London, where with his help a committee had been formed to assist the academic exiles, so that "practically everybody... had a position, except me." After the 1933 meeting of the British Association for the Advancement of Science, at which Lord Rutherford had remarked that the expectation of nuclear power was "the merest moonshine," Szilard began to ponder. "As the light changed to green and I crossed the street, it suddenly occurred to me that if we could find an element which is split by neutrons and which would emit *two* neutrons when it absorbed *one* neutron, such an element, if assembled in sufficiently large mass, could sustain a nuclear chain reaction.... The idea never left me."

He applied for a British patent on this notion beginning in 1934. The patent was issued, but it was sealed secret in 1936. In late 1938, from America, Szilard sought to withdraw it because "I couldn't make the process work." A month later he heard from Eugene Wigner about fission; "all the things which H. G. Wells had predicted appeared suddenly real to me." He canceled his request for withdrawal of the patent. The collected writings of Szilard are promised soon, to fill out this marvelous and romantic yet real tale, in which all of us are minor characters. The 50 pages here are hard to excel; surely they will be translated into theater.

So much represents the historian-as-documenter. The penetrating history of ideas is also a component of this volume. Peter Gay, Berlin-born Columbia historian, presents an essay, as luminous and as foreboding as a will-o'-the-wisp, on the Weimar background of the emigration, under the subtitle "The Outsider as Insider." He presents the period as a time when—with the roots of the past cut off by defeat, political murder and insane inflation—the bearers of the older culture stood empty-handed. There came a raging hunger for a lost wholeness that led in the end to the rejection of reason and discipline by the young, impatient with the inadequacy of the state structure. The "spirit of Weimar" was only a sojourner in Germany; it found its true home in exile. Such is Gay's bitter judgment.

Donald Fleming writes a tour de force, a history of the origins of the double helix. Today's molecular biology, which grew out of the bacteriophage work of the physicist Max Delbrück and the microbiologist Salvador Luria, both emigrants to America, carries a load of philosophical irony. The profound success that, loyal to the spirit of Niels Bohr's Copenhagen, the brilliant and dedicated Delbrück had seen as the eventual outcome of seeking a quantum-theoretical structure in biology was in fact achieved. It was hardly, however, a new physics; it held no trace of the deep complementarity half-prophesied by Bohr. When it came, it was no deeper in paradox than the mundane relation of lock to key. Comment on Fleming's strongly argued case by the combative and articulate principals is worth watching for.

The volume holds much more—too much to sketch. Now it is the generous account of a learned and philosophical sociologist on his first encounter with the American tradition of questionnaire and interview and all-embracing category. Again it is the story of the psychologists and the psychoanalysts, from Egon Brunswik (whose school is yet to rise but is implied in the statistical style of the most modern theories) to the naturalization of Freud through his many displaced disciples; Jung remains distant, awry, European. Then it is the vision of the Viennese positivists, the worldly success of the Bauhaus in diaspora but its subordination to the momentum of American institutions (with 15 pages of architectural illustrations from Germany and America) and the flowering of *Kunstgeschichte* American style. Even this list does not present all that is in the volume. Americans who want to understand any part of the American intellectual edifice today, particularly those who personally recall its growth over the past few decades, will need to seek out the relevant piece here. Missing, though, is a clear antecedent for Herman Kahn.

Let Szilard have the last word; it is a word we need to grasp. "Hitler came into office... and I had no doubt what would happen.... I noticed that the Germans always took a utilitarian point of view. ... You see, the moral point of view was completely absent, or very weak.... On that basis did I reach the conclusion... not because the forces of the Nazi revolution were so strong, but rather because I thought that there would be no resistance whatsoever."

THE DESIGN OF DESIGN, by Gordon L. Glegg. Cambridge University Press ($4.95). DESIGN BY ACCIDENT, by James F. O'Brien. Dover Publications, Inc. ($3). To link these two books is somewhat more than a play on the words in the titles: it is an effort to join disparate yet kindred audiences. The first book is a small, personal, generally wise and only occasionally stilted set of half a dozen lectures on the nature of engineering design. Glegg is a consulting engineer of wide experience, familiar with machine design, structures, materials, processing and control. His method is to present a set of cases from his own experience, thoughtfully viewed and made to cohere into a statement of principles for design. He starts at the proper beginning: the problem. The designer is wrong to accept the problem uncritically; he needs to be wary both of the impossible and of the unspecified obvious. The words "quiet," "light," "cheap," "typical," "instantaneous" and "simultaneous" are dangerous.

After the problem comes the designer. His mind needs three qualities: the inventive, the artistic and the rational. Each category has a chapter of exemplification. The inventive has a major rule: "Do not be conditioned by tradition.... The subconscious conditioning of engineering by locomotive design may have started with the model railway in the nursery." A recent success is a small rail conveyor whose wheels have a vertical axle! Innovation must be unforced; it was a sculptor who made the ball-point pen and a cork salesman who gave us the safety razor. The artistic is more subtle still; it is a matter of style. The hula hoop, the watch escapement and the buzz-bomb engine clearly have it. There in the blue Surrey sky the author watched the powerful and costly fighter planes barely catching the cheap buzz bombs. The pulse-jet engine of the buzz bomb transformed energy in a stream; the piston engine of the fighter plane did so through a dozen interfaces of oil and metal on the way to drive the plane. The bicycle wheel is a "neat partnership of materials and mathematics," using hard-drawn wire always in tension yet transmitting torques by tangential mounting to the hub. The bicycle wheel is to the cartwheel as a slender suspension span is to the thick-tubed bridge over the Firth of Forth.

The rational has two spheres: computer logic and human judgment. Computerize as much as you can, most of all to release young designers from detailed tasks that are never inherently difficult but are often faultily solved. Once the main idea is present, a step-by-step, careful, logical path will lead to success. Here, however, enters the greatest enemy of the designer: the assumption masquerading as a logical conclusion. If a hook tightly bolted to a plate lifts a weight, does its shank bear more tension? Yes if the shank is rigid and the plate that holds it is elastic; no if the plate is rigid and the shank is elastic. It is not the calculation that can go wrong; it is the assumed nature of the problem. It is better to check. Beware particularly the commercial "obvious need." Is there a need?

Finally, no one corrects the designer. He must be critical, even though he is personally involved, and jealous of his inventions. The Mulberry harbors—huge concrete barges that were floated to the Normandy invasion coast and sunk as breakwaters—had a terrible flaw that was found by tests just in time. They had a lengthwise central partition and flood valves on both sides. Naturally the two halves filled unevenly in the swell of the sea; the breakwater was liable to sink uselessly sideways. It is remarkable how this engineer's homily comes to be pretty good advice for almost any human activity.

The second book is very different. It celebrates the modern understanding that "Chance is beloved of Art, and Art of Chance" (cited by Aristotle from Agathon). The book contains some 200 photographs of design by chance, organized in nine classes by the physical means of production. There are tree forms, where pigment is pressed, blown or dragged across surfaces; crack patterns, where coats of paste and glue are allowed to dry in tension and are made visible with pigment; drips and dribbles in the mode so brilliantly used by Jackson Pollock; fallen droplets of color in starred dots or broad splashes and runs, and folded and crumpled sheets of paper and plastic. A few photographs are in color. Together with the striking designs are many smaller photographs showing analogues in nature, often on a quite different scale, for instance the treelike pattern of the Amazon and its tributaries. The most unexpected example is a design made by the soot from a candle flame deposited in engrossing ways on a piece of drawing paper. Details of the materials and methods are given, but it would be ingenuous to believe that these delicate processes, so complex and so dependent on the details of the environment and unspecified subjective decisions, could be repeated in any but the most general way. Still, the practice might bring designers from both cultures a great deal of pleasure and some insight. The author-artist himself has missed an opportunity by paying almost no attention to the physical reasons for the effects he gets.

BIOLOGIE ET ÉCOLOGIE DES PREMIERS FOSSILES, by Henri Termier and Geneviève Termier. Masson et Cie, Éditeurs (66 francs). All books on paleontology have a page on which is entered the geologic column, stretching from 600 million years ago, the time of trilobites, to our epoch. This brief work in French, by two prolific paleontologists of Paris, is different. It too devotes a page to such a column, but here the uppermost entry is marked with that same date and reads "Apparition du squelette." Below that line we go down through a dozen rather detailed entries until we reach the appearance of life, placed between 3.1 and 3.8 billion years ago. The center of interest in this little treatise lies in the deep past, although the story proceeds until insects and mammals arise to populate the continents. The motif of the work is the slow spread of life from one ecological zone to another, starting with the beginnings in shallow waters—never, or very rarely, dry—and spreading outward to the continental shelf and inward to the regularly dried intertidal region. From there life invested the emerging lands, first in and around the deep inland seas and finally in the rivers and on the solid ground. Last of all, it is argued, life spread to the deep sea and by now even to the true abyss. The watery tale is told with proper background: the physical circumstances, the methods of inference from fossil assemblages, the present needs and behavior of the important surviving members of the ancient lineages, and then the history.

There is a neat summary of trilobite life. Those eyes were not always blind in stone; they were compound eyes of a relatively few large elements, with a not very well defined organization. They could not have formed clear images; their visual field seems to have been made up of long, narrow bands, displaying objects in the horizontal plane.

The longest chapter in the rise of life was written in layers of the finest sediment, intimately interleaved with the remains of the microscopic organisms that laid it down year by year in columns and masses. In the most ancient times these thin layers of lime and dark carbonaceous chert record the colonies of blue-green algae and their many microscopic allies. These are the forms of the shallow salty waters; they are the only life we know with certainty during the first three-quarters or more of the planet's history. Sections and masses of these structures, called stromatolites, appear in the interesting illustrations reproduced here. It is curious that such structures grow even today, and that the widespread coral reefs, built by colonial marine animals rather than by algae, mirror the same architecture, although in a style more baroque. The first signs of nonmicroscopic animal life, so far only tracks and burrows, seem to begin in the stromatolites of 1.5 billion years ago. Once skeletons are invented—first, surely, an exoskeleton—the fossil pace quickens, and soon we reach the conventional post-Cambrian era, which occupies most even of this book. The long, soft past is well hidden.

There are obvious imperfections in this original work. The authors are willing to believe the fancies of some physicists, having learned too well the lesson of the physicists' erroneous rejection of continental drift. The theory of an expanding earth, however, should meet skepticism even deeper than the question mark it gets here. On the other hand, not one micrograph, let alone an electron micrograph, is presented to represent the marvelous microflora of the most ancient cherts. The material is cited indirectly, but the authors appear impressed less by the wonderful recent finds than by the biochemical residues demonstrated in the same ancient rocks. An English version of this unique book, perhaps with these flaws repaired, is much to be wished.

VESTIGES OF THE NATURAL HISTORY OF CREATION, by Robert Chambers. Introduction by Sir Gavin de Beer. Humanities Press, Inc. ($7.50). In July, 1844, Charles Darwin completed his 200-page sketch of the origin of species. That piece held the substance of the great volume he was to publish 15 years later, only after he had been shocked into decision by the paper from Alfred Russel Wallace. In October, 1844, *Vestiges,* a speculative system of the world centered on evolution, was published. Its painstakingly anonymous author was a self-taught successful publisher and amateur geologist of Edinburgh. The book was on everyone's mind; it ran through 10 editions in 10 years. It displayed the name of its true author only in the edition published in 1884 after Chambers' death, the last edition before this facsimile version. Evolution according to Chambers is not Darwin's evolution; there is no mechanism at all. Vague references to the law of organic development are taken as being satisfactory. Nonetheless, the argument is lively, and it has a genuine sweep of unity, from the stars that are suns to the nebular hypothesis to the rise of fossil life to the succession of living forms up to the family of the simians, who develop into man. "The human race is *one,*" Chambers adds.

In Chambers' book the astronomy and geology are deft enough, the biology wide-ranging but naïve, the sociology-anthropology struggling, the sympathy real. Darwin put it well: "That strange, unphilosophical, but capitally written book . . . has made more talk than any work of late, and has by some been attributed to me." Students of Darwin and his times will find the book indispensable. Here is one ringing extract: "The inorganic has one final comprehensive law, GRAVITATION. The organic, the other great department of mundane things, rests in like manner on one law, and that is—DEVELOPMENT." There is also an aside (based on a theory of sea-limited migration) that may win readers: "The United States might be expected to make no great way in civilization until they be . . . peopled to the Pacific; . . . the greatest civilizations of that vast territory will be found in the peninsula of California."

MATHEMATICAL IDEAS IN BIOLOGY, by J. Maynard Smith. Cambridge University Press ($5). Brief, good-humored and incisive, this is a first-rate phrase book for anyone with a little knowledge who wishes to use mathematics as a language. Its concern is to "explain how equations can be written down [rather] than to describe how they can be solved." The author is a well-known population geneticist, and there are two short chapters on that not very easy subject; they begin with the foundations of probability and continue to an account of the recurrence relations of selection theory and how to solve them. The usual biological statistics are not treated; the topic already has many good texts. Indeed, significance tests, Professor Smith says with justice, "have been oversold." The coupled differential equations of feedback theory, the basis of the oscillations seen in predator-prey populations, enzyme concentration and forearm position—are here in all those contexts, with an informal account of the nature of the results. The famous Turing proposal for the banded results of diffusion is presented as well, and the book's most abstract level is reached with a page—too brief—on the motion of system points in phase space, a really useful method of organizing the understanding of those nonlinear oscillating systems. The antistatistical mood is broken by a fine chapter on target theory and the usefulness of the Poisson distribution—quite without any fancy tests.

The book begins on the high note of the consequences of scaling, carried through to the well-known results—all animals, big and small, can jump the same height—and beyond them to running, diving, flying and the study of gaits. The number of results that appear in these first few pages are splendid advertisements for the whole approach. The book has 17 commonsense final pages on the honest practice of calculus, a workable refresher course, although too hard for a starter. The problems are happily chosen, and terse answers are given.

The style of the work is exemplified by the section on the control of limb movement, which demonstrates clearly that antagonistic muscle pairs without velocity-dependent forces must mean not control but oscillatory hunting. Inertia is the root cause of such mechanical jitter, and its absence (in examples as wavy as gene-frequency changes and predator numbers) means that intrinsic delays must enter the feedback loop.

THE SAGER WEATHERCASTER, by Raymond M. Sager. October House Inc ($5.95). Almost every household barometer has forecasts boldly written on the dial: fair for high pressure and rainy for low. There is reason even in this oversimplified approach. This small, well-made booklet, containing a set of concentric plastic dials, carries the idea of the barometer face to a much more adequate degree of complexity. You set the dials for the cloud state, the barometer pressure and its change, and the wind direction. Out comes a code number, one of some 6,000, and keyed to that number you build up a forecast for the next day or so, learning the weather, the wind speed and the wind direction. It has been in use since 1942, and it seems practical for many people and fun for even more. The device is designed for the northern Temperate Zone; one feels it should be adjustable for region. As it stands, it seems to represent East Coast experience best. No hurricanes out of a west wind! The developer is a forecaster of long New York City experience. A few more pages outlining the theory and the regularities behind the design would have been welcome.

LONDON LABOUR AND THE LONDON POOR, by Henry Mayhew. With a new introduction by John D. Rosenberg. Dover Publications, Inc. ($14). A new four-volume paperbound facsimile of the 1861 edition of this teeming, personal, etched, nonfiction Dickens of a book, with all the contemporary engravings. A single volume of selections prepared by John L. Bradley was reviewed here at length in July, 1966. Volume III treats of the street folk, the happiest of whom is the telescope exhibitor and the proudest a stilt walker who claims to have been the first ever seen in Ireland, not so very long before Yeats's poem.

INDEX OF ADVERTISERS

AUGUST 1969

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

MILITARY TECHNOLOGY AND NATIONAL SECURITY

THE ESSENCE OF SECURITY. Robert S. McNamara. Harper & Row, Publishers, 1968.

ABM: AN EVALUATION OF THE DECISION TO DEPLOY AN ANTIBALLISTIC MISSILE SYSTEM. Edited by Abram Chayes and Jerome B. Wiesner, introduction by Edward M. Kennedy. Signet Books, 1969.

STRATEGIC AND FOREIGN POLICY IMPLICATIONS OF ABM SYSTEMS: HEARINGS BEFORE THE SUBCOMMITTEE ON INTERNATIONAL ORGANIZATION AND DISARMAMENT AFFAIRS OF THE COMMITTEE ON FOREIGN RELATIONS, UNITED STATES SENATE. NINETY-FIRST CONGRESS, FIRST SESSION ON THE STRATEGIC AND FOREIGN POLICY IMPLICATIONS OF ANTIBALLISTIC MISSILE SYSTEMS. U.S. Government Printing Office, 1969.

"GENETIC DRIFT" IN AN ITALIAN POPULATION

FREQUENCIES OF PEDIGREES OF CONSANGUINEOUS MARRIAGES AND MATING STRUCTURE OF THE POPULATION. I. Barrai, L. L. Cavalli-Sforza and A. Moroni in *Annals of Human Genetics,* Vol. 25, Part 4, pages 347–377; May, 1962.

THE PROBABILITY OF CONSANGUINEOUS MARRIAGES. L. L. Cavalli-Sforza, M. Kimura and I. Barrai in *Genetics,* Vol. 54, No. 1, Part 1, pages 37–60; July, 1966.

EXPERIMENTS WITH AN ARTIFICIAL POPULATION. L. L. Cavalli-Sforza and G. Zei in *Proceedings of the International Congress of Human Genetics,* edited by James F. Crow and James V. Neel. The Johns Hopkins Press, 1967.

METALLIDING

VAPOR-PLATING: THE FORMATION OF COATINGS BY VAPOR DEPOSITION TECHNIQUES. C. F. Powell, I. E. Campbell and B. W. Gonser. John Wiley & Sons, Inc., 1955.

THE ELECTROCHEMISTRY OF MOLTEN SALTS. Thomas B. Reddy in *Electrochemical Technology,* Vol. 1, Nos. 11–12, pages 325–351; November–December, 1963.

COATINGS OF HIGH-TEMPERATURE MATERIALS. Edited by Henry H. Hausner. Plenum Press, 1966.

THE SIZE AND SHAPE OF ATOMIC NUCLEI

NUCLEAR SIZES. L. R. B. Elton. Oxford University Press, 1961.

PHYSICS OF THE NUCLEUS. M. A. Preston. Addison-Wesley Publishing Company, Inc., 1962.

NUCLEAR STRUCTURE. Aage Bohr and Ben R. Mottelson. W. A. Benjamin, Inc., 1969.

THE PETROGLYPHS OF SIBERIA

THE LENA RIVER PICTURES: CLIFF-GLYPHS NEAR SHISHKINO VILLAGE. A. P. Okladnikov and V. D. Zaporozhskaya (in Russian). Moscow-Leningrad, 1959.

THE DEER WITH GOLDEN HORNS. A. P. Okladnikov (in Russian). Leningrad, 1964.

THE SOVIET FAR EAST IN ANTIQUITY: AN ARCHAEOLOGICAL AND HISTORICAL STUDY OF THE MARITIME REGION OF THE U.S.S.R. A. P. Okladnikov. The University of Toronto Press, 1965.

THE ANGARA RIVER PETROGLYPHS. A. P. Okladnikov (in Russian). Moscow-Leningrad, 1966.

IMAGES OF THE ANCIENT AMUR RIVER. A. P. Okladnikov (in Russian). Novosibirsk, 1968.

KERATINS

THE KERATINS. Harold P. Lundgren and Wilfred H. Ward in *Ultrastructure of Protein Fibers,* edited by Rubin Borasky. Academic Press, 1963.

THE CHEMISTRY OF KERATINS. W. G. Crewther, R. D. B. Fraser, F. G. Lennox and H. Lindley in *Advances in Protein Chemistry,* edited by C. B. Anfinsen, Jr., M. L. Anson, John T. Edsall and Frederic M. Richards, Vol. 20, pages 191–346; 1965.

THE KERATINIZATION OF EPIDERMAL SCALES, FEATHERS AND HAIRS. R. I. C. Spearman in *Biological Reviews,* Vol. 41, No. 1, pages 59–96; February, 1966.

THE WEDDELL SEAL

ANIMALS IN AQUATIC ENVIRONMENTS: DIVING MAMMALS AND BIRDS. P. F. Scholander in *Handbook of Physiology, Section 4: Adaptation to the Environment.* American Physiological Society, 1964.

AN ANALYSIS OF SOME BEHAVIORAL AND PHYSIOLOGICAL CHARACTERISTICS RELATED TO DIVING IN THE WEDDELL SEAL. Gerald L. Kooyman in *Antarctic Research Series, Vol. XI: Biology of the Antarctic Seas III,* edited by George A. Llano and Waldo L. Schmitt. American Geophysical Union, 1967.

DEEP DIVING. G. L. Kooyman and H. T. Andersen in *The Biology of Marine Mammals,* edited by H. T. Andersen. Academic Press, 1968.

RUDOLF DIESEL AND HIS RATIONAL ENGINE

THEORY AND CONSTRUCTION OF A RATIONAL HEAT MOTOR. Rudolf Diesel. Translated from the German by Bryan Donkin. Spon & Chamberlain, 1894.

DIESEL'S RATIONAL HEAT MOTOR. *Progressive Age,* Vol. 15, pages 575–578, December 1, 1897; Vol. 15, pages 602–607, December 15, 1897; Vol. 16, pages 5–6, January 1, 1898; Vol. 16, pages 30–35, January 15, 1898.

DIESEL'S RATIONAL HEAT MOTOR. E. D. Meier in *Journal of the Franklin Institute of the State of Pennsylvania,* Vol. 146, No. 4, pages 241–264; October, 1898.

RUDOLF DIESEL. Eugen Diesel in *From Engines to Autos: Five Pioneers in Engine Development and Their Contributions to the Automotive Industry,* by Eugen Diesel, Gustav Goldbeck and Friedrich Schildberger. Henry Regnery Company, 1960.

MATHEMATICAL GAMES

THE TEST OF SIMPLICITY. Nelson Goodman in *Science,* Vol. 128, No. 3331, pages 1064–1069; October 31, 1958.

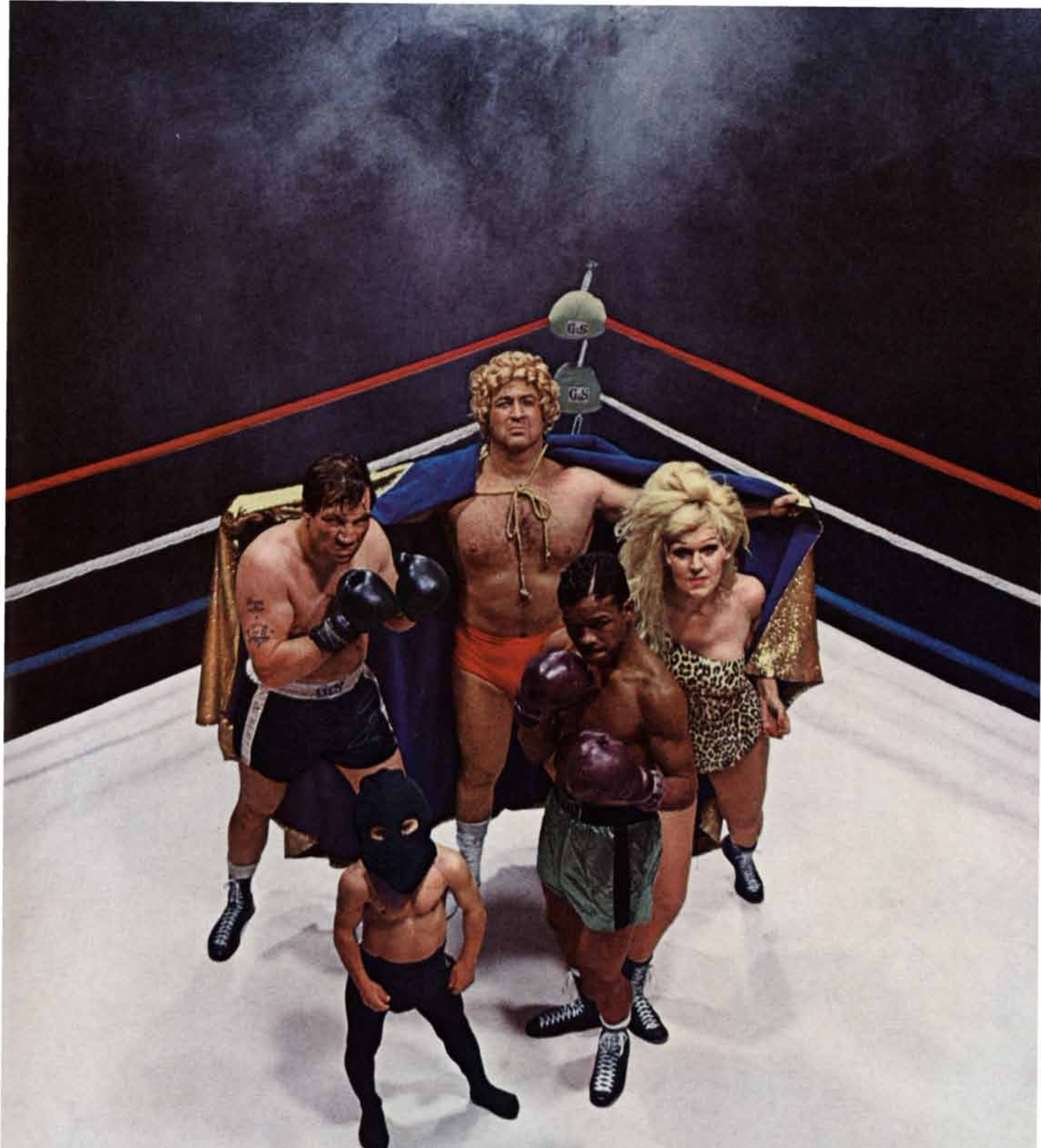
SCIENCE AND SIMPLICITY. Nelson Goodman in *Philosophy of Science Today,* edited by Sidney Morgenbesser. Basic Books, Inc., Publishers, 1967.

RAMIFICATIONS OF "GRUE." Mary Hesse in *The British Journal for the Philosophy of Science,* Vol. 20, No. 1, pages 13–25; May, 1969.

THE AMATEUR SCIENTIST

EYE, FILM, AND CAMERA IN COLOR PHOTOGRAPHY. Ralph M. Evans. John Wiley & Sons, Inc., 1959.

EXPERIENCES WITH COOLED COLOR EMULSIONS. Arthur A. Hoag in *Sky and Telescope,* Vol. 28, No. 6, pages 332–334; December, 1966.



Sometimes one head is better than two.

If you don't give lettuce seeds enough space, you wind up with two scrawny, stunted lettuces that nobody wants, instead of one big fat one. And the seeds are too small to space out well with mechanical planters. They come up in clumps, which you have to thin out by hand.

Now Union Carbide has developed a way to space the seeds out on tape and plant the tape. We make the tape of a plastic that dissolves in water. You plant it and water it and the seeds start to grow. Just the right distance apart.

Planting lettuce this way gives farmers big savings on thinning costs. And you can plant all the seeds at the same depth, so they come up all together. Which means harvesting costs less and yields are bigger.

Seed tape also works with tomatoes, celery, asparagus, sugar beets and other crops with small, light seeds. So we expect to be making a lot of it, along with all the other things we make to help people grow more food. Now if we could add something to it that speeds up germination....

UNION CARBIDE

THE DISCOVERY COMPANY