SCIENTIFIC AMERICAN

THE JOURNAL BEARING

$1.25

July 1975

The best GT car in its class.

Road & Track names Datsun 280-Z
"Best sports-GT car, $5,500 to $8,000"

The new fuel injected Datsun 280-Z makes its bow in America with a rare honor indeed. It's been named one of the 10 best cars for a changing world in the June issue of *Road & Track* magazine. At $6,284;* the Z took top spot in the "Sports-GT cars, $5,500 to $8,000" category.

The new 280-Z has computerized fuel injection for instant acceleration, great mileage and better emission control. The 280 carries on the Z-Car heritage of superior technology with its new 2800cc overhead cam engine, fully independent suspension and transistorized ignition, as well as a long inventory of standard comfort and performance features. Test-drive the new 280-Z and 280-Z 2+2.

*Manufacturer's suggested retail price for 280-Z two-passenger with standard 4 speed transmission, excluding tax, license, transportation and dealer prep if any.

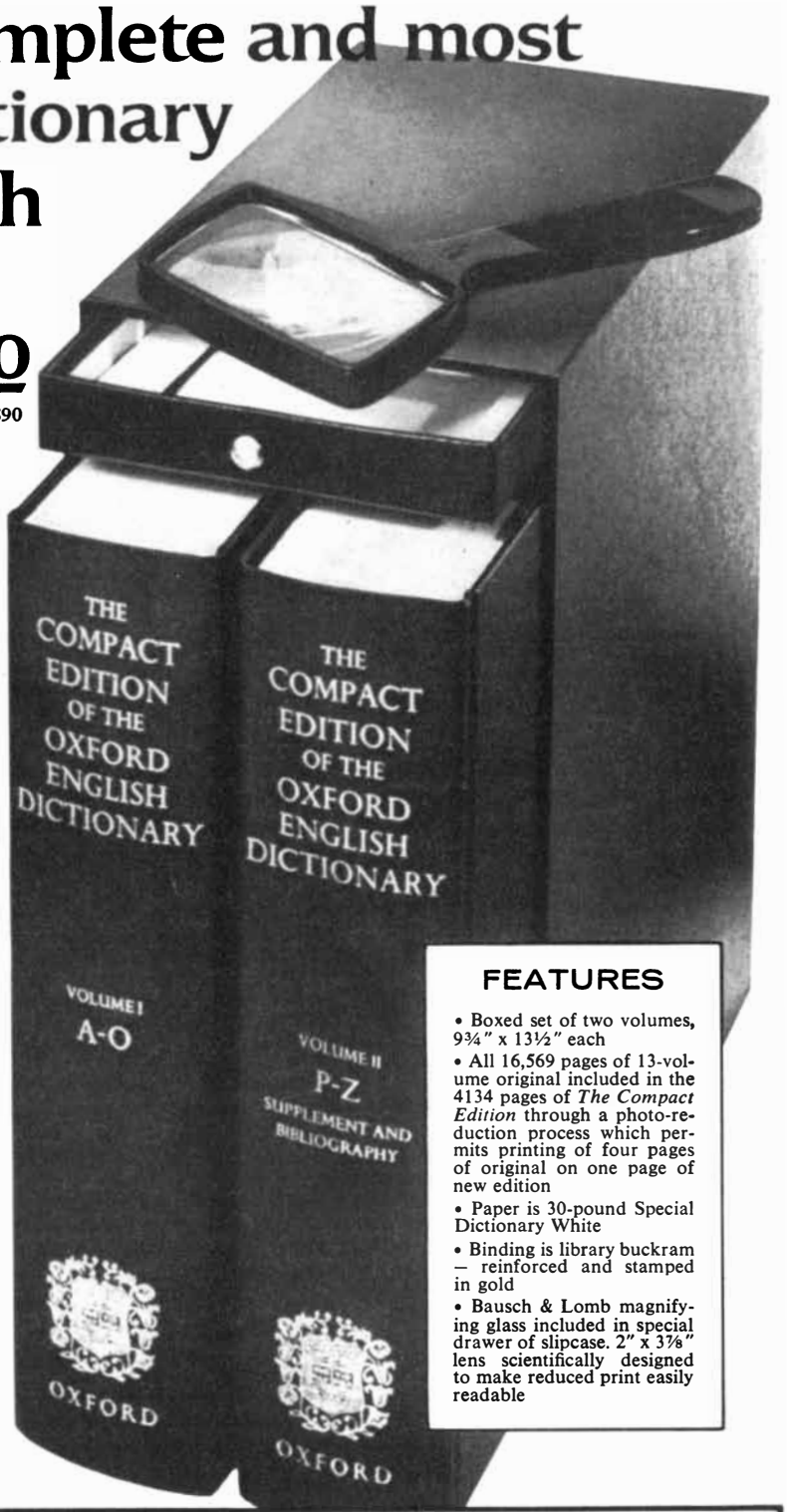
Datsun 280-Z

As your introduction to membership in the BOOK-OF-THE-MONTH CLUB®

The most complete and most scholarly dictionary of the English language for only $17.50

Publisher's list price: $90

THE SUGGESTED TRIAL: You simply agree to buy four Club choices within a year at substantial savings on most books you choose



THE OXFORD ENGLISH DICTIONARY is generally regarded as the final arbiter of the meaning, origin, history and usage of words in the English language. Until recently, it had been available only as a thirteen-volume set, priced at $350. Now, through the combination of an ingenious method of micrographic reproduction and a fine Bausch & Lomb optical lens, every single one of its 16,569 pages, fifty million words and close to two million illustrative quotations appears, in easily readable form, in the two volumes of *The Compact Edition*.

> The *New York Times* book critic Christopher Lehmann-Haupt has said of this edition: "It is something of a miracle. . . . *The Compact Edition* is easier to work with than the original with its 13 separate volumes."

Even more extraordinary, as a trial member of the Book-of-the-Month Club you may obtain *The Compact Edition of the Oxford English Dictionary* for only $17.50. And as long as you remain a member, you will receive the Book-of-the-Month Club *News*, a literary magazine announcing the coming Selection and describing other important books, most of which are available at substantial discounts – up to **40%** on more expensive volumes. All of these books **are identical to the publishers' editions in content, format, size and quality.**

If you continue after your trial membership, you will earn at least one Book-Dividend® Credit for every Selection or Alternate you buy. These Credits entitle you to obtain a wide variety of books, called Book-Dividends, at astonishing savings – *at least 70% of publishers' list prices.* Under this unique system, more than 100 Book-Dividends will be available to choose from every year, enabling you to acquire a well-rounded library at a small fraction of what you would otherwise have to pay.

FACTS ABOUT MEMBERSHIP

- You will receive the Book-of-the-Month Club *News*, a literary magazine published by the Club fifteen times a year. The *News* describes the coming Selection and scores of Alternates, and will be sent to you approximately every three and a half weeks.
- If you wish to purchase the Selection, do nothing and it will be shipped to you automatically.
- If you do not want the Selection – or you would like one of the Alternates, or no book at all – simply indicate your decision on the reply form always enclosed with the *News* and mail it so we receive it by the date specified.
- If, because of late mail delivery of the *News*, you should receive a Selection without having had 10 days to decide whether you want it, that Selection may be returned at Club expense.

FEATURES

- Boxed set of two volumes, 9¾" x 13½" each
- All 16,569 pages of 13-volume original included in the 4134 pages of *The Compact Edition* through a photo-reduction process which permits printing of four pages of original on one page of new edition
- Paper is 30-pound Special Dictionary White
- Binding is library buckram – reinforced and stamped in gold
- Bausch & Lomb magnifying glass included in special drawer of slipcase. 2" x 3⅞" lens scientifically designed to make reduced print easily readable

Book-of-the-Month Club, Inc.
Camp Hill, Pennsylvania 17012

5-A170-7

Please enroll me as a member of the Book-of-the-Month Club and send me *The Compact Edition of the Oxford English Dictionary*, billing me $17.50 (in Canada $19 – publisher's price $99). I agree to buy at least four Selections or Alternates during the first year I am a member, paying in most cases special members' prices. My membership is cancelable any time after I buy these four books. A shipping charge is added to all shipments.

MR. / MRS. / MISS 53
(Please print plainly)

Address ..... *Apt* .....

City .....

State ..... *Zip* .....

Western Electric Reports:

Moving phone calls bit by bit.

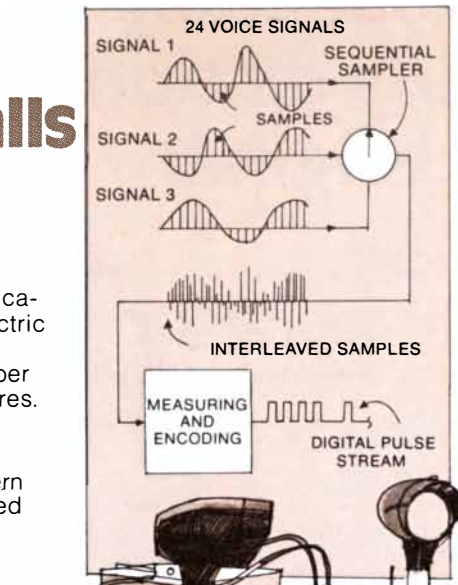
To meet the growing demand for communications facilities, the people at Western Electric and Bell Labs have developed digital techniques, which dramatically increase the number of phone calls that can be carried over existing wires.

In digital communications, a voice signal is sampled 8,000 times a second. Each sample represents the amplitude of the voice's wave pattern on a scale from 1 to 256. This measurement is coded in binary form as a series of pulses or "bits." And the code is transmitted to the receiving end where it's decoded to faithfully recreate the voice. Because this is a sampling technique, the pulses representing a number of voice signals can be interleaved. For example, the T1 System, workhorse of the Bell System's evolving digital network, transmits 24 simultaneous conversations on two pairs of wire.

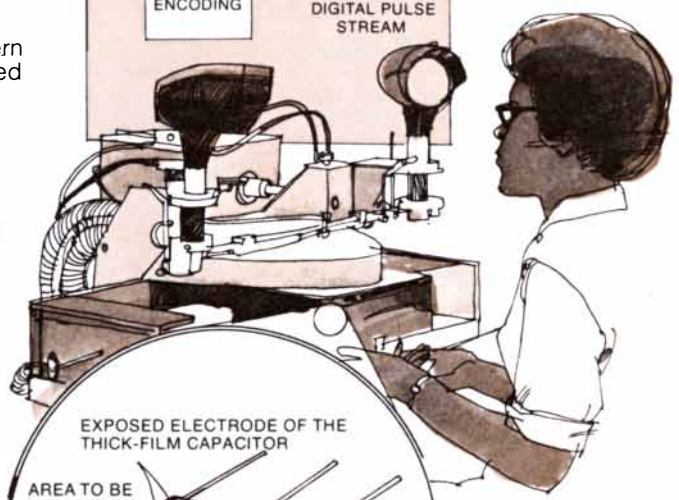
Development of digital techniques has demanded close coordination between designer and manufacturer. Interleaving 24 conversations on wire pairs originally intended to carry a single voice signal meant designing the T1 System to fit the characteristics of cable already in place. It meant manufacturing components that operate with clockwork precision, since the system must transmit a "bit" precisely every 648 nanoseconds. (The time it takes light to travel about 650 feet.) And because the stream of pulses must be regenerated at about one mile intervals—often in manholes under busy city streets—the components must be extremely stable.

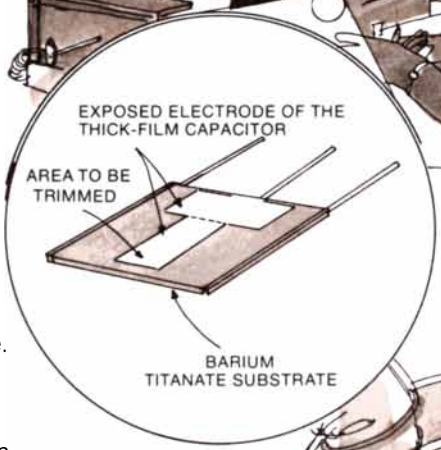
Engineers at Western Electric's plant in Massachusetts are working with Bell Labs on a wide range of design and manufacturing innovations. For example, previous timing circuits used in the regenerator for the T1 System were tuned manually. Western Electric engineers have developed a computerized process that tunes the circuits faster and more accurately. Meanwhile, Bell Labs has developed even higher capacity digital systems. The latest can interleave 4,032 simultaneous conversations on a pair of coaxial conductors.

Benefit: Digital communications techniques are one more way the Bell System is working to meet your communications needs reliably and economically.

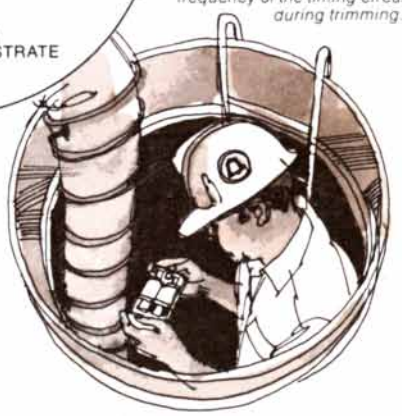


The T1 System samples 24 voice signals and encodes the measurements in binary form for transmission over a conventional pair of telephone wires as a stream of pulses.





The timing circuit is an inductor-capacitor. It is brought to a specific frequency by abrading the exposed electrode of the thick-film capacitor. A computer controls the process by measuring the frequency of the timing circuit during trimming.



The automatically adjusted timing circuit helps make the latest regenerator smaller, less expensive and even more reliable than its predecessors.

Western Electric

We're part of the Bell System.
We make things that bring people closer.

Established 1845 **SCIENTIFIC AMERICAN** July 1975 Volume 233 Number 1

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. 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: $15 PER YEAR, U.S., POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

THE COVER

The painting on the cover shows a journal bearing: the basic kind of bearing in which a shaft (the journal) rotates within a sleeve (the bearing). In some journal bearings the sleeve rotates on the shaft, or both the shaft and the sleeve have rotational motion (see "The Journal Bearing," page 50). The latter is the case with the bearing on the cover, a grooved journal bearing that is at one end of the connecting rod for a diesel engine. The rod links the crankshaft and one of the pistons of the engine, and it therefore is made with two journal bearings, one at each end. The grooved bearing carries as its journal a piston pin; the crankshaft is the journal in the other bearing. Most journal bearings are lubricated by a squeezing effect or a hydrodynamic effect on the lubricant in the small space between the journal and the bearing; such a bearing does not need to be grooved. Certain piston-pin bearings do not develop these effects because of the way they are loaded, and so they are made with grooves to facilitate the flow of lubricant. The inside diameter of this piston-pin bearing is 38 millimeters, or 1.5 inches.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
15–23	Jerome Kuhl
24	Stanley N. Cohen
26–30	Bunji Tagawa
32	Stanley N. Cohen
33	Bunji Tagawa (*top and bottom*), John F. Morrow (*middle*)
35–41	Allen Beechel
42	C. A. Burnham and Gordon L. Brownell
51	Ben Rose
52–60	Dan Todd
61	Willard W. Bach, General Motors Technical Center
62–63	Dan Todd
64	Thomas P. Schreiber, General Motors Technical Center
68	Lou Bory Associates
70	Ilil Arbel
71–72	Carol Donner
73–77	Ilil Arbel
81	Henry B. Garrett, Rice University
82–90	George V. Kelvin
92–93	Ralph Morse
94–95	Lorelle M. Raboni
96	Ralph Morse
97–100	Lorelle M. Raboni
102	Martin H. Zimmermann, Harvard Forest (*top*); Lorelle M. Raboni (*bottom*)
105	Edward E. Davis, Stanford Research Institute (*top*); Jack Colvard Jones, University of Maryland
106	Ilil Arbel (*top*), British Columbia Research Council (*bottom*)
107	British Columbia Research Council
108	British Columbia Research Council (*top*); F. E. Kellogg, British Columbia Research Council (*middle and bottom*)
109–110	Ilil Arbel
112	Escher Foundation, Haags Gemeentemuseum, The Hague
113–118	Ilil Arbel
121–124	Jerome Kuhl

QUALITY PAPERBACK BOOK SERVICE

The easy, economical way to buy the best paperbacks in print

More than 7000 distinguished new paperbacks are published each year. No wonder you miss so many good ones.

HOW CAN YOU KEEP TRACK of all the worthwhile paperbacks published every year? You can't. But we can. We're the Quality Paperback Book Service. We offer our members the pick of the paperbacks. And we offer a 20% saving on all Main Selections.

Now it will be easy to get all the good paperbacks you want. Because now there's the Quality Paperback Book Service. Join us today.

Any 3 books or sets for only $1 each

You simply agree to buy six more within the next year and as a member you'll immediately qualify for our Free Book Plan

HOW MEMBERSHIP WORKS

1. You will receive the *QPB Review* fifteen times a year—about every 3½ weeks. This informative catalog describes the Main Selection plus at least fifty Alternates. **The Main Selection is always offered at a 20% discount.**

2. If you want the Main Selection, do nothing. It will be shipped to you automatically. If you want one or more of the Alternates—or no book at all—just indicate your decision on the reply form always enclosed with the *Review* and return it by the date specified.

3. Free books. For every book or set you buy (exclusive of your three introductory choices for $1 each) you earn at least one Bonus Point, and the Bonus Points you accumulate entitle you to free books and sets. (You pay only shipping charges.)

4. Return privilege. If your *Review* is delayed in the mail and therefore you receive the Main Selection without having had ten days to notify us if you did not want it, you may return that Selection at our expense.

QUALITY PAPERBACK BOOK SERVICE, INC.
Middletown, Pennsylvania 17057

199. Pub Price **$3.95**

204. Pub Price **$5.95**

216. Pub Prices Total **$7.90**

132. CAT'S CRADLE, SLAUGHTERHOUSE-FIVE *and* BREAKFAST OF CHAMPIONS *by* KURT VONNEGUT, JR. (3 Vols.) Pub prices total **$7.95**

171. AMERIKA, THE TRIAL *and* THE CASTLE *by* FRANZ KAFKA (3 Vols.) Pub prices total **$6.85**

119. Pub Price **$8.85**

197. Pub Price **$4.95**

177. FIFTY YEARS OF RUSSIAN PROSE: From Pasternak to Solzhenitsyn. *Edited by* KRYSTYNA POMORSKA. (2 Vols.) Pub prices total **$7.90**

151. Pub Price **$8.95**

184. Pub Price **$4**

196. Pub Price **$4.95**

201. Pub Price **$3.95**

102. Pub Price **$4.95**

207. Pub Price **$5.95**

212. Pub Price **$4.95**

198. Pub Price **$7.95**

Prices shown are publishers' list prices

QUALITY PAPERBACK BOOK SERVICE, INC.
Middletown, Pennsylvania 17057 **5-QB170-7**

Please enroll me as a member of the Quality Paperback Book Service and send me the three books or sets whose numbers I've indicated in the boxes below. Bill me only $3 for all three, plus shipping. My only obligation is to purchase six more books or sets during the coming year, receiving a 20% discount on every Main Selection I take. My membership is cancelable any time after I buy the six additional books or sets. A moderate shipping charge is added to all shipments.

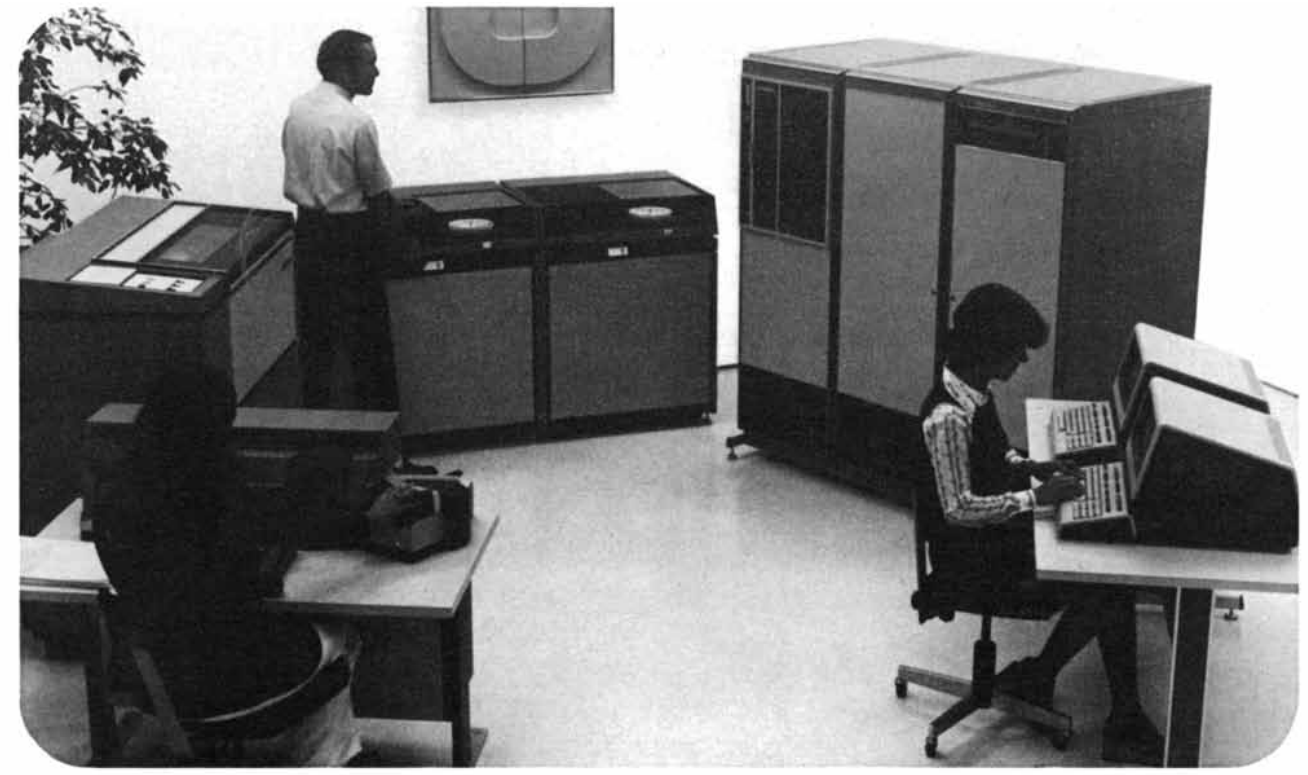
INDICATE BY NUMBER YOUR THREE BOOKS OR SETS [ ] [ ] [ ]

Mr. / Mrs. _____ 5
(Please print plainly)

Address _____ Apt. _____

City & State _____ Zip _____

Q-13



It says "Productivity" in five languages.

Multilingual and readily accessible from up to 32 terminals concurrently, the HP 3000CX Mini DataCenter saves users' time and avoids galling delays. Data base management is an important part of its impressive repertoire.

From data acquisition to administrative reports, HP's 3000CX Mini DataCenter greatly simplifies the entire range of computation tasks in a scientific environment.

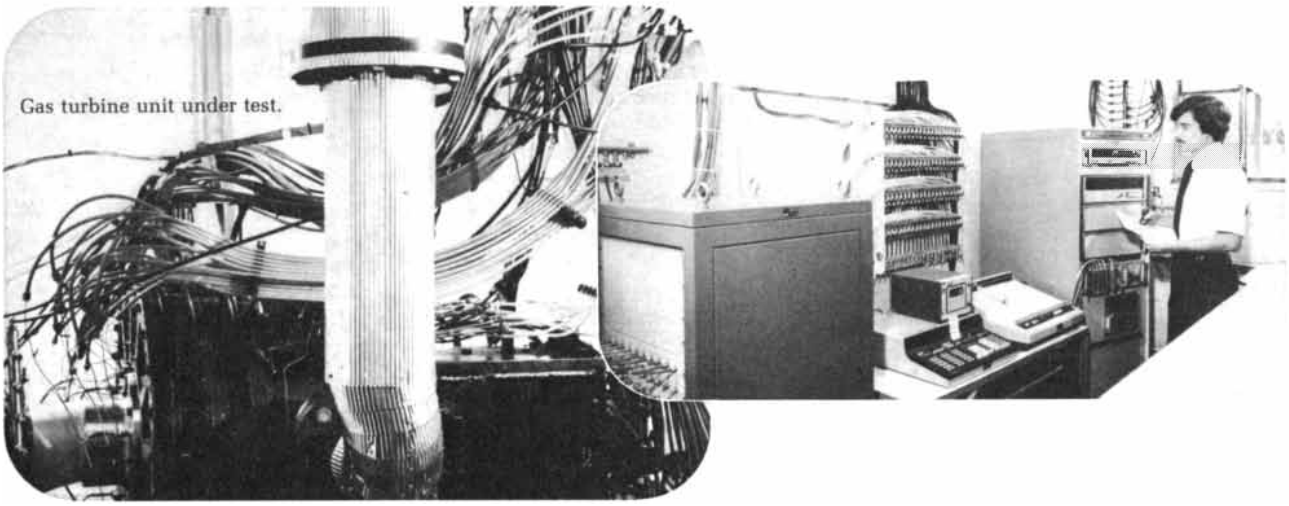
From as many as 32 interactive terminals, all users can access literally the entire system. Working with a single multiprogramming operating system and fully supported by advanced file management, they can develop programs; enter or retrieve data; perform complex inquiries and updates; access any peripheral; control remote instruments; and connect, if necessary, with central EDP.

Data base management lets all users store initial values and intermediate results in multiple data bases; generate conclusions from a terminal simply by entering logical expressions and ranges of values; and complete a search for related data in moments rather than days.

Mini DataCenters completely overcome the language barrier built into many scientific data systems. Any user can run a program in any of five languages, choosing precisely the right one to get the job done efficiently. A scientist may choose BASIC/3000 for data acquisition because it is the easiest language yet devised for interactive terminals and the most powerful implementation of BASIC currently available. Switch to FORTRAN/3000 for complex data analyses with scientific notation and formatted I/O; to COBOL/3000 for administrative processing; or to RPG/3000, which allows clerical personnel to produce reports quickly. And SPL/3000, HP's ALGOL-like system language, eliminates the need for assembly language routines by providing bit-level stack and device control. Programs can be prepared using more than one language—run in timesharing, real-time, or batch modes—and be shared by all users immediately after the first execution.

Through a Real-Time Programmable Controller, users can develop and execute data acquisition and priority-scheduled programs from their terminals, and transfer data between several HP data acquisition systems linked in a distributed system. A 2780/3780 Emulation Subsystem links Mini DataCenters to large EDP systems for access to centralized data as needed, leaving the full multilingual productivity of the 3000 systems available to respond immediately to the needs of the laboratory.

The 3000CX Mini DataCenter allows scientists to access, analyze, and report information, interactively from their own terminals, on their own terms. Full capability configurations start at $99,500*.



Calculator-based system speeds Chrysler's data acquisition.

At Chrysler, gas turbine research and development relies on an HP data acquisition system for improved data collection, computer-compatible format, and on-line data analysis.

One facet of Chrysler Corporation's pollution abatement research is the development of a gas turbine engine. Testing such an engine is complex: A single dynamometer test of turbine performance requires as many as 18,000 individual measurements of temperature, pressure, and running clearance at 300 points in the engine.

Under the circumstances, standard data collection techniques were taking too many men too long. Chrysler's research engineers decided to speed the pace of development, and chose a Hewlett-Packard 3050 Automatic Data Acquisition System as the most cost-effective solution to the problem.

The system, controlled by an HP 9820A programmable calculator, permits each dynamometer test to be completed by a single technician rather than two, and within three days instead of 12 or 13—despite the fact that the system handles the output of two dynamometer test cells simultaneously, and acquires about 40 percent more data from each cell than was acquired manually. An additional week's delay previously experienced at the computer center while test data were laboriously keypunched for computer analysis is now reduced to less than a day, because the HP system automatically punches the test data on paper tape in computer-compatible format.

As important as these improvements are, Chrysler engineers appreciate even more the HP system's capability to perform on-line data analysis. The system converts the raw transducer outputs into meaningful engineering units, provides continuous compensation for the thermocouple measurements, recalibrates the pressure transducers twice daily, and monitors test limits. Most importantly, it calculates engine work *vs* efficiency or slip factors and plots the results on-line. With this kind of information available *while the test is going on*, Chrysler engineers can adjust procedures as needed to ensure the validity of results.

The versatility of calculator-based systems can be further enhanced by the Hewlett-Packard Interface Bus (HP-IB), a standard interface system that allows interconnected system components to communicate effectively in an orderly and unambiguous manner.

Price for the HP calculator-based automatic data acquisition system starts at $14,100*.

For more information on these products, write to us. Hewlett-Packard, 1502 Page Mill Road, Palo Alto, California 94304.

Mail to: Hewlett-Packard, 1502 Page Mill Road, Palo Alto, CA 94304.
Please send me further information on

( ) HP 3000CX Mini DataCenter
( ) HP 3050 Automatic Data Acquisition System
( ) HP Programmable Desk-Top Calculators

Name _____ Title _____
Company _____
Address _____
City _____ State _____ Zip _____

HEWLETT hp PACKARD

Sales and service from 172 offices in 65 countries

*Domestic USA prices only.

00546

LETTERS

Sirs:

I should like to enlarge on the question of the therapeutic efficacy of thioctic acid touched on by Walter Litten in his elegant review of the amanitas and amanitin [SCIENTIFIC AMERICAN, March]. There is a little published evidence based on animal experiments with thioctic acid, and a fair amount of evidence has been accumulating for its use in man.

George L. Floersheim gave thioctic acid (five and 50 milligrams per kilogram of body weight) to mice before or after injection of alpha amanitin and found "no prophylactic or curative effect." F. R. Alleva gave thioctic acid (1.5 milligrams per kilogram) to dogs fed *Amanita phalloides,* with similar results.

In neither set of experiments was any effort made to prevent hypoglycemia, which appears as a toxic effect of thioctic acid and of amanitin itself. In Alleva's study hypoglycemia was indeed found in the dogs for which values were available on the day of death.

The experience in man may be summarized as follows. At the time of Jiří Kubička's summary article in 1968, 20 out of 69 patients who had received thioctic acid for amanita poisoning had died. Kubička's analysis convinced him that in these subjects the dosage had been too low, the initiation of treatment too late, or both. In any event at least 20 out of 69, or some 30 percent, of poisoning cases might be expected to die if there were *no* treatment.

In the years from 1969 through 1972 the European literature reports 43 cases of amanita poisoning in which thioctic acid was given in larger doses. Of these 41 lived.

Finally, the U.S. experience to date (not including the successful treatment of Clayton Brown, described by Litten) reveals that of 11 cases of amanitin poisoning treated with thioctic acid (from five to 30 milligrams per kilogram) and large amounts of glucose 10 have lived. The 11th was in coma and received only a single dose "at the time of cerebral death."

Because of these encouraging early results the Food and Drug Administration has approved the continued trial of thioctic acid. Supplies are available at the National Institutes of Health (301–496–6268 or 301–656–4000) and at the University of California Medical Center at San Francisco (Dr. Charles E. Becker, 415–648–6016).

FREDERIC C. BARTTER, M.D.

Clinical Director
National Heart and Lung Institute
Bethesda, Md.

Sirs:

The article "The Most Poisonous Mushrooms" in your March issue suggests that Leo J. Tanghe and Donald M. Simons identified the greenish mushrooms they collected in Rochester, N.Y., and southern New Jersey as *Amanita phalloides* solely by morphological criteria. I should like to report that I have analyzed toxins from mushrooms they have supplied me from the two locations. By chromatographing crude extracts on a column of Sephadex LH-20 and monitoring the optical density at 303 nanometers, three phallotoxins and three amatoxins showed up as distinct peaks. These toxins were then identified by thin-layer chromatography against known standards supplied by Heinz Faulstich of the Max Planck Institute for Medical Research in Heidelberg. Mushrooms from both locations contained phallacidin, phalloidin, phalloin and alpha-, beta- and gamma-amanitin in amounts comparable to those found in European *A. phalloides* by Faulstich.

ROGERS YOCUM

The Biological Laboratories
Harvard University
Cambridge, Mass.

Sirs:

My wife and I were most interested in Walter Litten's article on the amanitas in the March issue of *Scientific American.* For the past 10 years we have been finding, photographing, identifying and eating wild mushrooms. On November 7, 1970, I found several dozen *Amanita phalloides* mushrooms growing in a red pine plantation in Berks County, Pa. I took photographs, collected samples and subsequently received confirmation of my identification from Alexander H. Smith of the University of Michigan.

A week after this find I took my wife to the same plantation and there were several hundred *A. phalloides* spread over the area. Since then I have been able to find them every fall, spread over a much greater area.

My purpose in writing is that on page 94 of his article Litten says, "Rumor has it that *A. phalloides* makes a tasty dish." Since I found the mushroom so easy to identify even though I had never seen one before, I was curious as to why anyone would be damn fool enough to eat it. Accordingly I fried a small piece and can confirm the fact that it is delicious. There was no danger, because I did not swallow any, and there is nothing magical about mushroom poisoning. It is obvious that someone eats the mushroom, finds it delicious and says, "It tastes so good, it can't be poisonous."

C. B. MOORE

Moore Products Co.
Spring House, Pa.

ERRATUM

In "Earthquake Prediction," by Frank Press (SCIENTIFIC AMERICAN, May), the entire set of colored curves in the illustration on page 19 was inadvertently turned upside down. In each case the curve, indicating the anomalous uplift of the earth's crust, rises before the onset of the earthquake, indicated by the vertical arrow.

Scientific American, July, 1975; Vol. 233, No. 1. 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 secretary; George S. Conn, treasurer; Arlene Wright, assistant 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 Harry T. Morris, Advertising Director, 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 30 cents.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. For change of address, notify us at least four weeks in advance. Send both old and new addresses and enclose an address imprint from a recent issue. (Date of last issue on subscription is shown at upper right-hand corner of label.)

NAME

NEW ADDRESS

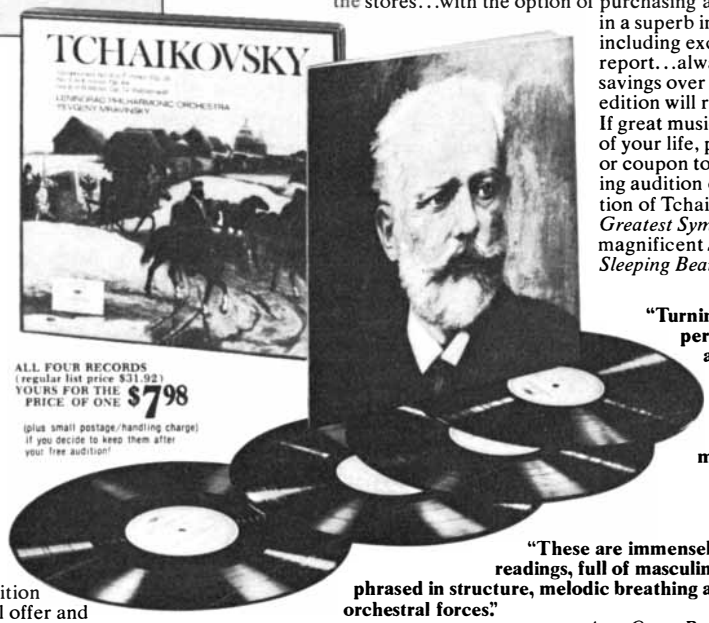
OLD ADDRESS

Now . . . a rare opportunity to audition the definitive interpretations of

Tchaikovsky's 3 Greatest Symphonies

A strictly limited edition of the master's incomparable Fourth, Fifth and Sixth Symphonies superbly performed on world-renowned Deutsche Grammophon recordings by Yevgeny Mravinsky and the Leningrad Philharmonic Orchestra

Unavailable in this country for over five years, this connoisseur's edition was exclusively commissioned for reissue by The International Preview Society:



Plus a Special Added Bonus: Tchaikovsky's Swan Lake and Sleeping Beauty Ballet Suites

Yevgeny Mravinsky's inspired renditions of Tchaikovsky's exquisite *Fourth, Fifth* and *Sixth Symphonies* have been hailed as perfection itself. And now, although these desirable recordings have been unavailable in this country for many years, they can be yours as a new member of the celebrated International Preview Society. You will also receive two other masterworks of extraordinary beauty, Tchaikovsky's legendary *Swan Lake* and *Sleeping Beauty Suites* with Herbert von Karajan conducting the Berlin Philharmonic Orchestra. You may audition these brilliant works for 10 days without charge or obligation. And you may *keep all four records for the price of just one*—a savings of a full 75% off their suggested list price! The strictly limited reissue of this edition was expressly commissioned for this special offer and is available *only* through The International Preview Society.

The Society's exclusive Selection Committee report, which accompanies the recordings, describes why these towering renditions were chosen for this special introductory offer. Here is a rare opportunity for music lovers to experience the *definitive* performances of Tchaikovsky's best-loved and finest symphonies. Under Mravinsky's masterful direction, the famed Leningrad Philharmonic achieves a sensitivity and precision beyond compare, performing these symphonic masterworks *as the composer meant them to be heard!* The genius of the dean of Russian conductors is richly conveyed by the superb Deutsche Grammophon records themselves—silent-surfaced European pressings manufactured to the highest quality control standards with the resulting realism of a live, concert hall performance. All four recordings come in a handsome hinged presentation case, and are accompanied by an illustrated brochure and an *exclusive* Committee report.

This introductory offer is indeed limited, so please mail the attached card or coupon on this page today. Send no money. Listen for 10 days... free! Then either return the albums and pay nothing, or keep them for only $7.98 (plus a small postage/handling charge). This is a savings of a full $23.94 *off* the retail price of comparable recordings!

Be the first to preview the greatest new recordings...free!

The International Preview Society is a unique record program dedicated to bringing you the finest classical compositions in their most splendid performances. Only when the Society's Carnegie Hall Selection Committee—headed by Julius Bloom, Executive Director of Carnegie Hall—discovers a new album of extraordinary merit and importance will it be sent to you for your free audition. By replying now, you assure yourself of the privilege of receiving, for future 10-day free previews, special *advance editions* of major new albums months before they will be available to the American public...all chosen exclusively for The International Preview Society by the Carnegie Hall Selection Committee. Since the Society is *not* a record "club," there is no minimum number of records you must purchase. Nor will you receive a "record-of-the-month" like clockwork. Only when the Committee discovers a new album of the most exceptional musical interest, artistic excellence and technical quality, will it be sent for your free audition. So there will often be months when no album at all is offered by The International Preview Society.

In every case, you will preview these records *long before* they reach the stores...with the option of purchasing any album you wish, in a superb imported pressing, including exclusive Committee report...always at substantial savings over what the ordinary edition will retail for months later. If great music is an important part of your life, please mail the card or coupon today for your stimulating audition of the ultimate rendition of Tchaikovsky's *Three Greatest Symphonies* and his magnificent *Swan Lake* and *Sleeping Beauty* ballet suites.

"Turning to the Mravinsky performances one is in another world. They tingle with the excitement and electricity that a great live performance generates."
Robert Layton, The Gramophone

"These are immensely powerful virtuoso readings, full of masculinity...beautifully phrased in structure, melodic breathing and balance of orchestral forces."
Ates Orga, Records And Recordings

All 4 Records For the Price of 1

...plus future free previews of important new recordings months before they are available to the public!

FREE TRIAL OFFER...SEND NO MONEY...
MAIL COUPON TODAY!

The International Preview Society
175 Community Drive, Great Neck, N.Y. 11025

Pleasesend, for my free audition, the superb recordings of Tchaikovsky's *Three Greatest Symphonies* by Yevgeny Mravinsky, plus the *Swan Lake* and *Sleeping Beauty* ballet suites by Herbert von Karajan. I may return all four records after 10 days and owe nothing, or keep them and pay just $7.98* (plus a small postage/handling charge). This is *$23.94 off* the retail price of comparable recordings, and this limited edition is available only from The International Preview Society. I will also receive, at intervals of one or more months, free 10-day previews of albums chosen by the Carnegie Hall Selection Committee, far in advance of general release. I may keep any album for far less than suggested list price, or return it without obligation. *I am not obligated to buy any minimum number of albums, and I may cancel this arrangement at any time.*

Name _____ (please print)

Address _____ Apt. No. _____

City _____ State _____ Zip _____

*Sales tax added for New York residents.

39506

Offer limited to the 48 connecting states. Offer expires 1/1/76. Limit: one membership per household. Only new members eligible NOTE: All applications are subject to review, and The International Preview Society reserves the right to reject any applications.

THE AUTHORS

KOSTA TSIPIS ("The Accuracy of Strategic Missiles") is at the Massachusetts Institute of Technology, where he teaches and does research in his two major fields of interest: high-energy particle physics and the arms race. A native of Athens, he came to the U.S. in 1954 to attend Rutgers University as a Fulbright scholar. After receiving his B.S. in electrical engineering and his M.S. in atomic physics from Rutgers he went on to obtain his Ph.D. from Columbia University in 1966. He joined the M.I.T. faculty that year and became a U.S. citizen a year later. Once at M.I.T., he writes, "I developed an increasing interest in the problems of science in public affairs and particularly in arms control. I became executive secretary of the U.S. Pugwash group in 1971, lectured at the Third Italian Summer School on Disarmament in 1970 and organized a symposium on the dynamics of the arms race in 1972. In 1973 I joined the Stockholm International Peace Research Institute (SIPRI) as a senior researcher and at the same time accepted the post of research associate at the Center for International Studies at M.I.T."

STANLEY N. COHEN ("The Manipulation of Genes") is associate professor of medicine at the Stanford University School of Medicine. A graduate of Rutgers University and the University of Pennsylvania School of Medicine, he joined the Stanford faculty in 1968 after spending several years teaching and doing research in molecular biology at the Albert Einstein College of Medicine. His research has also involved a stint at the National Institute of Arthritis and Metabolic Diseases. A specialist in bacterial plasmids, Cohen was a member of the National Academy of Sciences committee that recently called for the voluntary deferral of certain potentially hazardous experiments involving recombinant DNA molecules.

WERNER BRANDT ("Positrons as a Probe of the Solid State") is professor of physics and director of the Radiation and Solid State Laboratory at New York University. His doctorate is from the University of Heidelberg. After spending two postdoctoral years at the Niels Bohr Institute in Copenhagen, he came to the U.S. to work in the radiation-physics laboratory of E. I. du Pont de Nemours & Company. He has been at N.Y.U. since 1961. Among his main research interests are "phenomena initiated by the penetration of atomic radiation through matter." An earlier article by him for SCIENTIFIC AMERICAN ("Channeling in Crystals," March, 1968) covered another aspect of that subject.

JOHN C. BIERLEIN ("The Journal Bearing") is a project engineer at the Research Laboratories of the General Motors Corporation. He holds a B.S. from Michigan State University and an M.S. and a Ph.D. (in metallurgical and materials engineering) from the University of Wisconsin. "I am still continuing my education," he writes, "and have passed the halfway point for my M.B.A. at the University of Detroit in the field of management and finance." Among his leisure activities he reports that he enjoys "folk dancing with my wife, HO model railroad building and bicycling with my children and trying to predict the financial markets."

RICHARD J. WURTMAN ("The Effects of Light on the Human Body") is professor of endocrinology and metabolism at the Massachusetts Institute of Technology. He also lectures at the Harvard Medical School, where he received his M.D. in 1960. Before going to M.I.T. in 1967 he was associated with the Massachusetts General Hospital and the National Institute of Mental Health. His research covers a wide range of subjects, including "catecholamines and other biogenic amines, the pineal gland, the adrenal medulla, biologic rhythms and their control by environmental lighting, rhythms in the metabolism of amino acids, the physiologic control of menarche and ovulation in the human, neuroendocrinology and neuropharmacology, L-dopa and brain amino acid metabolism." He adds that by the time his article appears he expects to be "lying on a beach in Falmouth, Mass., making manifest my belief in the biologic value of sunlight."

ARTHUR A. FEW ("Thunder") is on the faculty of Rice University, where he teaches courses in atmospheric dynamics, air pollution and acoustics while pursuing his primary research interest in the electric and acoustic phenomena of the earth's atmosphere. He is a graduate of Southwestern University, where he majored in physics, and the University of Colorado, where he received his master's degree in basic science. He was a graduate student in the department of space science at Rice when, he writes, "my interest turned from stellar atmospheres to the problems of the earth's atmosphere." The mechanism of thunder subsequently became the topic of his Ph.D thesis.

THOMAS A. McMAHON ("The Mechanical Design of Trees") is associate professor of applied mechanics at Harvard University. After receiving his B.S. in engineering physics from Cornell University in 1965, he went to the Massachusetts Institute of Technology, where his Ph.D. thesis was "on the dynamics and fluid mechanics of the intra-aortic balloon, a device for providing mechanical assistance to the failing circulation." He reports that since he joined the Harvard faculty in 1969 his research has focused primarily on "the study of how the muscles work in running. The study of the mechanical design of trees is an outgrowth of work I started on scaling principles in both animals and trees." McMahon also writes fiction. His novel *Principles of American Nuclear Chemistry* was published by Atlantic Monthly Press and Little, Brown and Company in 1970.

R. H. WRIGHT ("Why Mosquito Repellents Repel") writes: "I am a Canadian and a graduate of the University of British Columbia (B.A., 1928) and McGill University (Ph.D., 1931). My professional career divides neatly into three parts: 15 years spent teaching chemistry at the University of New Brunswick; then 15 years as head of the Division of Chemistry at the British Columbia Research Council, mostly dealing with industrial problems, many having to do with air or water pollution, and finally 15 years (or very nearly) doing straight research on smells and smelling by both insects and vertebrates, from 1962 to 1971 as head of the Olfactory Response Investigation at B.C. Research (a title invented to indicate that I was being paid to do whatever I liked) and since then as a nearly whole-time interest and a part-time job. In 1972 I was in Greece and Crete for the Food and Agriculture Organization, studying attractants for the olive fly, *Dacus oleae*, and more recently I have been working with the Vancouver Police Dog Squad. Later this year a book on training the family dog will be published, having been written partly by me but mainly by my good friends Sergeant Campbell and Constable Laughy of the Vancouver Police."

DAVID LAYZER, who in this issue reviews *The Science and Politics of I.Q.*, by Leon J. Kamin, is professor of astronomy at Harvard University.

Digital Watch Breakthrough!

The new CDR display dramatically increases legibility and battery life and opens a new era of watch technology.

Would you do this with your solid-state watch? Of course not. Practically all solid-state watches require care and pampering. Not the Sensor 770. You can dunk it, drop it and abuse it without fear during its unprecedented five year parts and labor warranty.

At night or during the day, the Sensor's large, constantly "alive" CDR display is clear and easy to read.

A glance at your solid-state watch won't give you the time. Sound incredible? If it's an LED (light-emitting diode) watch, you've got to press the button first. If it's an LCD (liquid crystal display) watch, you must have plenty of light at just the right angle.

Now there's a new solid-state display technology called CDR (crystal diffusion reflection) incorporating the best features of the LED and the LCD displays. You can easily and constantly read your watch under any light conditions without strain or inconvenience.

The new CDR display takes the properties of the field-effect liquid crystal display, puts a strong reflective substance behind two closely-alligned polarization lenses, and the resulting large digits can be read clearly from practically any angle. When engaged, an integrated light source illuminates the display at night. The Sensor's constantly "alive" high-contrast display makes legibility outstanding under **all** light conditions.

Press the button on the Sensor 770 and the date and seconds appear in large black numerals—easy to read in any light.

A WORRY-FREE WATCH

Solid-state watches pose their own problems. They're fragile, they must be pampered, and they require frequent service. Not Sensor! Here are just five common solid-state watch problems you can forget about with this advanced space-age timepiece:

1. Forget about batteries Sensor is powered by a single EverReady battery that will actually last years without replacement. In fact, if your battery fails during the first five years, we will replace it free of charge. A low-power indicator tells you when to change the battery one month in advance and you simply open the hatch at the back of your watch and replace the battery yourself.

2. Forget about water Take a shower or go swimming. The Sensor is so water-resistant that it withstands depths of up to 100 feet.

3. Forget about shocks A three foot drop onto a solid hardwood floor or a sudden jar. Sensor's solid case construction, dual strata crystal, and cushioned quartz timing circuit make it the most rugged solid-state quartz watch ever produced.

4. Forget about service The Sensor 770 has an unprecedented five-year parts and labor unconditional warranty. Each watch goes through weeks of aging, testing and quality control before assembly and final inspection. Service should never be required, but if it should anytime during the five year warranty period, we will pick up your Sensor at your door and send you a loaner watch while yours is repaired—all at our expense.

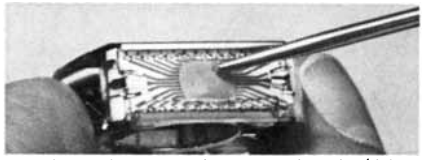
5. Forget about changing technology The Sensor is literally years ahead of every other watch in durability and technology. But should Sensor's technology improve anytime during the next five years, you may trade in your watch for Sensor's newer model under JS&A's liberal trade-in policy.

COMPARED TO EVERY OTHER

The $275 Pulsar uses the LED technology which requires pressing a button each time you want to review the time. Even the $500 solar-powered Synchronar watch, in our opinion, can't compare with the Sensor and its 5-year warranty. And no solid-state watch can compare to Sensor's quality, accuracy, ruggedness and exceptional value.

PLENTY OF ADVANCED FUNCTIONS

Sensor's five functions give you everything you really need in a solid-state watch. Your watch displays the hours and minutes constantly. Depress a button and your watch displays the seconds and date constantly. There's also an AM/PM indicator. To adjust the time, insert a ball-point pen into the four-channel time-control switch. Each channel independently controls one time function. In short, you can change the hours without affecting the date, and the minutes without affecting the hours.



A pin points to the new decoder/driver integrated circuit which takes the input from the oscillator countdown integrated circuit and computes the time while driving the display. This single space-age device replaces thousands of solid-state circuits and provides the utmost reliability—all unique to Sensor.

Sensor's accuracy is unparalleled. All solid-state digitals incorporate a quartz crystal. So does the Sensor. But crystals change frequency from aging and shock. And to reset them, the watch case must be opened and an air-tight seal broken which may affect the performance. In the Sensor, the crystal is first aged before it is installed, and secondly, it is actually cushioned in the case to absorb tremendous shock. The quartz crystal can also be adjusted through the battery compartment without opening the case. In short, your watch should be accurate to within 5 seconds per month and maintain that accuracy for years without adjustment and without ever opening the watch case.

STANDING BEHIND A PRODUCT

JS&A is America's largest single source of electronic calculators, digital watches and other space-age products. We have selected the Sensor as the most advanced American-made, solid-state timepiece ever produced. And we put our company and its full resources behind that selection. JS&A will unconditionally guarantee the Sensor—even the battery—for five years. We'll even send you a loaner watch to use while your watch is being repaired should it ever require repair. And our liberal trade-in policy guarantees that new watch technology will never leave you behind.

Wear the Sensor for one full month. If you are not convinced that the Sensor is the most rugged, precise, dependable and the finest quality solid-state watch in the world, return it for a prompt and courteous refund.

To order your Sensor, credit card buyers may simply call our toll-free number below or mail us a check in the amount indicated below plus $2.50 for postage, insurance and handling. (Illinois residents add 5% sales tax) We urge you, however, to act promptly and reserve your Sensor 770 today.

Stainless steel w/leather strap $99.95
Stainless steel w/metal band $109.95
Gold plated w/leather strap. $119.95
Gold plated w/metal band. $129.95

JS&A NATIONAL SALES GROUP

DEPT. SA 4200 DUNDEE ROAD
NORTHBROOK, ILLINOIS 60062
CALL TOLL-FREE. . . . (800) 323-6400
In Illinois call . . . . . (312) 498-6900

Established 1845 SCIENTIFIC AMERICAN July 1975 Volume 233 Number 1

The Accuracy of Strategic Missiles

The U.S. has initiated the development of ultra-accurate missiles capable of destroying the land-based missile force of the U.S.S.R. The technology behind the new "counterforce" strategy is reviewed

by Kosta Tsipis

On January 10 of last year Secretary of Defense James R. Schlesinger announced the intention of the U.S. to develop a new generation of long-range ballistic missiles capable of delivering nuclear warheads against distant targets with unprecedented accuracy. His announcement was widely interpreted at the time as signifying a decision on the part of the Nixon Administration to pursue a "counterforce" strategy, that is, a nuclear-war strategy that relies at least in part on the ability to destroy the land-based offensive missiles of the U.S.S.R. in their reinforced-concrete silos. That decision presumably continues to represent the policy of the Ford Administration.

What is the technological basis of the new counterforce strategy? In this article I shall review the procedures and devices currently available for achieving high accuracy in the delivery of strategic nuclear warheads. I shall also describe several new approaches to the problem, made possible largely by recent advances in the microminiaturization of electronic components, that may lead to significant improvements in missile accuracy. The new techniques apply potentially not only to the existing U.S. arsenal of land-based and sea-based ballistic missiles, which are unpowered during most of their flight, but also to an emerging new class of strategic weapons: continuously powered long-range "cruise" missiles. Finally, I shall attempt to assess the probable impact of the U.S. decision to acquire a counterforce capability on efforts to limit the strategic-arms race, on the future course of relations between the U.S. and the U.S.S.R. and on the altered likelihood that a nuclear war will break out between the two superpowers.

The task of delivering an explosive projectile accurately enough to destroy an intended target establishes several requirements: One must first be able to distinguish the target from its surroundings; one must then be able to determine the exact position of the target with respect to the launching point of the missile; finally, one must be able to launch and guide the projectile in such a way as to make the "error radius" (the probable distance by which the warhead would miss the target) smaller than the "kill radius" (the distance within which the explosive charge of the warhead would destroy the target). Obviously accuracy is a relative concept that involves factors other than simply the distance between the target and the point of impact of the warhead.

A nuclear warhead with an error radius of one kilometer, say, would be considered accurate if it were to be used against an airfield, since it would destroy such a target even if it landed one kilometer away; it would be considered inaccurate if it were to be used against a missile silo, however, because the silo would probably survive a nuclear blast at that distance. In the first case the kill radius would be larger than the error radius; in the second case the error radius would be larger than the kill radius. Thus accuracy depends on the "hardness" of a target. It also depends on the explosive yield of the weapon. A conventional 1,000-kilogram gravity bomb that lands 10 meters from a building will destroy it, whereas a 100-kilogram bomb that lands the same distance from the building will not destroy it and hence is not regarded as being accurate enough against that particular target. In general if the kill radius of a weapon is larger than or equal to the error radius for a given target, the weapon is said to be accurate, whereas if the kill radius is smaller than the error radius, the weapon is said to be inaccurate.

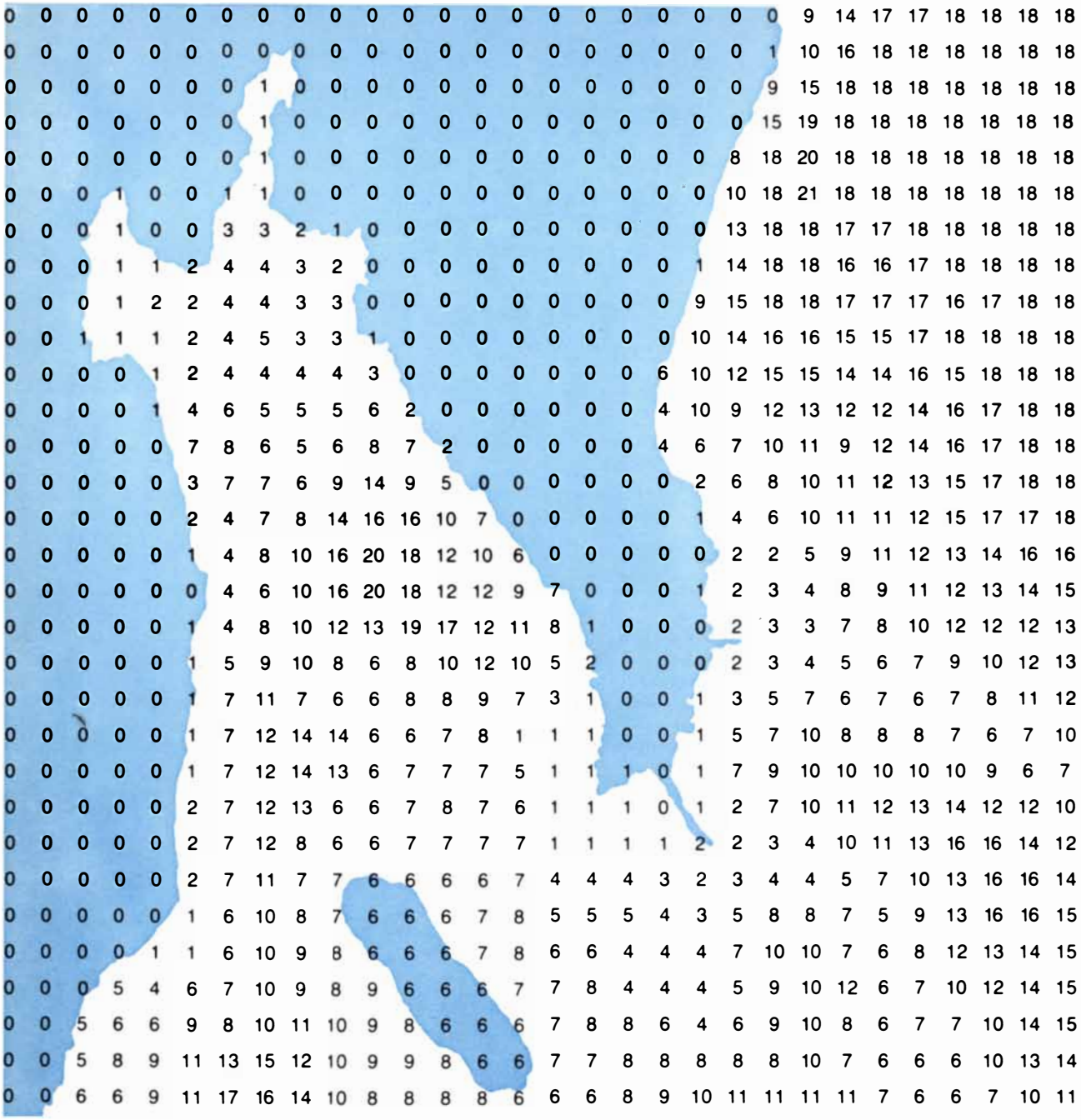
The significance of the relative magnitudes of kill radius and error radius can be illustrated best by considering how a nuclear explosion would affect a "soft" target, such as an airfield or a city, compared with how it would affect a "hard" target, such as a missile silo.

The destructive effects of a nuclear weapon result from the almost instantaneous release of an enormous amount of energy. The explosion of a one-megaton nuclear weapon, for example, releases within a few billionths of a second an amount of energy equivalent to the energy released by a million tons of TNT (some 10^{15} calories), generating in the immediate vicinity of the explosion temperatures on the order of millions of degrees Celsius. The intense heat causes

the air around the point of the explosion to expand suddenly, giving rise to a shock wave at which the pressure reaches as much as 100,000 pounds per square inch; the pressure then decreases abruptly as the shock wave propagates outward from the point of the explosion.

The same nuclear warhead can of course be used against either a city or a reinforced concrete missile silo, but the performance characteristics of the warhead and the missile that are relevant to each type of attack are quite different.

Shock-wave "overpressures" of five pounds per square inch will demolish an ordinary brick house. Since overpressure is proportional to the energy released by a nuclear charge and is inversely proportional to the cube of the distance from the point of explosion, the larger the energy yield of the weapon is,



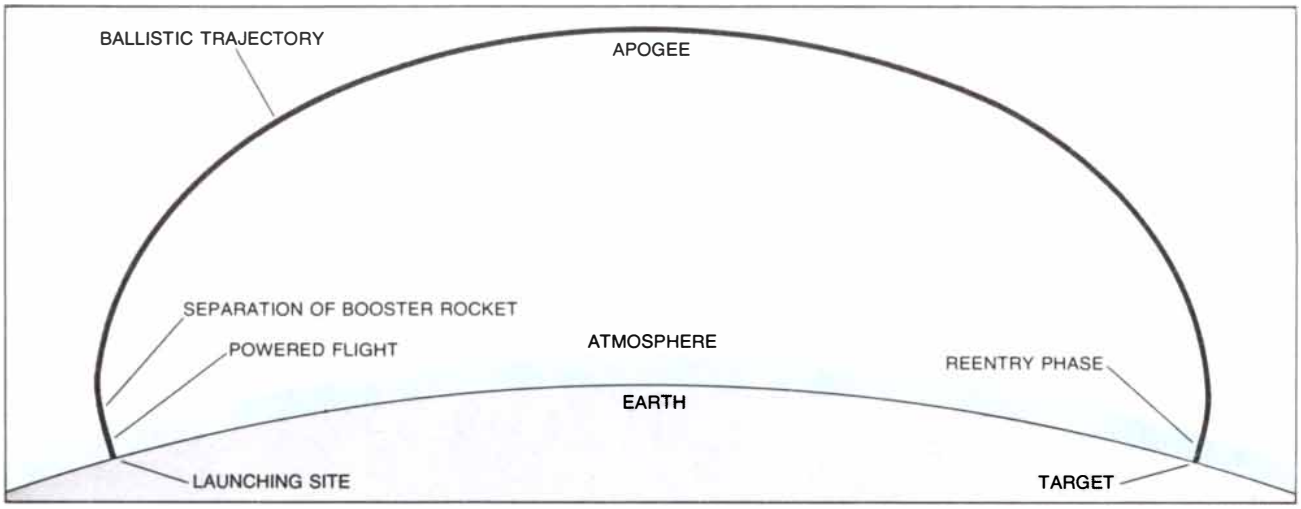
DIGITAL MAP of the terrain in the vicinity of Rockport, Me., consists of an array of numbers representing the variation of ground elevation above sea level as a function of location. Each number gives the average elevation (in multiples of three meters) of a square area 100 meters on a side. Water is shown in color. Maps with comparable resolution could easily be compiled with the aid of radar or laser altimeters on board orbiting reconnaissance satellites. A maneuverable nuclear warhead with a terminal-guidance system based on the principle of terrain-matching could have such a map stored in its computer memory in the form of sequences of numbers. By comparing the data stream from its own on-board altimeter with the stored digital map, an "active" guidance system of this type could supplement a missile's basic inertial-guidance system by generating corrections in the trajectory of the warhead as it reenters the atmosphere. Such a system, now under development in the U.S., should be capable of achieving near-perfect accuracy. (The map is oriented with north at bottom, which is presumably how an incoming nuclear warhead would "see" the terrain.)

the farther from its point of impact the perimeter of total destruction will be. A one-megaton nuclear weapon, for example, creates an overpressure of at least five pounds per square inch as much as four kilometers from the point of explosion; hence it would destroy all houses in an area of 50 square kilometers around its point of impact. Accordingly a missile aimed at a city does not have to be very accurate, since the loss of property and life caused by the weapon would be immense no matter where in the city it lands. Furthermore, the foregoing analysis does not even take into account the very widespread destruction likely to be caused by direct thermal effects and by radiation.

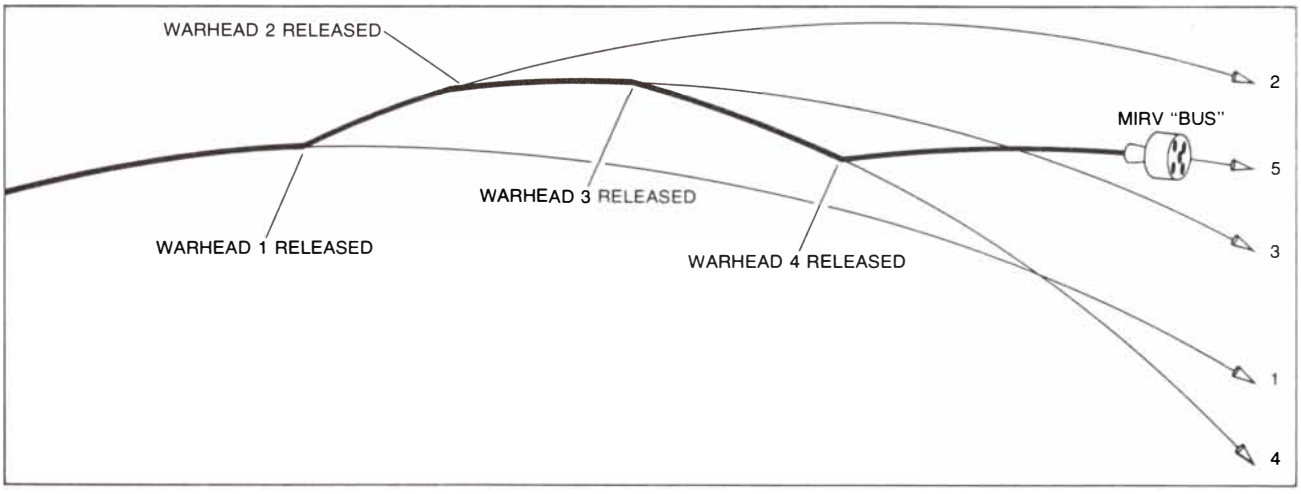
An attack against urban or industrial centers therefore requires the delivery of large amounts of thermal energy and the creation of modest overpressures over very large areas. This can be achieved either by delivering a high-yield nuclear weapon somewhere in the general neighborhood of the target or by scattering several small weapons at random over the area. These are the operational requirements of a "counter-city" nuclear-weapons system.

A nuclear-weapons system intended to implement a counterforce strategy has a different set of performance criteria. Here the weapon must deliver a very large amount of energy against a very small target. Since a missile silo is designed to withstand overpressures of hundreds of pounds per square inch (and is largely immune to thermal effects), it will remain intact unless the weapon lands in its immediate vicinity; such a weapons system must obviously be very accurate.

As a matter of fact, the lethality of a warhead directed against a silo rises



THREE DISTINCT SEGMENTS in the flight of a long-range ballistic missile are illustrated. In the first segment (*left*) the booster rocket accelerates the missile past the earth's atmosphere to an altitude of about 100 kilometers, at which point the rocket engine shuts off and the reentry vehicle bearing the nuclear warhead (or warheads) separates from the rocket. In the second segment (*middle*) the reentry vehicle "coasts" through space on a roughly elliptical ballistic trajectory under the influence of the earth's gravitational field. In the final segment (*right*) the vehicle reenters the atmosphere, where it encounters strong aerodynamic forces that "bend" the trajectory toward the vertical. A typical flight of 10,000 kilometers from launching site to target would last about 30 minutes.



MIRV SYSTEM employed by the latest U.S. strategic missiles is based on the use of a "bus" that carries the MIRV's, or multiple independently targetable reentry vehicles, beyond thrust termination into the ballistic segment of the missile's trajectory. The bus contains its own inertial-guidance system and small rockets that enable it to change velocity. At some point the guidance system causes the bus to release a warhead and then to change velocity; the bus continues to release warheads, changing velocity each time, until all the warheads have been released, each on a different trajectory toward a different target. The U.S. has already MIRVed some 850 of its land-based and submarine-based strategic missiles; the U.S.S.R. has just begun to MIRV some of its land-based missiles.

much more rapidly with improvements in accuracy than it does with increases in yield. Hence it is the accuracy of the missile delivering the warhead that is crucial in a counterforce nuclear arsenal, rather than the size of the missile.

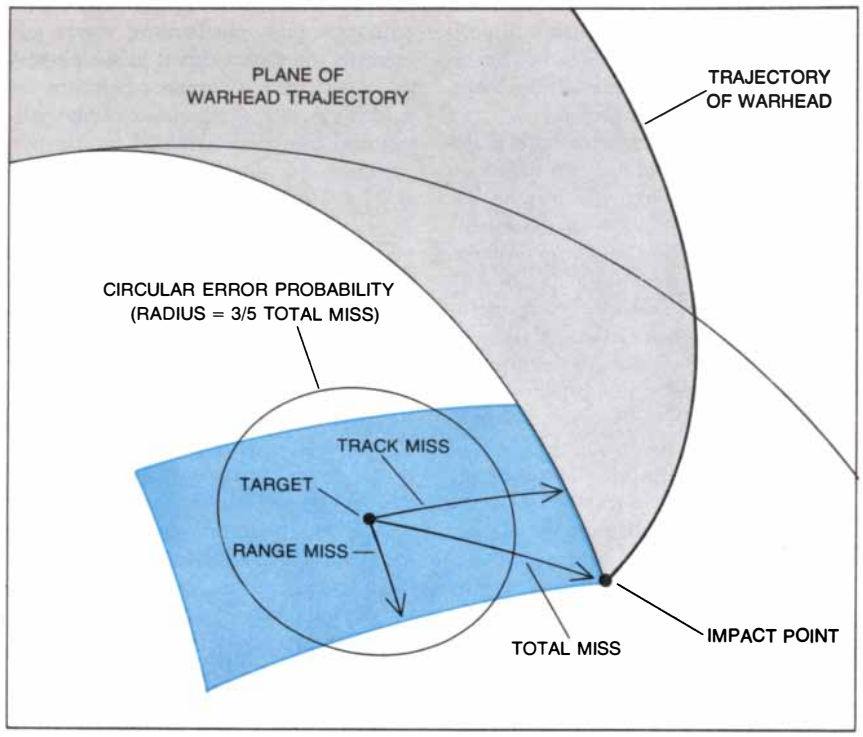
A ballistic missile consists of a booster rocket, arranged in two or more stages, with a reentry vehicle containing the nuclear warhead mounted on the top of the rocket. Once such a missile is launched it follows a trajectory that can be divided into three distinct parts [*see top illustration on opposite page*]. First there is the powered-flight portion, during which the rocket burns fuel, accelerating the missile through the dense lower atmosphere to a height of about 100 kilometers and the same distance downrange toward the target, whereupon the rocket is shut off and the reentry vehicle is separated from the booster. Next there is the ballistic-flight portion of the trajectory, during which the reentry vehicle, under the influence of the earth's gravitational field, "coasts" on an approximately elliptical path through the near-vacuum of space, reaching an altitude of some 1,000 kilometers before it starts falling back toward the earth. Finally, there is the reentry of the vehicle into the atmosphere, during which (in the last two or three minutes of its flight toward the target) the reentry vehicle experiences strong aerodynamic forces. For a typical distance of 10,000 kilometers between the launching site and the target such a flight would last about 30 minutes, of which the first six or eight minutes would constitute the powered boosting phase.

To deliver a warhead accurately against a target with a ballistic missile one must be able to guide the missile along a suitable trajectory that ends at the target. To do that the position of the missile must be known continuously up to the point where the booster separates from the reentry vehicle. With current technology the task can be accomplished in either of two ways: by radiation sensing or by inertial sensing.

In the radiation-sensing approach one determines the exact position of the missile by tracking it with radars and by receiving at several ground stations radio signals emitted by a device on the missile in response to commands from the ground. An alternative, also based on radio signals, is to create a "radio environment" with ground-based transmitters in which the missile can "navigate itself."

In the inertial-sensing approach one measures the forces other than gravity that influence the motion of the missile; then, using Newton's law that relates force to acceleration, one determines the acceleration of the missile as a continuous function of time. Since the integral of a component of the missile's acceleration over time gives the corresponding component of the missile's velocity, and since distance is the product of velocity and time, once the acceleration is known as a function of time the velocity and the position of the missile can be calculated with the aid of an on-board computer and an accurate clock, given the known velocity and position at the instant of the launch.

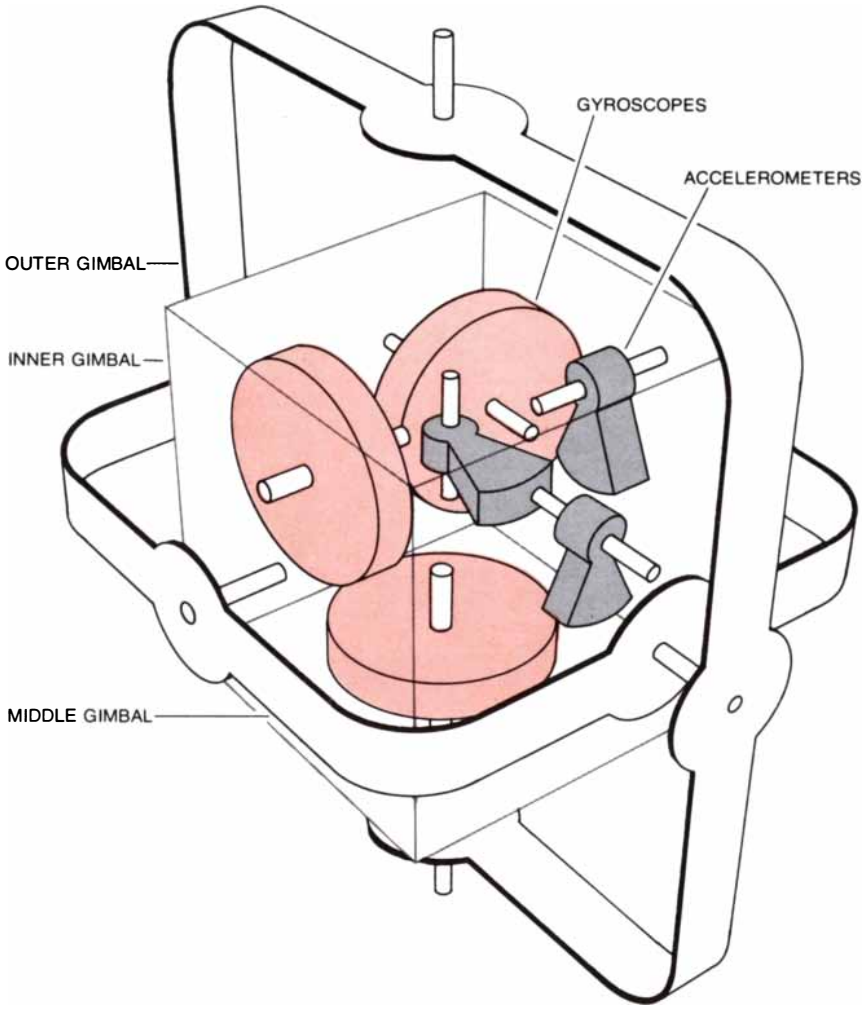
Because the ground installations necessary for radiation sensing would be vulnerable during a nuclear conflict, inertial sensing (which does not depend on signals from the world outside the missile) has been used for the guidance of most deployed strategic ballistic missiles. Inertial sensing employs an accelerometer to directly measure the forces acting on an enclosed test mass. Since acceleration is a vector quantity (that is, a quantity that has both magnitude and direction), three single-axis accelerometers with their sensitive axes at right angles to one another are installed in each missile [*see illustration on next page*]. In order to maintain the prelaunch alignment of the accelerometers during flight with respect to a fixed frame of reference on the earth's surface, either they are mounted on gimbaled platforms that are gyroscopically stabilized to maintain the original direction of each accelerometer or they are rigidly strapped to the frame of the missile. In the latter case additional accelerometers that measure angular acceleration are installed to determine the motion of the missile with respect to the fixed original coordinates, and a somewhat more complex on-board computer is used to convert the real-time linear accelerometer outputs to the original inertial frame of reference. In either configuration each accelerometer measures the magnitude of only one of three mutually perpendicular components of acceleration, from which the corresponding instantaneous velocity components of the missile are determined. In neither case, however, can the accelerometers measure the components of acceleration resulting from the force of gravity. The effects of this force on the vehicle must be determined analytically from the known magnitude and direction of the



"CIRCULAR ERROR PROBABILITY," or CEP, a standard measure of the accuracy of a long-range missile, is defined in geometric terms in this diagram. The probable maximum distance by which an impacting warhead can miss its intended target in the plane of its trajectory is called range miss; the comparable target miss at right angles to the trajectory plane is called track miss. (The trajectory plane is determined by three points: the launching site, the target and the center of the earth.) The total miss is the vector sum of the range miss and the track miss. Circular error probability is equal to three-fifths of the total miss.

gravitational force as a function of position above the earth's surface; the resulting data are then stored in the memory of the on-board computer.

There are two general methods to ensure that the reentry vehicle follows a correct trajectory that will lead to the target. One is to provide for the missile to have at the instant of thrust termination and booster separation specific values of the three velocity components that have been calculated in advance and stored in the memory of the missile's computer. The only quantities that can be controlled during the powered flight of the missile are the magnitude and direction of the propulsive force of the rocket. To affect the trajectory of the missile, therefore, the inertial-guidance system determines the acceleration, velocity and position of the missile and, by forming and sending commands to the automatic-pilot mechanism, steers and controls the rocket thrust in such a way as to make the differences between the actual velocity components of the missile and the predetermined termination velocities become zero at the same instant at a predetermined point in space. At that instant the rocket must be shut off. This method requires a minimal computational capability on board the missile, but it may lead to large inaccuracies because the magnitude of the rocket thrust is difficult to control. The second method involves controlling only the direction of the thrust and relies on a more sophisticated on-board computer to determine, given the launch and target coordinates and one more trajectory parameter, the termination velocity as a function of thrust history. All the solid-fueled U.S. ballistic missiles operate on the second method.



INERTIAL-GUIDANCE SYSTEM suitable for monitoring the exact position and velocity of a long-range ballistic missile continuously during the period between launch and thrust termination relies on data from three mutually perpendicular accelerometers; each of these extraordinarily sensitive devices measures the forces acting in a given direction on an enclosed test mass. In this design the accelerometers are mounted on a gimbaled platform that is gyroscopically stabilized to maintain the original direction of each accelerometer.

In either case the accurate delivery of a warhead by such a missile to a target depends on three additional factors: (1) the ability to specify the locations of launch and target points accurately in the same frame of reference, so that a set of suitable trajectories for the missile can be established; (2) the ability to steer the rocket thrust, so that the position and velocity of the missile at thrust termination are precisely those the missile must have in order to follow the chosen trajectory, and (3) the ability to predict the motion of the reentry vehicle during its entire flight, so that the velocity components at thrust termination can be correctly determined in advance. An inertial-guidance system of this general type cannot significantly affect the motion of the warhead during the reentry phase except by determining the angle, location and speed at which it will reenter the atmosphere.

None of the above conditions can be fulfilled perfectly. For example, shutting off the rocket thrust a thousandth of a second too late could result, depending on the terminal acceleration, in a miss of as much as 600 meters at the target. Errors due to imperfections of the accelerometers and gyroscopes, to possible misalignment of these instruments on their platforms, to inaccurate determination of the true vertical at the launch point, to uncertainties in the position and velocity of a movable launching platform (such as a submarine), to inaccurate information on the variation of the gravitational field along the missile's trajectory and, most important, to unexpected modifications of the warhead's trajectory caused by aerodynamic "loading" during reentry—all combine to cause the actual missile path to deviate from its planned trajectory and therefore to miss its target [*see illustration on opposite page*].

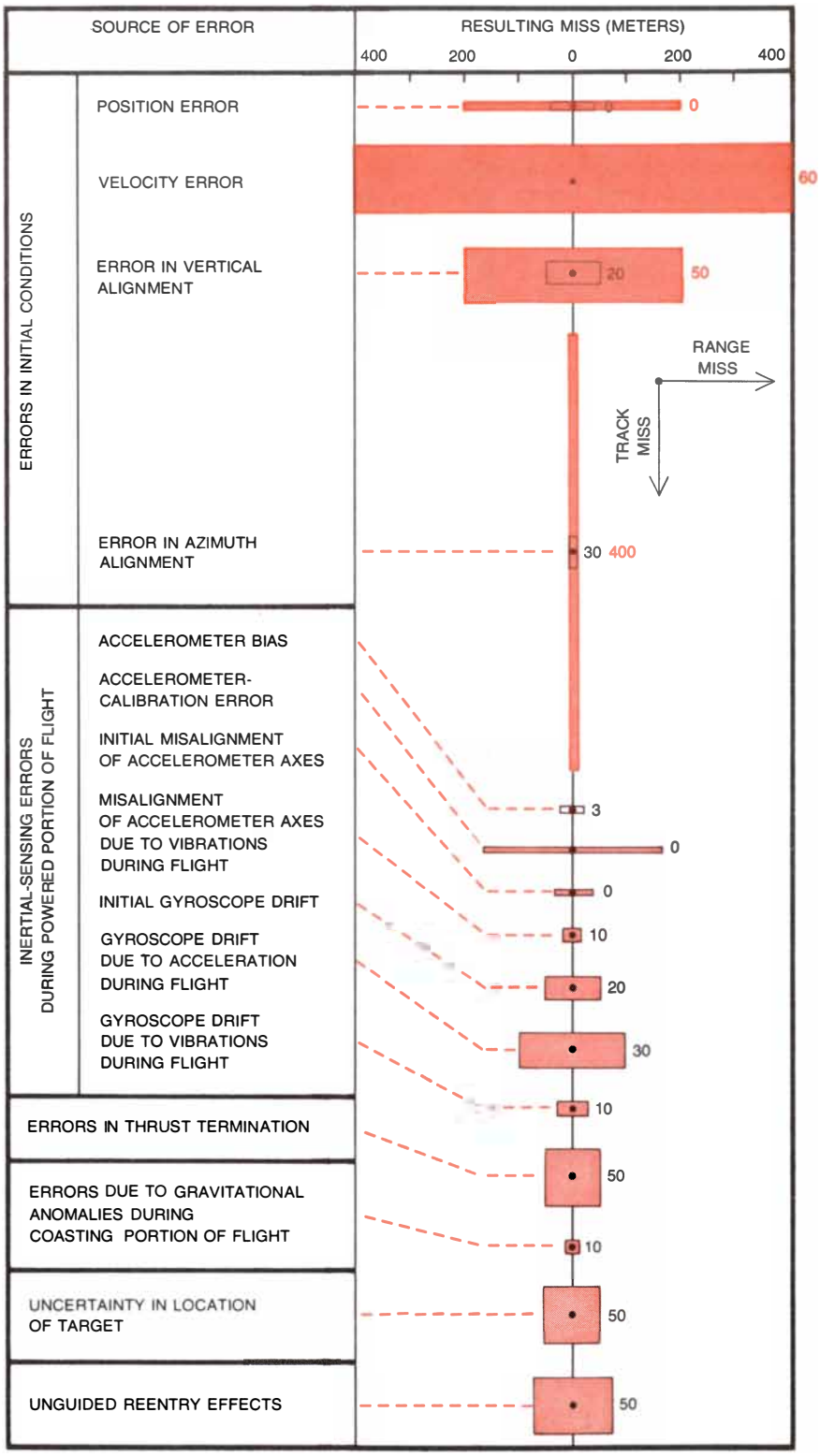
Target miss in the plane of a missile's trajectory is called range miss, and target miss perpendicular to this plane is called track miss [*see illustration on preceding page*]. Both range miss and track miss are calculated by taking the square root of the sum of the squares of individual effects, which are treated as uncorrelated standard deviations. The net miss is the vector sum of the track miss and the range miss; an important quantity, called the "circular error probability," or CEP, is equal to three-fifths of the net miss.

The inertial-guidance systems of the more advanced U.S. strategic ballistic missiles are already capable of exquisite accuracies and stabilities. For example, under certain conditions a dust particle

weighing five hundredths of a microgram (.00000005 gram) on the test mass of a single accelerometer can cause a 200-meter miss in range and a 70-meter miss in track. A shift of five angstroms (.00000005 centimeter) in the center of gravity of a spinning one-kilogram gyroscope wheel from its axis of symmetry can cause a range error of 100 meters and a track error of 50 meters. These instruments must remain "combat ready" for thousands of hours, then be able to tolerate without degradation of their performance the vibrations caused by the burning rocket fuel and still maintain the linearity of their outputs to 10 parts per million over a dynamic range of six orders of magnitude. This means that an accelerometer, for example, should be able to detect accelerations ranging from .01 centimeter per second per second to 10,000 centimeters per second per second and measure them correctly to within 10 parts per million.

A number of other factors can contribute to target miss. Some of them, such as the gravitational pull of the sun and the moon, the influence of the earth's magnetism and the interference of static electric fields, can be estimated and accounted for in establishing the instant of thrust termination; others, such as local weather conditions near the target or the exact location of a submerged submarine during missile launch, are much harder to remedy. As a result the present capabilities of the more advanced U.S. strategic missiles are very close to the limits of accuracy that an inertially guided ballistic missile can achieve without the additional enormous effort and expense necessary to develop a new generation of inertial-sensing instruments. There are, however, alternate methods of improving the accuracy of strategic weapons. Before proceeding to this "futuristic" portion of the article, however, it will be useful to briefly review the numbers and types of ballistic missiles currently deployed by the U.S. and the U.S.S.R. and to examine some of their operational capabilities [*see illustration on page 21*].

The first thing such a tabular representation reveals is that some U.S. missiles each have more than one warhead that can be targeted individually against separate targets. These are the MIRVed missiles (from multiple independently targetable reentry vehicles). A MIRVed missile differs from an ordinary one in that it consists of a rocket on top of which sits a vehicle containing several reentry vehicles. After thrust termination this vehicle (often called a "bus")



SOURCES OF INACCURACY in the targeting of a nuclear warhead delivered by an inertially guided ballistic missile are listed in this illustration along with a graphic representation of the estimated target miss associated with each error source. The dimensions used do not refer to any specific weapon or inertial-guidance system, but they are nonetheless thought to be realistic. The black outlines give the estimated range and track errors for a land-based missile system; the colored areas give the corresponding errors for a submarine-based missile system. The net target miss can be calculated by taking the square root of the sum of the squares of the individual effects, which are treated as uncorrelated standard deviations. The resulting circular error probability for the hypothetical weapon system shown (150 meters for land-based version, 400 meters for submarine-based version) is close to the performance publicly attributed to the more advanced U.S. long-range ballistic missiles.

separates and continues on a free-fall trajectory. It contains a guidance system and small rockets that enable it to change velocity. Somewhere along its ballistic trajectory the guidance system causes the rockets to fire and the bus to change velocity and release a warhead; then the bus changes velocity again and releases another warhead, until all warheads have been released, each on a different trajectory (determined by the velocity of the bus at the instant of release) toward a different target.

The lethality of a nuclear warhead against a silo can be expressed in terms of a parameter designated K, which equals the yield of the warhead in megatons to the two-thirds power divided by the square of the CEP in nautical miles. It follows that this measure of lethality rises much more rapidly with improvements in accuracy than it does with increases in yield [*see illustration below*]. Of course, K does not increase indefinitely with the shrinkage of CEP. Once the accuracy is such that the CEP is smaller than the radius of the crater excavated by the explosive force of the weapon, the lethality K has reached its maximum value, since any silo within such a crater is certainly destroyed. This maximum value of K (which turns out to be independent of the warhead yield) indicates that beyond a certain point increases in accuracy are not meaningful. There is an upper limit to how accurate a reentry vehicle can be, since once the error radius becomes smaller than the kill radius of the weapon any further improvement in accuracy does not increase its lethality.

The K factor can be used to describe both the lethality of a warhead and the lethality needed to destroy a target. Thus one can calculate the value of K required to destroy all the offensive-missile silos in either the U.S. or the U.S.S.R. with a given probability [*see illustration on page 22*]. At the moment it is clear that neither country has the capability of destroying by means of ballistic missiles alone the land-based strategic-missile force of the other country with any reasonable probability. The total $K \times S$ (where S is the number of silos) needed to destroy all the U.S. silos with a probability of 97 percent is more than 82,000, whereas even assuming perfect performance the total $K \times N$ (where N is the total number of warheads) that all the Russian missiles can deliver is about 4,000. Similarly, the total $K \times S$ needed to destroy all the currently deployed Russian silos with the same probability is 40,000, whereas the present U.S. missile force can deliver a maximum total $K \times N$ of about 21,000.

Moreover, the mere fact that the total $K \times N$ of one country may be larger than the total $K \times S$ of an opponent does not necessarily guarantee that the first country can destroy with certainty all the silo-based missiles of the second, because additional parameters such as the reliability of reentry vehicles, the timing of their arrival against a silo, the characteristics of a silo and the type of soil it is in may affect the lethality of a warhead, but in an indirect and less decisive fashion than the warhead's yield and accuracy.

The most effective way to increase the lethality, or K, of a warhead is to diminish the CEP, in other words, to make the missile more accurate. As I have already indicated, however, to reduce the target miss to less than 100 meters or so by striving to improve the inertial guidance is a difficult task. New technological advances, primarily in the fields of "large array" microelectronics and radiation sensors, have opened the way to new approaches to the problem of accurate delivery of a ballistic warhead by radiation-sensing guidance.

Large-array microcircuits are electronic circuits fabricated on very thin substrates that contain in the space of a few square centimeters the equivalent of many thousands of electronic logic circuits. Densities of a million devices per square centimeter of a substrate material 250 microns thick can now be achieved. Allowing for air spaces, one can conceive of practical devices that can store 10^8 bits of serially read-out information per cubic centimeter of electronic memory. The electronics needed for the guidance and control of a missile by radiation sensing consists essentially of logic circuits and memory banks. The new technology has made possible their microminiaturization and therefore the construction of elaborate systems that can be used to guide weapons automatically. Comparable advances have been made in the construction of small, light sensors of electromagnetic radiation that can provide the input signals for radiation-sensing guidance. These devices, which include multimode radars, microwave radiometers, infrared detectors and lasers, consume very small amounts of electric power compared with existing equipment.

Taken as a whole the new developments make possible what essentially is a new set of methods for the accurate

RELATIVE IMPORTANCE of accuracy and explosive yield in determining the lethality of a nuclear warhead designed to be used against a reinforced-concrete missile silo is suggested in this graph. Clearly the ability of the warhead to destroy such a "hardened" target increases much faster with improvements in accuracy than it does with increases in yield.

	MISSILE	EXPLOSIVE YIELD OF WARHEAD (MEGATONS)	C.E.P. OF REENTRY VEHICLE (NAUTICAL MILES)	LETHALITY PER REENTRY VEHICLE (K)	NUMBER OF REENTRY VEHICLES PER MISSILE	TOTAL NUMBER OF MISSILES	TOTAL LETHALITY OF MISSILE FORCE (K × N)
U.S.	MINUTEMAN III	~.16	.2	5	3	550	8,250
	MINUTEMAN II	1	.3	11	1	450	4,950
	TITAN	5	.5	12	1	54	648
	POSEIDON	.05	.3	1.5	10	496	7,440
	POLARIS	.20	.5	1	3	160	480
U.S.S.R.	SS-9	20	(1)	7	1	288	2,016
	SS-11, SS-13	1	(1)	1	1	970	970
	SS-N-6	1	(1–2)	1	1	528	528
	SS-N-8	1	(1–2)	1	1	80	80
	SS-7, SS-8	5	(1.5)	1.3	1	209	270

STRATEGIC-MISSILE FORCES deployed at present by the U.S. and the U.S.S.R. are compared in this table in terms of *K*, a measure of countersilo lethality that takes into account both the accuracy and the explosive yield of each side's nuclear warheads (*N*). Note effects of MIRVing on the total countersilo lethality (*K* × *N*) of the U.S. missiles. The numbers in parentheses are estimates.

delivery of strategic weapons. Of the many approaches made possible by the new technology three have special significance: stellar-guidance systems, terrain-matching systems and satellite-based global-positioning systems.

Stellar guidance is achieved by a device on board the missile that sights a star of known celestial coordinates and then repeatedly measures the angle between the star and the inertial-sensing platform. The expected direction of the star with respect to the platform for perfect guidance can be stored in the memory of the guidance computer. Any deviation from this value, caused by an accumulation of errors during launch, early powered flight and a portion of the coasting trajectory, can be corrected by firing small rockets on board the MIRV bus. If the major component of the anticipated target miss can be identified (in the case of a submarine-launched missile, for example, errors in the position of the launching platform dominate), considerable improvement in accuracy can be achieved by sighting on a star early in the ballistic flight and correcting deviations from the expected values of the angles between the inertial platform and the star. Such a star-fix system is expected to complement the inertial-guidance system of the missiles designed for the new U.S. Trident submarines.

Another method for correcting the accumulated errors of an inertial-guidance system is to use terrain-matching during the warhead's reentry into the atmosphere. Active guidance of a reentering vehicle can achieve essentially zero miss, since the warhead can be maneuvered to hit the target. Movable aerodynamic surfaces or an offset center of gravity can supply deflecting forces that can be commanded by the guidance system to move the warhead up or down or right or left as needed to hit the target. The commands would be generated by a guidance system that could utilize terrain-matching, terminal guidance based on pattern recognition, or some other technique.

The technique of terrain-matching makes use of the fact that the numerical values of certain measurable time-independent terrain variables, such as altitude above sea level or ground reflectivity at a given wavelength, vary as a function of location on the ground. For example, if one divides a map into a grid with squares that correspond to an area on the ground and records in each square on the map the average ground elevation, the result will be the array of numbers that represents the variation of elevation in digital form [*see illustration on page 15*]. The unit cells in the example shown are too large to show details such as man-made structures, but if the elevation of the terrain were determined and recorded with finer resolution, then the array of numbers could show the presence of such structures as houses, water towers, lighthouses and so on. Radar altimeters now in commercial use can resolve objects on the ground as small as three meters across from an altitude of several thousand meters; their vertical resolution can be as good as 30 centimeters from the same altitude. Laser altimeters have even better resolution: from an altitude of 2,000 meters they can provide resolutions smaller than 10 centimeters in the vertical plane and smaller than 20 centimeters in the horizontal plane. The variation of elevation as a function of position over any terrain could be mapped in advance with similar resolutions from an overflying reconnaissance vehicle.

Similar "digital maps" can be compiled that record the reflectivity of the ground at various wavelengths. The earth receives radiation from space and reflects it differentially as a function of the structure or the material at different locations on the surface. A forest will reflect radio waves differently from a road or an airfield, just as velvet reflects visible light differently from a mirror or a piece of paper. By recording the reflectivity of the ground as a function of location one can construct a digital map of the reflectivity of an area that is analogous to a digital map of elevation. Such maps can be easily stored in computer memories on board the reentering warhead as sequences of numbers that are functions of the actual coordinates of the terrain.

A reentry vehicle equipped with an altimeter can determine the variation in altitude of the terrain it overflies. By

HARDNESS OF SILO (POUNDS PER SQUARE INCH)		K REQUIRED TO DESTROY SILO WITH KILL PROBABILITY OF		NUMBER OF SILOS (S)	TOTAL K × S REQUIRED TO DESTROY SILOS WITH KILL PROBABILITY OF	
		97 PERCENT	90 PERCENT		97 PERCENT	90 PERCENT
U.S.	1,000	108	71	550	59,400	39,050
	300	45	30	450	20,250	13,500
	300	45	30	54	2,430	1,620
U.S.S.R.	300	45	30	~400	18,000	12,000
	100	20	13	~1,100	22,000	14,300

TOTAL $K \times S$ REQUIRED to destroy the land-based strategic missiles (S) of each of the nuclear superpowers in a counterforce attack with a kill probability of 97 and 90 percent is estimated in this table. Clearly neither side at present has the capability to destroy, by means of long-range ballistic missiles alone, the land-based strategic-missile force of the other side with any reasonable probability.

comparing this information with a digital map of the terrain stored on board, the guidance system can identify the exact position of the warhead. With these data available it can generate necessary corrections in the direction of motion of the warhead that will place it on a trajectory terminating at the target. Thus terrain-matching during the reentry of a maneuverable warhead can supplement inertial guidance by compensating for all errors of inertial sensing that occurred earlier in the flight and for atmospheric effects. With inertial guidance that results in a miss of a few hundred meters in the vicinity of the target this terrain-matching process can be quite simple, since the location error is quite small and the orientation of the warhead trajectory toward the target is known in advance. A related approach would be to store in the computer of the reentry vehicle the image of the spatial pattern formed by the target against the background. Then a sensing system could be constructed that would "recognize" the target and aim the nuclear warhead at it with pinpoint accuracy.

A combination of these new guidance systems and the older inertial guidance has made possible a new weapons system: the strategic cruise missile. A modestly accurate inertial-guidance system that will probably incorporate strapped-down gyroscopes, supplemented by terrain-matching and pattern-recognition terminal guidance, is designed to guide the first long-range strategic cruise missile now being developed by the U.S. This new weapons system is essentially a small, pilotless, continuously powered, air-breathing vehicle some six meters long and half a meter in diameter (the size of an ordinary torpedo); it is equipped with retractable aerodynamic surfaces and a small turbofan engine, and it is designed to travel subsonically, carrying a nuclear warhead, more than 2,500 kilometers. The weapon is being developed in two versions: an air-launched cruise missile (ALCM) and a submarine-launched cruise missile (SLCM). It could be launched from the torpedo tube of an ordinary submarine or even from a cargo plane flying outside the territorial space of an opponent's country. The guidance package of the weapon will be able to guide it continuously and land it on any fixed target with an error of only a few meters.

A system of satellites that is not subject to surprise hostile action (and is not jammable) can also provide the basis for missile guidance by radiation sensing. Such a system would involve 24 satellites on polar orbits positioned in such a manner that any point on the earth would have at least four of them in sight at all times. The satellites could broadcast coded signals that would be received by passive equipment on board a reentry vehicle or a cruise missile. Signals from four satellites would determine three "time differences of arrival" at the receiver and thereby fix its position accurately in three dimensions with respect to the satellites. If in addition the satellites were to broadcast information defining their orbits, the receiver could determine its position with respect to the earth with similar accuracies. With this position information available, and with velocity information derived from it, the guidance system of the reentry vehicle could maneuver the warhead to land on the target. Local jamming of the receiver near the target would of course be possible, but it could probably be negated by providing the receiver with a directional antenna.

At present it appears feasible to deploy a satellite-based global-positioning system that would enable a missile to determine its position while in flight with an accuracy of between seven and 10 meters anywhere on the surface of the earth independently of the relative positions of launching and target point. Such a system, which could be ready by 1980, would make possible the delivery of warheads over intercontinental ranges with an average error of 10 meters.

It was pointed out earlier in this article that nuclear weapons require a small error radius only if they are intended to be very accurate against hardened missile silos. Nuclear weapons intended for use against soft targets, such as industrial and urban complexes, airports, naval bases and transportation centers, do not need a small error radius in order to be effective. Compared with the present inaccuracy of strategic ballistic missiles, which have an error radius of about .3 kilometer, the efficacy of a nuclear weapon against an airfield five kilometers in extent, say, would not benefit from improved accuracy. Similarly, the lethality against other soft targets, for which the kill radius of the existing warheads already far exceeds the inaccuracy of the missile, would also not increase decisively with improved accuracies. Thus the only putative targets of the new ultra-accurate missiles would be the missile silos of the strategic forces of an adversary. It follows that the application of these new technologies to improve the accuracy of the U.S. strategic missiles will almost certainly be interpreted as reflecting the adoption by the U.S. of a strategic doctrine that anticipates the efficacious destruction of Russian land-based missiles as one of its options.

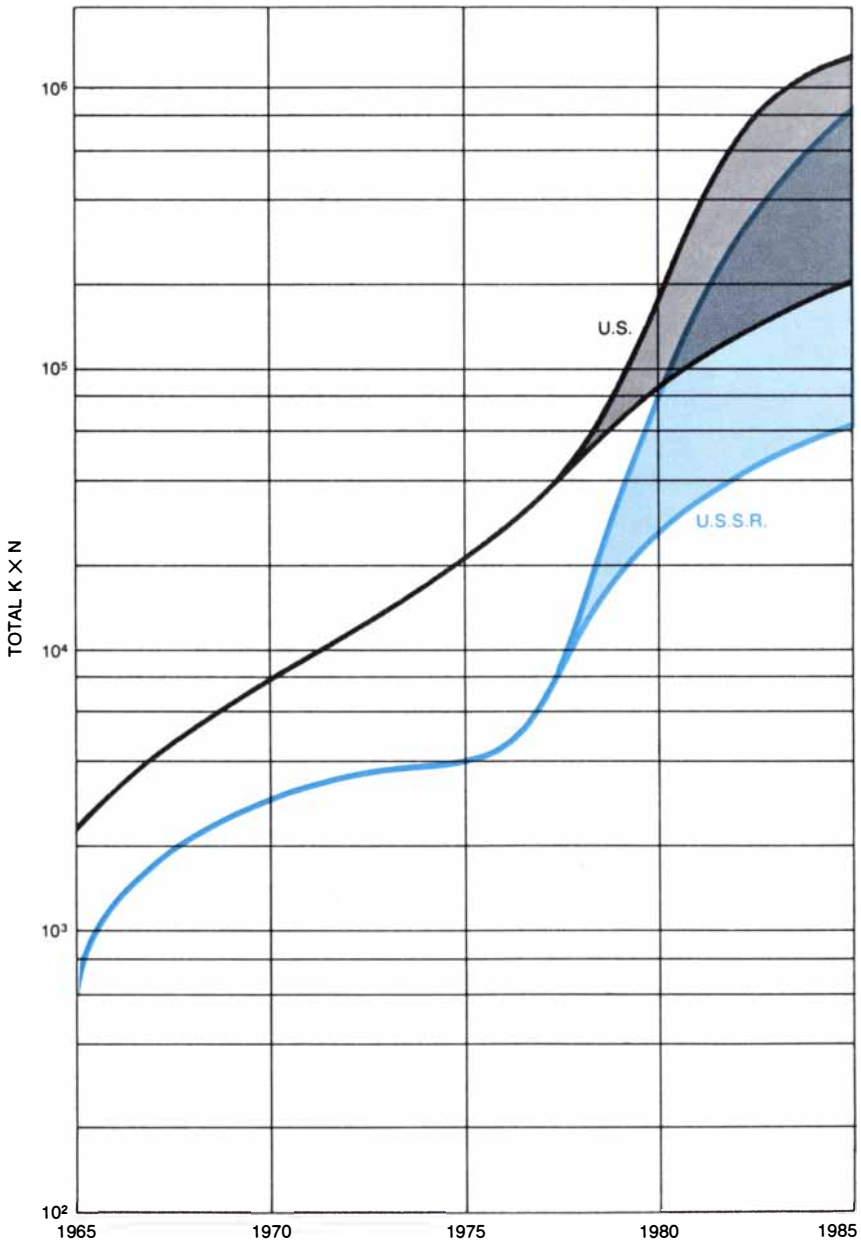
The deployment of U.S. missiles capable of such accuracies will trigger one

of two possible reactions on the part of the Russians. Either the U.S.S.R. will adopt a "launch on warning" policy for its land-based missiles, thereby increasing the strategic instability and the probability of accidental launch of these missiles, or it will resort to mobile land-based strategic missiles. The latter option would of course render U.S. improvements in accuracy futile; at the same time it would further complicate the already complex task of verification by national means of inspection of the number of missiles each country possesses. Since all arms-limitation efforts are currently based on such verification, increased accuracy will probably affect future arms-limitation efforts adversely. In addition, since a country would be reluctant to conclude an arms-limitation agreement while in a tangible position of strategic inferiority, and since increases in the accuracy of U.S. missiles would augment the present asymmetry in performance of the two arsenals to a point incompatible with the requirement of "visible parity" that both countries advocate as being essential for strategic stability and arms control, such emplaced accuracy would further minimize chances for a future negotiated limitation of strategic nuclear weapons.

Of course, the technology that has enabled the U.S. to achieve such improvements in missile accuracy is not the monopoly of this country; it seems likely that the Russians will eventually master it and achieve comparable performance for their strategic weapons. In addition many technological advances that can be employed to improve missile accuracies have important civilian applications. The global-positioning satellite system, for example, will enable commercial ships and planes to determine their position with accuracies of 10 meters, a substantial and useful improvement over present capabilities. It would be both impossible and counterproductive to attempt to control the development of undesirable weapons by advocating the selective arrest of technological progress. Rather one should look to control measures narrowly addressing qualitative improvements of strategic-weapons systems.

In attempting to justify the development of the new ultra-accurate weapons Secretary Schlesinger has said: "Our planning objectives should be to assure that no potential adversary achieves unilateral advantage over the United States." There are at least two flaws in this anticipatory reaction to a hypothetical future threat. First, the action taken will undoubtedly spur an adversary, who must harbor similar feelings about the security of his country, to continue the competition in strategic weapons. Second, the action taken ultimately will lessen the security of one's own land-based missiles. Thus preemptive use of new technologies to improve the accuracy of the U.S. nuclear strategic weapons will not remove the threat against the U.S. strategic arsenal. What *would* protect the strategic forces of both countries would be a freeze of the *quality* of the deployed strategic arsenals.

By not proceeding further with improvements in accuracy both the U.S. and the U.S.S.R. could avoid the inexorable threat to their land-based strategic missiles. To freeze qualitative improvements it suffices to gradually limit missile testing in both countries. The number of missile tests each country can perform every year is both verifiable by national means of inspection and quantitatively negotiable. Such an agreement can assume the form of a treaty, which when ratified will create a permanent climate of strategic stability conducive to better relations between the two countries and to a more stable world order.



ESTIMATED COUNTERFORCE LETHALITY ($K \times N$) of the strategic-missile arsenals of the U.S. and the U.S.S.R. with respect to each other's land-based missile force can be projected on the basis of announced U.S. programs and possible Russian responses. Every U.S. improvement in this area has been followed a few years later by a matching Russian effort.

PLASMID *pSC*101 is shadowed with platinum-palladium and enlarged 230,000 diameters in an electron micrograph made by the author. A plasmid is a molecule of DNA that exists apart from the chromosome in a bacterium and replicates on its own, often carrying the genes for some supplementary activity such as resistance to antibiotics. This plasmid, a small one made by shearing a larger plasmid native to the bacterium *Escherichia coli*, is a circular, or closed-loop, molecule of DNA about three micrometers in circumference that carries the genetic information for replicating itself in *E. coli* and for conferring resistance to the antibiotic tetracycline. It was the "vehicle" for the first gene-manipulation experiments by the author and his colleagues. Foreign DNA was spliced to it and the plasmid was introduced into *E. coli*, where it replicated and expressed both its own and the foreign DNA's genetic information.

THE MANIPULATION OF GENES

Techniques for cleaving DNA and splicing it into a carrier molecule make it possible to transfer genetic information from one organism to an unrelated one. There the DNA replicates and expresses itself

by Stanley N. Cohen

Mythology is full of hybrid creatures such as the Sphinx, the Minotaur and the Chimera, but the real world is not; it is populated by organisms that have been shaped not by the union of characteristics derived from very dissimilar organisms but by evolution within species that retain their basic identity generation after generation. This is because there are natural barriers that normally prevent the exchange of genetic information between unrelated organisms. The barriers are still poorly understood, but they are of fundamental biological importance.

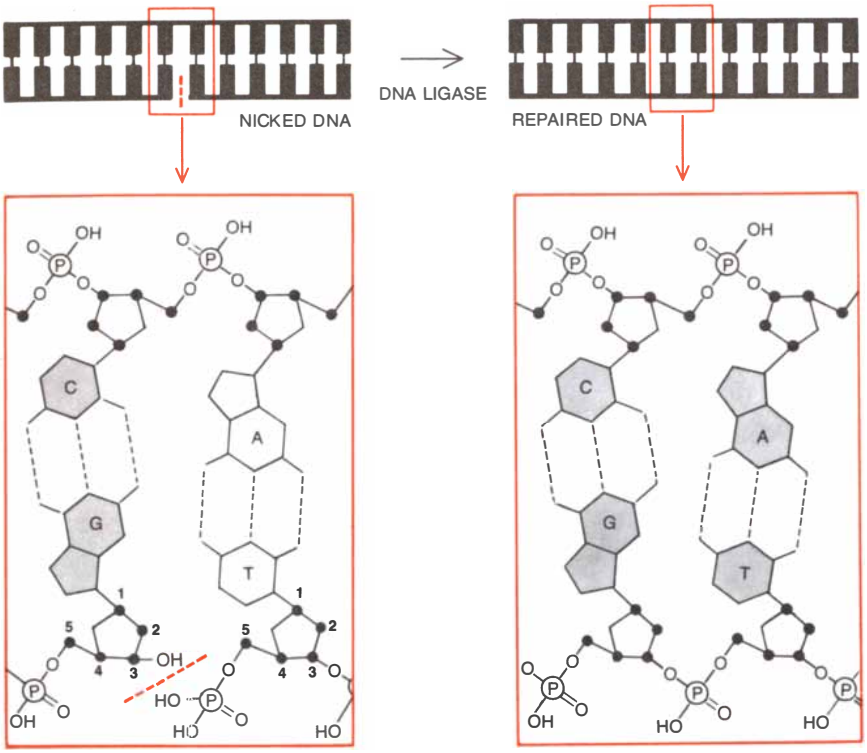
The basic unit of biological relatedness is the species, and in organisms that reproduce sexually species are defined by the ability of their members to breed with one another. Species are determined and defined by the genes they carry, so that in organisms that reproduce asexually the concept of species depends on nature's ability to prevent the biologically significant exchange of genetic material—the nucleic acid DNA—between unrelated groups.

The persistence of genetic uniqueness is perhaps most remarkable in simple organisms such as bacteria. Even when they occupy the same habitat most bacterial species do not exchange genetic information. Even rather similar species of bacteria do not ordinarily exchange the genes on their chromosomes, the structures that carry most of their genetic information. There are exceptions, however. There are bits of DNA, called plasmids, that exist apart from the chromosomes in some bacteria. Sometimes a plasmid can pick up a short segment of DNA from the chromosome of its own cell and transfer it to the cell of a related bacterial species, and sometimes the plasmid and the segment of chromosomal DNA can become integrated into the chromosome of the recipient cell. This transfer of genes between species by extrachromosomal elements has surely played some role in bacterial evolution, but apparently it has not been widespread in nature. Otherwise the characteristics of the common bacterial species would not have remained so largely intact over the huge number of bacterial generations that have existed during the era of modern bacteriology.

In 1973 Annie C. Y. Chang and I at the Stanford University School of Medicine and Herbert W. Boyer and Robert B. Helling at the University of California School of Medicine at San Francisco reported the construction in a test tube of biologically functional DNA molecules that combined genetic information from two different sources. We made the molecules by splicing together segments of two different plasmids found in the colon bacillus *Escherichia coli* and then inserting the composite DNA into *E. coli* cells, where it replicated itself and expressed the genetic information of both parent plasmids. Soon afterward we introduced plasmid genes from an unrelated bacterial species, *Staphylococcus aureus,* into *E. coli,* where they too expressed the biological properties they had displayed in their original host; then, applying the same procedures with John F. Morrow of Stanford and Howard M. Goodman in San Francisco, we were able to insert into *E. coli* some genes from an animal: the toad *Xenopus laevis.*

We called our composite molecules DNA chimeras because they were conceptually similar to the mythological Chimera (a creature with the head of a lion, the body of a goat and the tail of a serpent) and were the molecular counterparts of hybrid plant chimeras produced by agricultural grafting. The procedure we described has since been used and extended by workers in several laboratories. It has been called plasmid engineering, because it utilizes plasmids to introduce the foreign genes, and molecular cloning, because it provides a way to propagate a clone, or line of genetically alike organisms, all containing identical composite DNA molecules. Because of the method's potential for creating a wide variety of novel genetic combinations in microorganisms it is also known as genetic engineering and genetic manipulation. The procedure actually consists of several distinct biochemical and biological manipulations that were made possible by a series of independent discoveries made in rapid succession in the late 1960's and early 1970's. There are four essential elements: a method of breaking and joining DNA molecules derived from different sources; a suitable gene carrier that can replicate both itself and a foreign DNA segment linked to it; a means of introducing the composite DNA molecule, or chimera, into a functional bacterial cell, and a method of selecting from a large population of cells a clone of recipient cells that has acquired the molecular chimera.

In 1967 DNA ligases—enzymes that can repair breaks in DNA and under certain conditions can join together the loose ends of DNA strands—were discovered almost simultaneously in five laboratories. A DNA strand is a chain of nucleotides, each consisting of a deoxyribose sugar ring, a phosphate group and one of four organic bases: adenine, thymine, guanine and cytosine. The sugars and phosphates form the backbone of the strand, from which the bases project. The individual nucleotide building blocks are connected by phosphodiester bonds between the carbon atom at position No. 3 on one sugar and the carbon atom at position No. 5 on the adjacent sugar. Double-strand DNA, the form found in most organisms, consists of two



DNA LIGASE is an enzyme that repairs "nicks," or breaks in one strand of a double-strand molecule of DNA (*top*). A strand of DNA is a chain of nucleotides (*bottom*), each consisting of a deoxyribose sugar and a phosphate group and one of four organic bases: adenine (*A*), thymine (*T*), guanine (*G*) and cytosine (*C*). The sugars and phosphates constitute the backbone of the strand, and paired bases, linked by hydrogen bonds (*broken black lines*), connect two strands. The ligase catalyzes synthesis of a bond at the site of the break (*broken colored line*) between the phosphate of one nucleotide and the sugar of the next nucleotide.

chains of nucleotides linked by hydrogen bonds between their projecting bases. The bases are complementary: adenine (*A*) is always opposite thymine (*T*), and guanine (*G*) is always opposite cytosine (*C*). The function of the ligase is to repair "nicks," or breaks in single DNA strands, by synthesizing a phosphodiester bond between adjoining nucleotides [*see illustration above*].

In 1970 a group working in the laboratory of H. Gobind Khorana, who was then at the University of Wisconsin, found that the ligase produced by the bacterial virus T4 could sometimes catalyze the end-to-end linkage of completely separated double-strand DNA segments. The reaction required that the ends of two segments be able to find each other; such positioning of two DNA molecules was a matter of chance, and so the reaction was inefficient. It was clear that efficient joining of DNA molecules required a mechanism for holding the two DNA ends together so that the ligase could act.

An ingenious way of accomplishing this was developed and tested independently in two laboratories at Stanford: by Peter Lobban and A. Dale Kaiser and by David Jackson, Robert Symons and Paul Berg. Earlier work by others had shown that the ends of the DNA molecules of certain bacterial viruses can be joined by base-pairing between complementary sequences of nucleotides that are naturally present on single-strand segments projecting from the ends of those molecules: *A*'s pair with *T*'s, *G*'s pair with *C*'s and the molecules are held together by hydrogen bonds that form between the pairs. The principle of linking DNA molecules by means of the single-strand projections had been exploited in Khorana's laboratory for joining short synthetic sequences of nucleotides into longer segments of DNA.

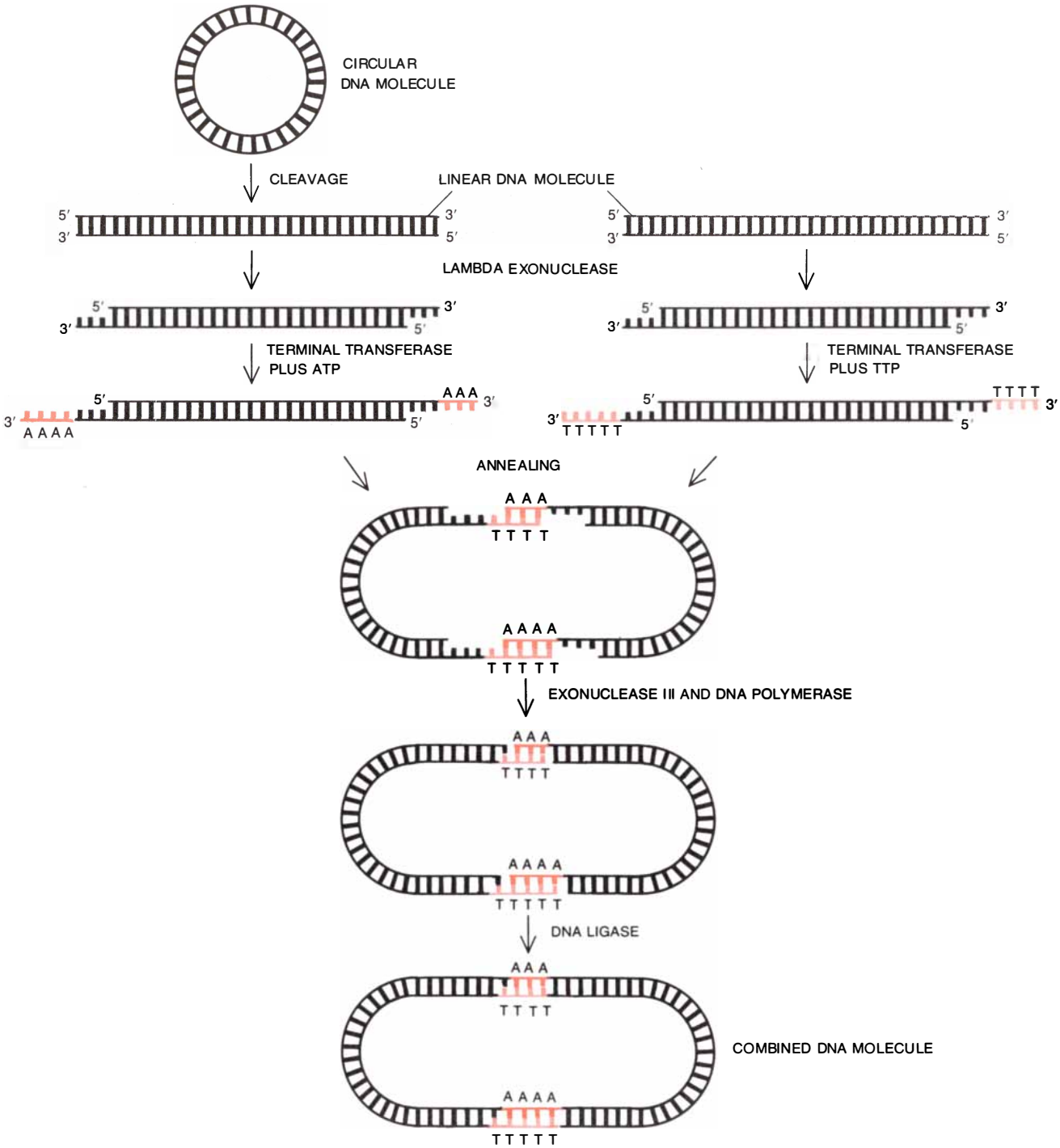
The Stanford groups knew too that an enzyme, terminal transferase, would catalyze the stepwise addition, specifically at what are called the 3′ ends of single strands of DNA, of a series of identical nucleotides. If the enzyme worked also with double-strand DNA, then a block of identical nucleotides could be added to one population of DNA molecules and a block of the complementary nucleotides could be added to another population from another source. Molecules of the two populations could then be annealed by hydrogen bonding and sealed together by DNA ligase. The method was potentially capable of joining any two species of DNA. While Lobban and Kaiser tested the terminal-transferase procedure with the DNA of the bacterial virus P22, Jackson, Symons and Berg applied the procedure to link the DNA of the animal virus SV40 to bacterial-virus DNA.

The SV40 and bacterial-virus DNA molecules Berg's group worked with are closed loops, and the loops had first to be cleaved to provide linear molecules with free ends for further processing and linkage [*see illustration on opposite page*]. (As it happened, the particular enzyme chosen to cleave the loops was the *Eco RI* endonuclease, which was later to be used in a different procedure for making the first biologically functional gene combinations. At the time, however, the enzyme's special property of producing complementary single-strand ends all by itself had not yet been discovered.)

The cleaved linear molecules were treated with an enzyme, produced by the bacterial virus lambda, called an exonuclease because it operates by cutting off nucleotides at the end of a DNA molecule. The lambda exonuclease chewed back the 5′ ends of DNA molecules and thus left projecting single-strand ends that had 3′ termini to which the blocks of complementary nucleotides could be added. The next step was to add, with the help of terminal transferase, a block of *A*'s at the 3′ end of one of the two DNA species to be linked and a block of *T*'s at the 3′ ends of the other species. The species were mixed together. Fragments having complementary blocks at their ends could find each other, line up and become annealed by hydrogen bonding, thus forming combined molecules. To fill the gaps at the 5′ ends of the original segments the investigators supplied nucleotides and two more enzymes: exonuclease III and DNA polymerase. Finally the nicks in the molecules were sealed with DNA ligase.

The method of making cohesive termini for joining DNA molecules in the first successful genetic-manipulation experiments was conceptually and operationally different from the terminal-transferase procedure. It was also much simpler. It depended on the ability of one of a group of enzymes called restriction endonucleases to make complementary-ended fragments during the cleavage of DNA at a site within the molecule, instead of requiring the addition of new blocks of complementary nucleotides to DNA termini.

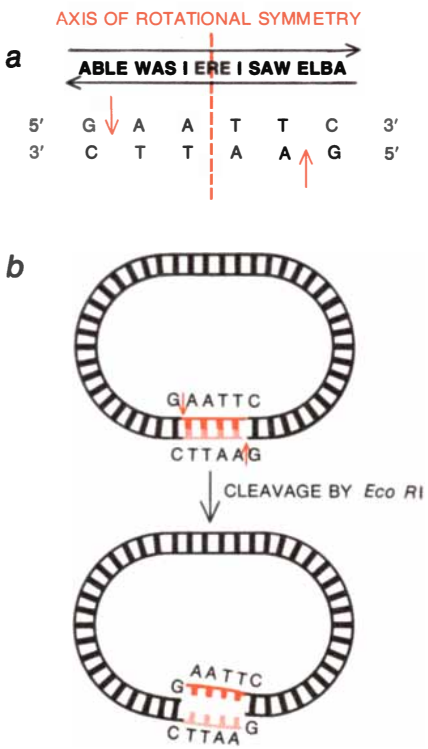
Viruses grown on certain strains of *E. coli* were known to be restricted in their ability to grow subsequently on other strains. Investigations had shown that this restriction was due to bacterial enzymes that recognize specific sites on a "foreign" viral DNA and cleave that DNA. (To protect its own DNA the bacterial cell makes a modification enzyme that adds methyl groups to nucleotides constituting the recognition sites for the restriction endonuclease, making them resistant to cleavage.) Restriction endonucleases (and modification methylases) are widespread in microorganisms; genes for making them were found on viral chromosomes and extrachromosomal plasmid DNA as well as on many bacterial chromosomes. During the early 1970's the nucleotide sequences at the cleavage sites recognized by several re-



TERMINAL-TRANSFERASE procedure for joining DNA molecules involves a number of steps, each dependent on a different enzyme. If one of the molecules to be joined is a closed loop, it must first be cleaved. The linear molecules are treated with lambda exonuclease, an enzyme that cuts nucleotides off the 5′ end of DNA strands (the end with a phosphate group on the No. 5 carbon). Then specific nucleotides are added to the 3′ end (the end with an OH group on the No. 3 carbon) by the action of the enzyme terminal transferase. One DNA species is supplied with adenosine triphosphate (ATP), the other with thymidine triphosphate (TTP), so that *A* nucleotides are added to one species and complementary *T* nucleotides to the other. When the two species are mixed, the complementary bases pair up, annealing the molecules. Nucleotides and the enzymes DNA polymerase and exonuclease III are added to fill gaps and DNA ligase is added to seal the DNA backbones. The result is a double molecule composed of two separate DNA segments.

striction endonucleases were identified. In every instance, it developed, the cleavage was at or near an axis of rotational symmetry: a palindrome where the nucleotide base sequences read the same on both strands in the 5′-to-3′ direction [*see illustration below*].

In some instances the breaks in the DNA strands made by restriction enzymes were opposite each other. One particular endonuclease, however, the *Eco RI* enzyme isolated by Robert N. Yoshimori in Boyer's laboratory in San Francisco, had a property that was of special interest. Unlike the other nucleases known at the time, this enzyme introduced breaks in the two DNA strands that were separated by several nucleotides. Because of the symmetrical, palindromic arrangement of the nucleotides in the region of cleavage this separation of the cleavage points on the two strands yielded DNA termini with projecting complementary nucleotide sequences: "sticky" mortise-and-tenon ter-



RESTRICTION ENDONUCLEASES cleave DNA at sites where complementary nucleotides are arranged in rotational symmetry: a palindrome, comparable to a word palindrome (*a*). The endonuclease *Eco RI* has the additional property of cleaving complementary strands of DNA at sites (*colored arrows*) four nucleotides apart. Such cleavage (*b*) yields DNA fragments with complementary, overlapping single-strand ends. As a result the end of any DNA fragment produced by *Eco RI* cleavage can anneal with any other fragment produced by the enzyme.

mini. The *Eco RI* enzyme thus produced in one step DNA molecules that were functionally equivalent to the cohesive-end molecules produced by the complicated terminal-transferase procedure.

The experiments that led to the discovery of the capabilities of *Eco RI* were reported independently and simultaneously in November, 1972, by Janet Mertz and Ronald W. Davis of Stanford and by another Stanford investigator, Vittorio Sgaramella. Sgaramella found that molecules of the bacterial virus P22 could be cleaved with *Eco RI* and would then link up end to end to form DNA segments equal in length to two or more viral-DNA molecules. Mertz and Davis observed that closed-loop SV40-DNA molecules cleaved by *Eco RI* would reform themselves into circular molecules by hydrogen bonding and could be sealed with DNA ligase; the reconstituted molecules were infectious in animal cells growing in tissue culture. Boyer and his colleagues analyzed the nucleotide sequences at the DNA termini produced by *Eco RI*, and their evidence confirmed the complementary nature of the termini, which accounted for their cohesive activity.

In late 1972, then, several methods were available by which one could join double-strand molecules of DNA. That was a major step in the development of a system for manipulating genes. More was necessary, however. Most segments of DNA do not have an inherent capacity for self-replication; in order to reproduce themselves in a biological system they need to be integrated into DNA molecules that can replicate in the particular system. Even a DNA segment that can replicate in its original host was not likely to have the specific genetic signals required for replication in a different environment. If foreign DNA was to be propagated in bacteria, as had long been proposed in speculative scenarios of genetic engineering, a suitable vehicle, or carrier, was required. A composite DNA molecule consisting of the vehicle and the desired foreign DNA would have to be introduced into a population of functional host bacteria. Finally, it would be necessary to select, or identify, those cells in the bacterial population that took up the DNA chimeras. In 1972 it still seemed possible that the genetic information on totally foreign DNA molecules might produce an aberrant situation that would prevent the propagation of hybrid molecules in a new host.

Molecular biologists had focused for many years on viruses and their relations with bacteria, and so it was natural that bacterial viruses were thought of as the most likely vehicles for genetic manipulation. For some time there had been speculation and discussion about using viruses, such as lambda, that occasionally acquire bits of the *E. coli* chromosome by natural recombination mechanisms for cloning DNA from foreign sources. It was not a virus, however, but a plasmid that first served as a vehicle for introducing foreign genes into a bacterium and that provided a mechanism for the replication and selection of the foreign DNA.

A ubiquitous group of plasmids that confer on their host bacteria the ability to resist a number of antibiotics had been studied intensively for more than a decade. Antibiotic-resistant *E. coli* isolated in many parts of the world, for example, were found to contain plasmids, designated *R* factors (for "resistance"), carrying the genetic information for products that in one way or another could interfere with the action of specific antibiotics [see "Infectious Drug Resistance," by Tsutomu Watanabe; SCIENTIFIC AMERICAN, December, 1967]. Double-strand circular molecules of *R*-factor DNA had been separated from bacterial chromosomal DNA by centrifugation in density gradients and had been characterized by biochemical and physical techniques [see "The Molecule of Infectious Drug Resistance," by Royston C. Clowes; SCIENTIFIC AMERICAN, April, 1973].

In 1970 Morton Mandel and A. Higa of the University of Hawaii School of Medicine had discovered that treatment of *E. coli* with calcium salts enabled the bacteria to take up viral DNA. At Stanford, Chang and I, with Leslie Hsu, found that if we made the cell membranes of *E. coli* permeable by treating them with calcium chloride, purified *R*-factor DNA could be introduced into them [*see illustration on opposite page*]. The *R*-factor DNA is taken up in this transformation process by only about one bacterial cell in a million, but those few cells can be selected because they live and multiply in the presence of the antibiotics to which the *R* factor confers resistance, whereas other cells die. Each transformed cell gives rise to a clone that contains exact replicas of the parent plasmid DNA molecules, and so we reasoned that plasmids might serve as vehicles for propagating new genetic information in a line of *E. coli* cells.

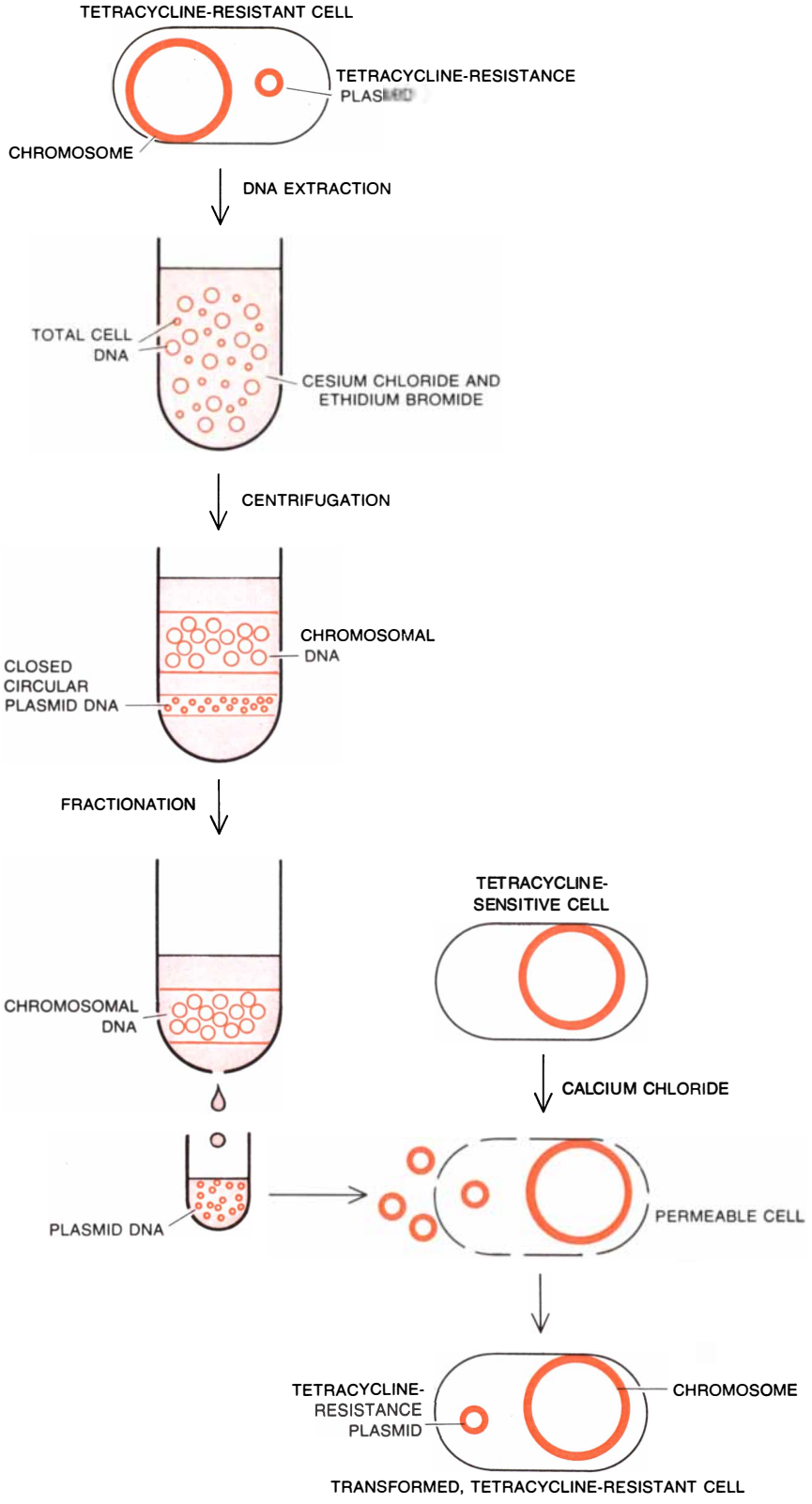
In an effort to explore the genetic and molecular properties of various regions of the *R*-factor DNA we had begun to take plasmids apart by shearing their DNA mechanically and then transforming *E. coli* with the resulting

fragments. Soon afterward we began to cleave the plasmids with the *Eco RI* enzyme, which had been shown to produce multiple site-specific breaks in several viruses. It might therefore be counted on to cleave all molecules of a bacterial plasmid in the same way, so that any particular species of DNA would yield a specific set of cleavage fragments, and do so reproducibly. The fragments could then be separated and identified according to the different rates at which they would migrate through a gel under the influence of an electric current.

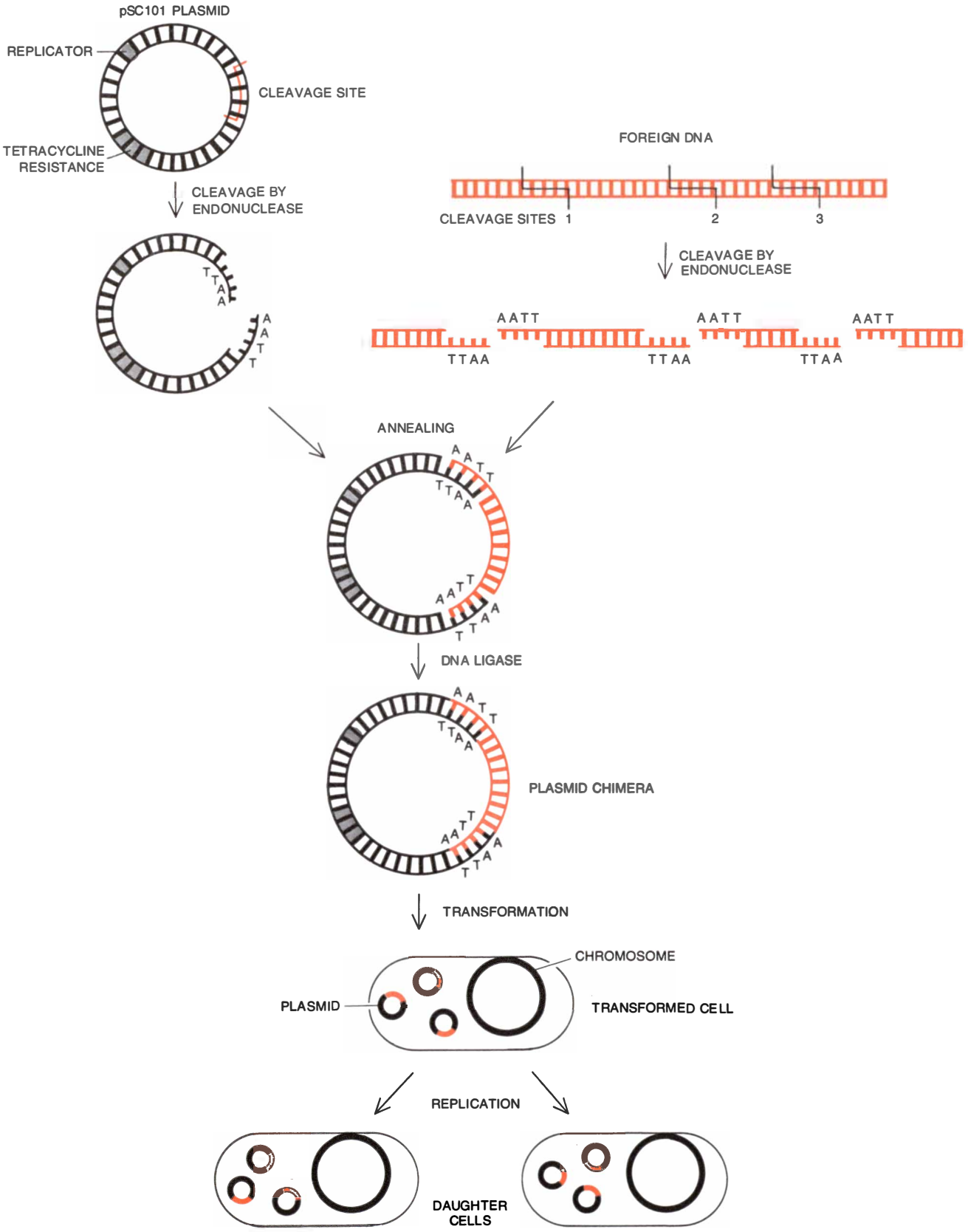
When the DNA termini produced by *Eco RI* endonuclease were found to be cohesive, Chang and I, in collaboration with Boyer and Helling in San Francisco, proceeded to search for a plasmid that the enzyme would cleave without affecting the plasmid's ability to replicate or to confer antibiotic resistance. We hoped that if such a plasmid could be found, we could insert a segment of foreign DNA at the *Eco RI* cleavage site, and that it might be possible to propagate the foreign DNA in *E. coli*.

In our collection at Stanford there was a small plasmid, *pSC*101, that had been isolated following the mechanical shearing of a large plasmid bearing genes for multiple antibiotic resistance. It was less than a twelfth as long as the parent plasmid, but it did retain the genetic information for its replication in *E. coli* and for conferring resistance to one antibiotic, tetracycline. When we subjected *pSC*101 DNA to cleavage by *Eco RI* and analyzed the products by gel electrophoresis, we found that the enzyme had cut the plasmid molecule in only one place, producing a single linear fragment. We were able to join the ends of that fragment again by hydrogen bonding and reseal them with DNA ligase, and when we introduced the reconstituted circular DNA molecules into *E. coli* by transformation, they were biologically functional plasmids: they replicated and conferred tetracycline resistance.

The next step was to see if a fragment of foreign DNA could be inserted at the cleavage site without interfering with replication or expression of tetracycline resistance and thus destroying the plasmid's ability to serve as a cloning vehicle. We mixed the DNA of another *E. coli* plasmid, which carried resistance to the antibiotic kanamycin, with the *pSC*101 DNA. We subjected the mixed DNA to cleavage by *Eco RI* and then to ligation, transformed *E. coli* with the resulting DNA and found that some of the transformed bacteria were indeed resist-



PLASMID DNA can be introduced into a bacterial cell by the procedure called transformation. Plasmids carrying genes for resistance to the antibiotic tetracycline (*top left*) are separated from bacterial chromosomal DNA. Because differential binding of ethidium bromide by the two DNA species makes the circular plasmid DNA denser than the chromosomal DNA, the plasmids form a distinct band on centrifugation in a cesium chloride gradient and can be separated (*bottom left*). The plasmid DNA is mixed with bacterial cells that are not resistant to tetracycline and that have been made permeable by treatment with a calcium salt. The DNA enters the cells, replicates there and makes the cells resistant to tetracycline.



FOREIGN DNA is spliced into the *pSC*101 plasmid and introduced with the plasmid into the bacterium *Escherichia coli*. The plasmid is cleaved by the endonuclease *Eco* RI at a single site that does not interfere with the plasmid's genes for replication or for resistance to tetracycline (*top left*). The nucleotide sequence recognized by *Eco* RI is present also in other DNA, so that a foreign DNA exposed to the endonuclease is cleaved about once in every 4,000 to 16,000 nucleotide pairs on a random basis (*top right*). Fragments of cleaved foreign DNA are annealed to the plasmid DNA by hydrogen bonding of the complementary base pairs, and the new composite molecules are sealed by DNA ligase. The DNA chimeras, each consisting of the entire plasmid and a foreign DNA fragment, are introduced into *E. coli* by transformation, and the foreign DNA is replicated by virtue of the replication functions of the plasmid.

ant to both tetracycline and kanamycin. The plasmids isolated from such transformants contained the entire *pSC*101 DNA segment and also a second DNA fragment that carried the information for kanamycin resistance, although it lacked replication functions of its own. The results meant that the *pSC*101 could serve as a cloning vehicle for introducing at least a nonreplicating segment of a related DNA into *E. coli.* And the procedure was extraordinarily simple.

Could genes from other species be introduced into *E. coli* plasmids, however? There might be genetic signals on foreign DNA that would prevent its propagation or expression in *E. coli.* We decided to try to combine DNA from a plasmid of another bacterium, the *pI*258 plasmid of *Staphylococcus aureus,* with our original *E. coli* plasmid. The staphylococcal plasmid had already been studied in several laboratories; we had found that it was cleaved into four DNA fragments by *Eco R*I. Since *pI*258 was not native to *E. coli* or to related bacteria, it could not on its own propagate in an *E. coli* host. And it was known to carry a gene for resistance to still another antibiotic, penicillin, that would serve as a marker for selecting any transformed clones. (Penicillin resistance, like combined resistance to tetracycline and kanamycin, was already widespread among *E. coli* strains in nature. That was important; if genes from a bacterial species that cannot normally exchange genetic information with the colon bacillus were to be introduced into it, it was essential that they carry only antibiotic-resistance traits that were already prevalent in *E. coli.* Otherwise we would be extending the species' antibiotic-resistance capabilities.)

Chang and I repeated the experiment that had been successful with two kinds of *E. coli* plasmids, but this time we did it with a mixture of the *E. coli*'s *pSC*-101 and the staphylococcal *pI*258: we cleaved the mixed plasmids with *Eco R*I endonuclease, treated them with ligase and then transformed *E. coli.* Next we isolated transformed bacteria that expressed the penicillin resistance coded for by the S. *aureus* plasmid as well as the tetracycline resistance of the *E. coli* plasmid. These doubly resistant cells were found to contain a new DNA species that had the molecular characteristics of the staphylococcal plasmid DNA as well as the characteristics of *pSC*101.

The replication and expression in *E. coli* of genes derived from an organism ordinarily quite unable to exchange genes with *E. coli* represented a breach in the barriers that normally separate biological species. The bulk of the genetic information expressed in the transformed bacteria defined it as *E. coli,* but the transformed cells also carried replicating DNA molecules that had molecular and biological characteristics derived from an unrelated species, *S. aureus.* The fact that the foreign genes were on a plasmid meant that they would be easy to isolate and purify in large quantities for further study. Moreover, there was a possibility that one might introduce genes into the easy-to-grow *E. coli* that specify a wide variety of metabolic or synthesizing functions (such as photosynthesis or antibiotic production) and that are indigenous to other biological classes. Potentially the *pSC*101 plasmid and the molecular-cloning procedure could serve to introduce DNA molecules from complex higher organisms into bacterial hosts, making it possible to apply relatively simple bacterial genetic and biochemical techniques to the study of animal-cell genes.

Could animal-cell genes in fact be introduced into bacteria, and would they replicate there? Boyer, Chang, Helling and I, together with Morrow and Goodman, immediately undertook to find out. We picked certain genes that had been well studied and characterized and were available, purified, in quantity: the genes that code for a precursor of the ribosomes (the structure on which proteins are synthesized) in the toad *Xenopus laevis.* The genes had properties that would enable us to identify them if we succeeded in getting them to propagate in bacteria. The toad DNA was suitable for another reason: although we would be constructing a novel biological combination containing genes from both animal cells and bacteria, we and others expected that no hazard would result from transplanting the highly purified ribosomal genes of a toad.

Unlike the foreign DNA's of our earlier experiments, the toad genes did not express traits (such as antibiotic resistance) that could help us to select bacteria carrying plasmid chimeras. The tetracycline resistance conferred by *pSC*101 would make it possible to select transformed clones, however, and we could then proceed to examine the DNA isolated from such clones to see if any clones contained a foreign DNA having the molecular properties of toad ribosomal DNA. The endonuclease-generated fragments of toad ribosomal DNA have characteristic sizes and base compositions; DNA from the transformed cells could be tested for those characteristics. The genes propagated in bacteria could also be tested for nucleotide-sequence homology with DNA isolated directly from the toad.

When we did the experiment and analyzed the resulting transformed cells, we found that the animal-cell genes were indeed reproducing themselves in generation after generation of bacteria by means of the plasmid's replication functions. In addition, the nucleotide sequences of the toad DNA were being transcribed into an RNA product in the bacterial cells.

Within a very few months after the first DNA-cloning experiments the procedure was being used in a number of laboratories to clone bacterial and animal-cell DNA from a variety of sources. Soon two plasmids other than *pSC*101 were discovered that have a single *Eco RI* cleavage site at a location that does not interfere with essential genes. One of these plasmids is present in many copies in the bacterial cell, making it possible to "amplify," or multiply many times, any DNA fragments linked to it. Investigators at the University of Edinburgh and at Stanford went on to develop mutants of the virus lambda (which ordinarily infects *E. coli*) that made the virus too an effective cloning vehicle. Other restriction endonucleases were discovered that also make cohesive termini but that cleave DNA at different sites from the *Eco R*I enzymes, so that chromosomes can now be taken apart and put together in various ways.

The investigative possibilities of DNA cloning are already being explored intensively. Some workers have isolated from complex chromosomes certain regions that are implicated in particular functions such as replication. Others are making plasmids to order with specific properties that should clarify aspects of extrachromosomal-DNA biology that have been hard to study. The organization of complex chromosomes, such as those of the fruit fly *Drosophila,* is being studied by cloning the animal genes in bacteria. Within the past few months methods have been developed for selectively cloning specific genes of higher organisms through the use of radioactively labeled RNA probes: instead of purifying the genes to be studied before introducing them into bacteria, one can transform bacteria with a heterogeneous population of animal-cell DNA and then isolate those genes that produce a particular species of RNA. It is also possible to isolate groups of genes that are expressed concurrently at a particular stage in the animal's development.

The potential seems to be even broader. Gene manipulation opens the pros-

pect of constructing bacterial cells, which can be grown easily and inexpensively, that will synthesize a variety of biologically produced substances such as antibiotics and hormones, or enzymes that can convert sunlight directly into food substances or usable energy. Perhaps it even provides an experimental basis for introducing new genetic information into plant or animal cells.

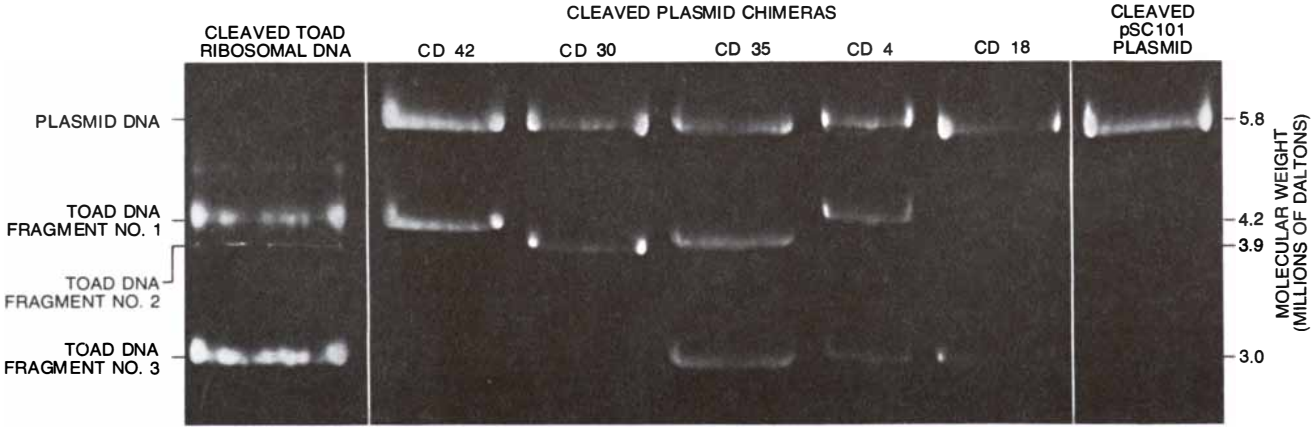
It has been clear from the beginning of experimentation in molecular cloning that the construction of some kinds of novel gene combinations may have a potential for biological hazard, and the scientific community has moved quickly to make certain that research in genetic manipulation would not endanger the public. For a time after our initial experiments the *pSC*101 plasmid was the only vehicle known to be suitable for cloning foreign DNA in *E. coli,* and our colleagues asked for supplies with which to pursue studies we knew were of major scientific and medical importance. Investigators normally facilitate the free exchange of bacteria and other experimental strains they have isolated or developed, but Chang and I were concerned that manipulation of certain genes could give rise to novel organisms whose infectious properties and ecological effects could not be predicted. In agreeing to provide the plasmid we therefore asked for assurance that our colleagues would neither introduce tumor viruses into bacteria nor create antibiotic-resistance combinations that were not already present in nature; we also asked the recipients not to send the plasmid on to other laboratories, so that we could keep track of its distribution.

When still other cloning vehicles were discovered, it became apparent that a more general mechanism for ensuring experimental safety in gene-manipulation research was advisable. The groundwork for such control had been established earlier: the National Academy of Sciences had been urged to consider the "possibility that potentially biohazardous consequences might result from widespread or injudicious use" of these techniques and had asked Paul Berg to form an advisory committee that would consider the issue. Berg too had been concerned about the potential hazards of certain kinds of experimentation for some years, and had himself decided to abandon plans to try to introduce genes from the tumor virus SV40 into bacteria because of the possible danger if the experiment were successful.

Berg brought together a number of investigators, including some who were then directly involved in molecular cloning, in the spring of 1974. In a report released in July and in a letter to leading professional journals the members of the committee expressed their "concern about the possible unfortunate consequences of indiscriminate application" of the techniques and formally asked all investigators to join them in voluntarily deferring two types of experiments (which had, as a matter of fact, been avoided by informal consensus up until that time). Experiments of Type I involved the construction of novel organisms containing combinations of toxin-producing capabilities or of antibiotic-resistance genes not found in nature. Type 2 experiments involved the introduction of DNA from tumor viruses or other animal viruses into bacteria; the committee noted that "such recombinant molecules might be more easily disseminated to bacterial populations in humans and other species, and might thus increase the incidence of cancer or other diseases."

The Academy committee was concerned largely because of our inability to assess the hazards of certain experiments accurately before the experiments were undertaken. Guidelines for safety had long been available in other areas of potentially hazardous research, such as studies involving known disease-causing bacteria and viruses, radioactive isotopes or toxic chemicals. Because of the newness of the microbial gene-manipulation methods, no such guidelines had yet been developed for work in this area, however; there was the possibility that potentially hazardous experiments might proceed before appropriate guidelines could be considered and implemented. We recognized that most work with the new methods did not and would not involve experiments of a hazardous nature but we recommended the deferral of Type I and Type II experiments until the hazards were more carefully assessed, until it was determined whether or not the work could be undertaken safely and until adequate safety precautions were available. The committee also proposed that an international meeting be held early in 1975 to consider the matter more fully.

Such a meeting was held in February at the Asilomar Conference Center near Pacific Grove, Calif. It brought together 86 American biologists and 53 investigators from 16 other countries, who spent three and a half days reviewing progress in the field of molecular cloning and formulating guidelines that would allow most types of new hereditary characteristics to be introduced into bacteria and

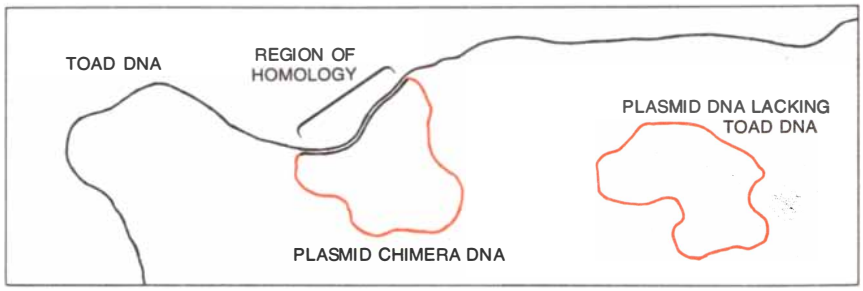


GEL ELECTROPHORESIS demonstrates the presence of toad DNA in chimeric plasmids. Fragments of DNA migrate through a gel at different rates under the influence of an electric current, depending on their size. Linear molecules of plasmid DNA (*right*) and the cleavage products of toad ribosomal DNA (*left*) therefore have characteristic sizes and migrate characteristic distances in a given time. The bands of DNA, visualized by a fluorescent dye, are photographed in ultraviolet. All five chimeric plasmids (*center*) contain a plasmid DNA molecule; in addition each chimera includes one or more fragments characteristic of original toad DNA.

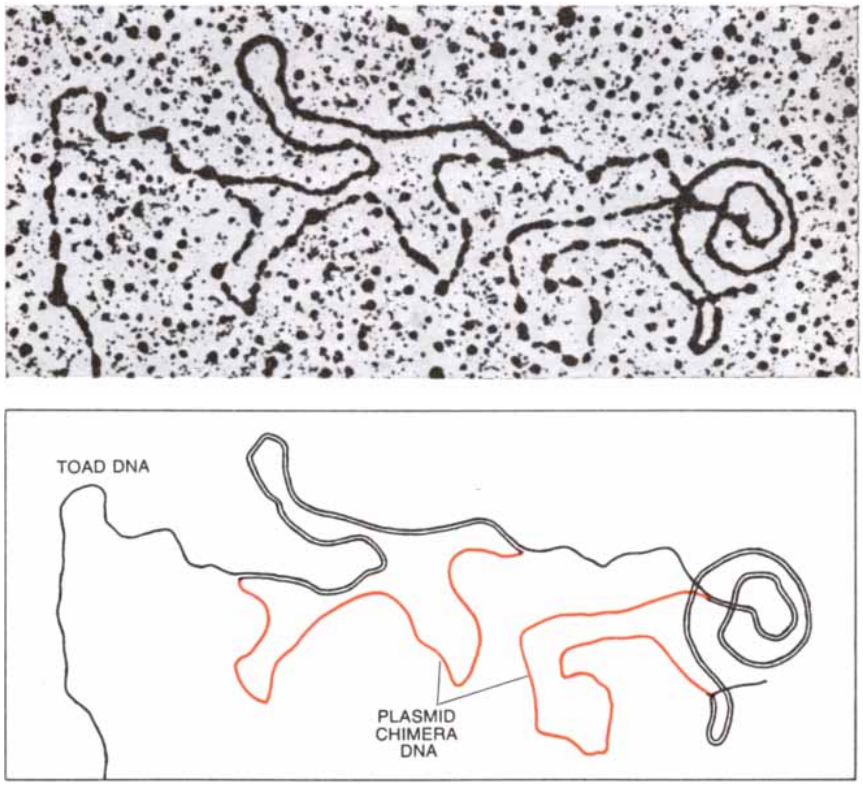
viruses safely. Invited nonscientists from the fields of law and ethics participated in the discussions and decisions at Asilomar, along with representatives of agencies that provide Federal funds for scientific research; the meetings were open to the press and were fully reported. The issues were complex and there were wide differences of opinion on many of them, but there was consensus on three major points. First, the newly developed cloning methods offer the prospect of dealing with a wide variety of important scientific and medical problems as well as other problems that trouble society, such as environmental pollution and food and energy shortages. Second, the accidental dissemination of certain novel biological combinations may present varying degrees of potential risk. The construction of such combinations should proceed only under a graded series of precautions, principally biological and physical barriers, adequate to prevent the escape of any hazardous organisms; the extent of the actual risk should be explored by experiments conducted under strict containment conditions. Third, some experiments are potentially too hazardous to be carried out for the present, even with the most careful containment. Future research and experience may show that many of the potential hazards considered at the meeting are less serious and less probable than we now suspect. Nevertheless, it was agreed that standards of protection should be high at the beginning and that they can be modified later if the assessment of risk changes.

Physical containment barriers have long been used in the U.S. space-exploration program to minimize the possibility of contamination of the earth by extraterrestrial microbes. Containment procedures are also employed routinely to protect laboratory workers and the public from hazards associated with radioactive isotopes and toxic chemicals and in work with disease-causing bacteria and viruses. The Asilomar meeting formulated the additional concept of biological barriers, which involve fastidious cloning vehicles that are able to propagate only in specialized hosts and equally fastidious bacterial strains that are unable to live except under stringent laboratory conditions.

In the past the scientific community has commonly policed its own actions informally, responding to ethical concerns with self-imposed restraint. Usually, but not always, society at large has also considered the public well-being in determining how knowledge obtained by basic scientific research should be applied. Extensive public scrutiny and open discussion by scientists and nonscientists of the possible risks and benefits of a particular line of basic research has been rare, however, when (as in this case) the hazards in question are only potential and, for some experiments, even hypothetical. As this article is being written it is still too early to know what the long-range outcome of the public discussions initiated by scientists working in genetic manipulation will be. One can hope that the forthright approach and the rigorous standards that have been adopted for research in the cloning of recombinant DNA molecules will promote a sharper focus on other issues relevant to public and environmental safety.



HETERODUPLEX ANALYSIS identifies regions of a toad DNA (*black*) that have been incorporated in a chimeric plasmid DNA molecule. DNA isolated from toad eggs and the DNA of the chimera are denatured, that is, each natural double-strand molecule is split into two single strands of DNA, by alkali treatment. The toad and the chimeric DNA's are mixed together, and any complementary sequences are allowed to find each other. The toad DNA incorporated in the chimeras has nucleotide sequences that are complementary to sequences in the DNA taken directly from the animal source. Those homologous sequences anneal to form heteroduplex double-strand DNA that can be identified in electron micrographs.



PRESENCE OF TOAD DNA in two separate chimeric plasmid molecules is demonstrated by an electron micrograph made by John F. Morrow at the Stanford University School of Medicine. As is indicated in the drawing (*bottom*), there are DNA strands from two plasmids and a strand of toad DNA. The micrograph shows thickened regions of DNA where nucleotide sequences are homologous and two single strands have been annealed. The toad DNA in the chimeras codes for ribosomes, and the space between the two heteroduplex regions is compatible with the spacing of multiple ribosomal genes in toad DNA.

Positrons as a Probe of the Solid State

When positrons are injected into a crystalline substance, they are annihilated in encounters with electrons. The resulting gamma rays yield unique information on the atomic structure of the substance

by Werner Brandt

The positron is the antimatter particle of the electron. The two particles have identical properties except that they have opposite electric charges. When a positron encounters an electron, the particles are annihilated, and their mass is converted into pure energy in the form of gamma rays. The process is a direct demonstration of several fundamental conservation laws of modern physics. Certain characteristics of the process, however, can be influenced if the particles meet in an atomic environment. As a consequence positrons can be used as a probe for exploring the nature of matter. It turns out that the probe is unusually sensitive and revealing when it is applied to the regular arrays of atoms characteristic of most solids. It is also beginning to have its uses in the study of living matter.

When a positron and an electron annihilate each other in condensed matter (a solid or a liquid), they always give rise to two gamma ray photons. The conversion of the particles' mass into energy exactly follows Einstein's equation $E = mc^2$, in which E is the energy liberated, m is the mass of the particles and c is the speed of light; mass and energy are thus conserved. The sum of the positron's positive charge (+1) and the electron's negative charge (−1) is zero. The gamma ray photons that result from the annihilation carry no charge; charge is thus conserved. In the annihilation events that are of interest here the spins of the particles are antiparallel and add up to zero. The gamma ray photons have no spin; spin is thus conserved. The two photons each have an energy of .511 million electron volts (MeV), and they leave the site of the annihilation in exactly opposite directions. Their net momentum is zero; momentum is thus conserved. The annihilation process therefore conserves energy, charge, spin and momentum. All the same, electrons could annihilate the protons in the nuclei of atoms without violating these laws, and if they did so, there would be no atoms. Protons, however, are 2,000 times as massive as electrons. Electrons can be annihilated only by antielectrons, that is, positrons. The fact that our world exists therefore proves yet another universal law, a law that might be called the conservation of light and heavy particles. In short, the positron-electron system exemplifies the most basic conservation laws found in nature.

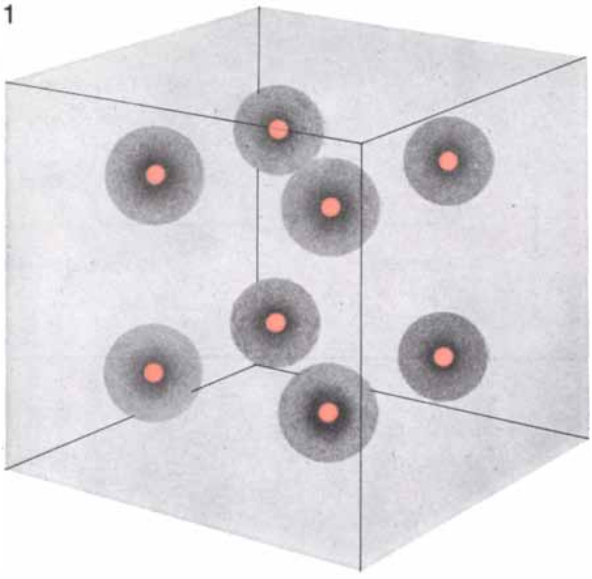
Since the discovery of the positron as the antiparticle of the electron some 40 years ago entire systems of antiparticle pairs have been predicted theoretically. Many of the antiparticles have been experimentally identified, among them the antiproton. It has been demonstrated in the laboratory that nuclear forces act between antiprotons and antineutrons just as they act between protons and neutrons to hold together the nuclei of ordinary atoms. If an antinucleus were dressed with a cloud of positrons, it would be an atom of antimatter. Since antimatter can exist, it has been posited that, for the sake of symmetry in the universe, there must be equal amounts of antimatter and ordinary matter. The Swedish physicist Oskar Klein has suggested a cosmology in which galaxies of antimatter are interspersed with galaxies of ordinary matter [see "Antimatter and Cosmology," by Hannes Alfvén; SCIENTIFIC AMERICAN, April, 1967]. If that is the case, some mechanism must be preventing the two types of galaxies from colliding so that they do not annihilate each other. One mechanism would be supplied if antimatter and matter repelled each other by antigravity.

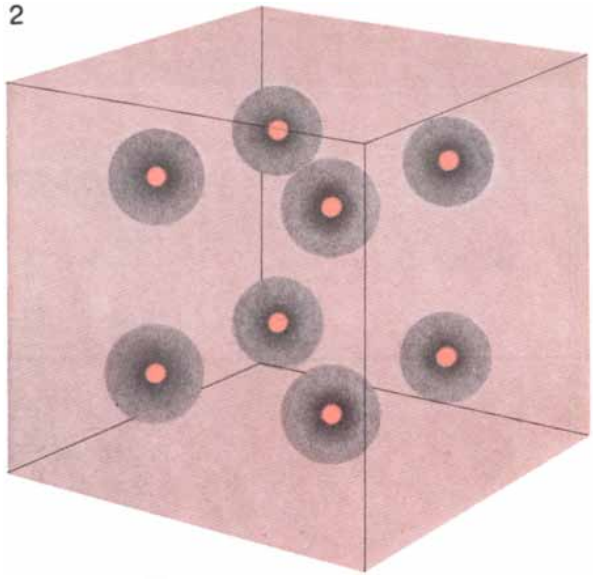
The positron has provided the first clues, however, that the gravitational interaction between particles having mass is universally attractive. The strong electric field immediately surrounding the nucleus of an atom gives rise to a cloud of electron-positron pairs in the region close to the nucleus. The late Leonard I. Schiff of Stanford University calculated that if positrons were gravitationally repelled by the matter in our world, then the ratio of the gravitational mass (the weight) to the inertial mass (the resistance to change of motion) would be measurably smaller for elements with heavy nuclei than it is for elements with light nuclei. Conversely, if positrons were gravitationally attracted to all kinds of matter in the same way, the ratio would be the same for all elements. The Hungarian physicist Roland von Eötvös measured such mass ratios at about the turn of the century, and he found that the ratios were the same to an accuracy of one part in 10^8 for all the elements investigated. His results were confirmed by Robert H. Dicke of Princeton University in 1963 to an accuracy of one part in 10^{11}. Eötvös' results alone were sufficiently accurate to show unequivocally that matter and antimatter, at least as they have been sampled in the form of positron-electron pairs in the vicinity of atomic nuclei, both obey the law of universal gravitational attraction.

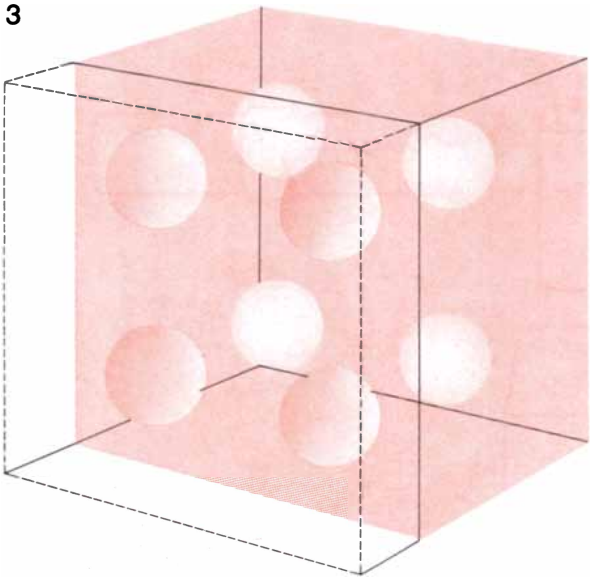
In the course of such heady research into the basic stuff of which the world is made it was noticed that the characteristics of the annihilation of positrons by electrons that were bound in atoms in solids can vary slightly among various solids. If the annihilation laws are fundamental and unvarying, such changes must reflect differences in the electronic milieu inside the solids. Beginning with the pioneering work of Sergio De Bene-

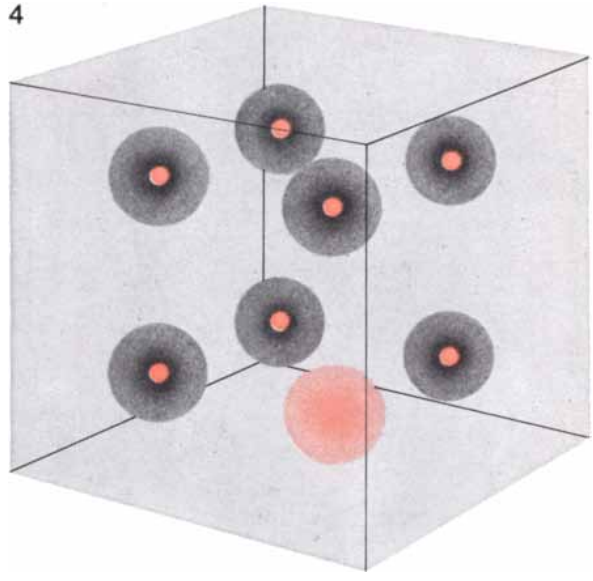
detti at the Carnegie Institute of Technology in 1950, a small group of investigators began to develop an intriguing new tool for materials research: the positron method. Employing positrons as a probe, they launched a systematic effort to explore the electronic structure of solid matter.

The positron method is confronted with the inherent difficulty that the positive charge of the positrons distorts the configuration of the electrons in their vicinity, the very configuration one wishes to study. Much ingenious thought has been devoted to finding ways of accounting for the distortion and correcting for it. Over the past few years, however, the research has entered a new phase as a result of the discovery that positrons, because of their positive charge and small mass, are extremely sensitive to structural imperfections in solids. The positrons' sensitivity is related to the fact that, since some of the de-





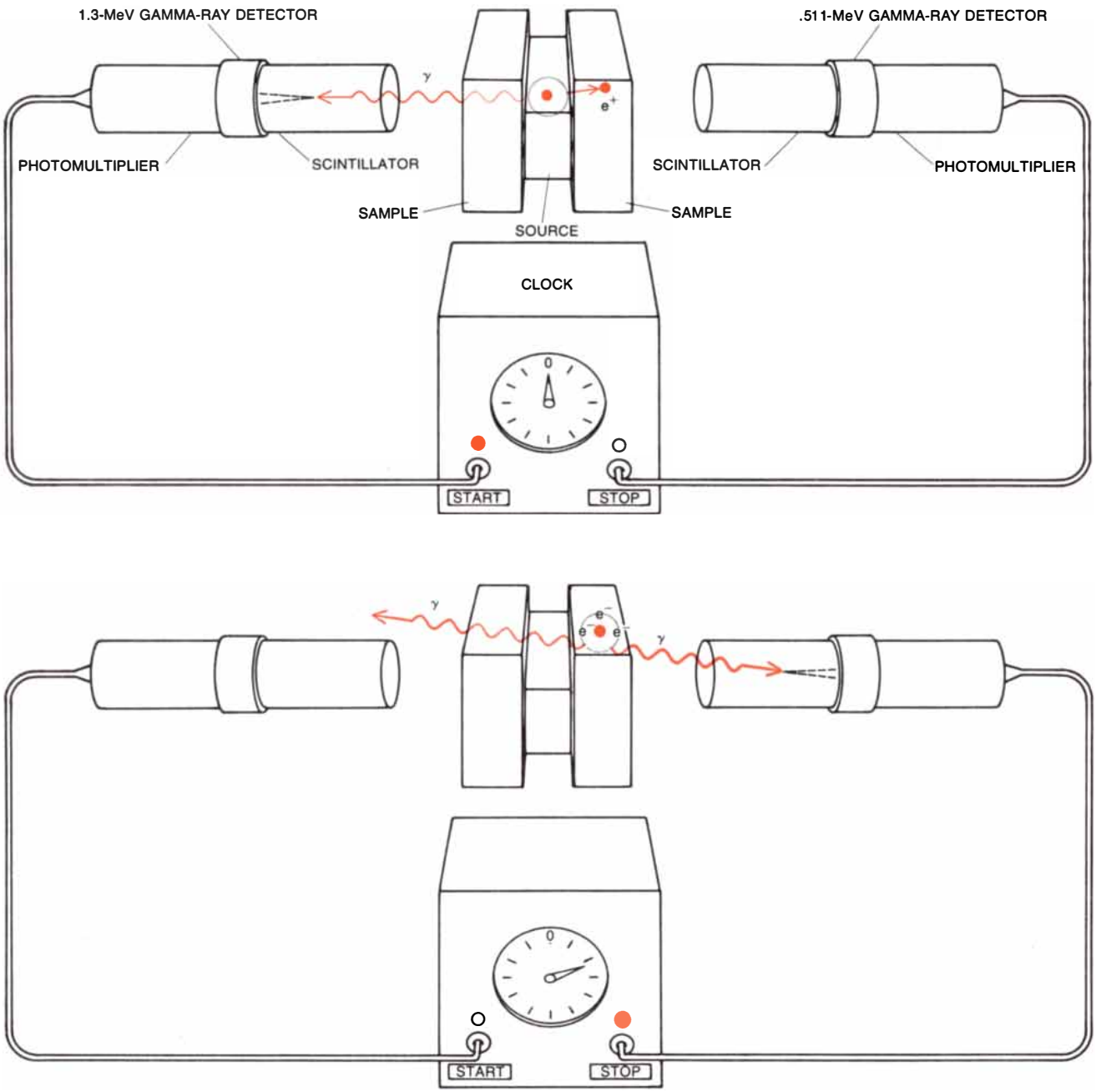




DENSITY DISTRIBUTION OF A POSITRON, the antimatter counterpart of the electron, in a metal conductor is determined by the fact that both the positron and the electron behave as waves as well as particles. A pure metal with a regular cubic structure (*1*) can be represented as an array of positive atomic nuclei (*dark color*) surrounded by a cloud of negatively charged electrons (*gray*). The density of the gray represents the density distribution of the electrons at any point, and in a conductor there are always some free electrons forming an electron gas within the bulk of the crystal. When a positron (*light color*) is injected into the metal, it distributes itself uniformly within the metal (*2*). It is repelled, however, by the positively charged atomic nuclei, and so it forms a region of reduced positron density centered on each atomic nucleus. The result is that the positron has a "Swiss cheese" density distribution, which is shown alone for purposes of clarity (*3*). If one of the atoms in the metal lattice is missing (*4*), the vacancy acts as a negative center to which the positron is attracted. Positron lives longer in the vacancy than in the bulk of the crystal before it is annihilated.

fects are negatively charged, they respond to them quickly and can be annihilated in a milieu that differs from the one in a perfect crystal.

Advanced solid-state technology is intimately linked to the physics of imperfections in crystals, that is, irregularities in the regular lattice of atoms in a perfect crystal. The performance of electronic solid-state devices depends on the types of defect that are incorporated into them in vanishingly small concentrations. A metal consisting of a perfect crystal would be too ductile for any mechanical requirement, so that ever since men first learned to harden tools on anvils and in open fires the properties of metals have been tailored for specific purposes by deliberately introducing structural defects through tempering by heat, through the addition of impurities, through working the metal mechanically and through alloying it with other metals. On the other hand, metal implements become brittle and break when they are fatigued as the result of unwanted defects building up after repeated stress and strain. The color of blue diamonds and the darkening of photochromic eyeglasses in sunlight stem from



LIFETIME OF A POSITRON IS MEASURED in this apparatus. A positron (e^+) is emitted from a source, sodium 22, sandwiched between two slabs of the sample material under investigation. A gamma ray (γ) having an energy of 1.3 million electron volts (MeV) is emitted simultaneously with the positron and is detected with a scintillation counter mounted on a photomultiplier. Signal from the ray's detection starts an electronic clock (*top*). Depending on the structure of the sample, the positron lives for 100 to 1,000 picoseconds within it before it encounters an electron (e^-) and is annihilated into two gamma rays, each with an energy of .511 MeV. (One picosecond is 10^{-12} second.) One of the gamma rays is detected by another scintillation counter and stops the clock (*bottom*). Clock readings of a few million annihilations are stored electronically and are analyzed by computer to find the mean lifetime of positrons.

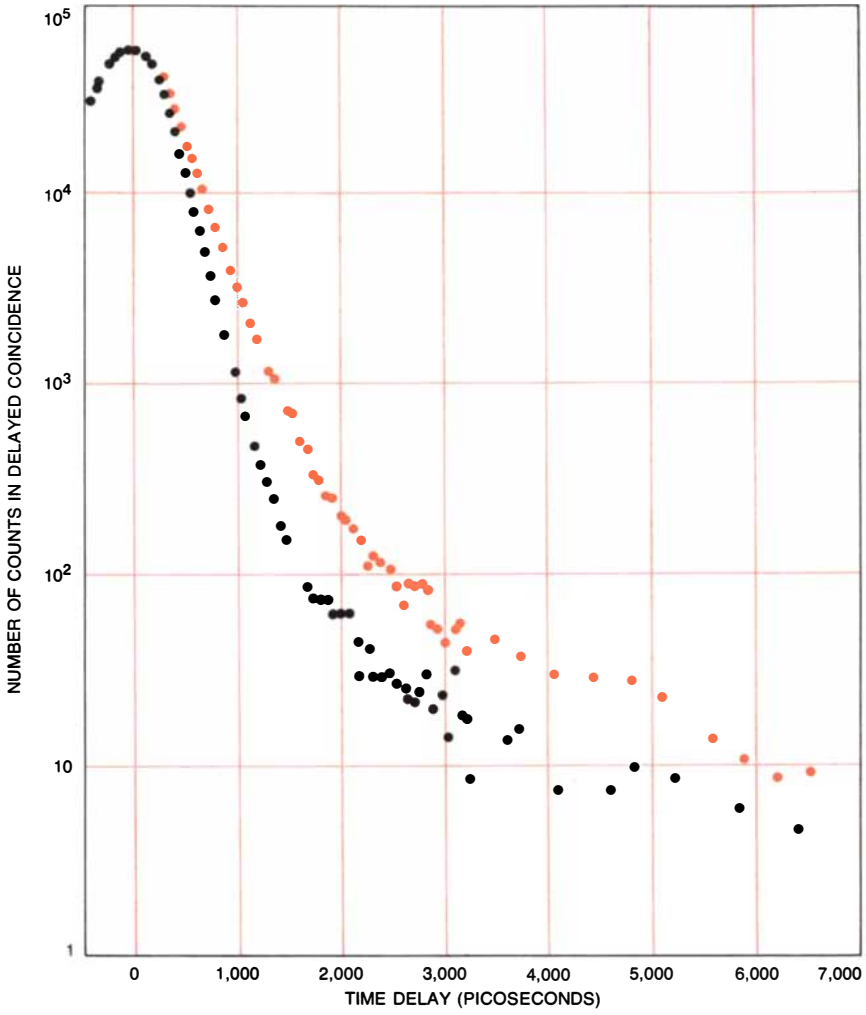
defects called color centers in the electronic structure of these materials.

The positron method has begun to contribute to our understanding of all these phenomena in the past five years. Now that the need for new sources of energy confronts us with special urgency, scientists and engineers working on materials for the nuclear power plants of the future are preparing to resort to the positron method to monitor the performance of solids under the extreme conditions of temperature and radiation that must be sustained by the materials of the reactors.

Any piece of matter that can be lifted in one hand—a brick, a book, a glass of wine—is made up of some 10^{24} atoms. If we want to describe and control the properties of a material consisting of so many particles, we cannot do so by keeping track of each of them. Even if that were possible by some computational tour de force, the amount of the data amassed in the process would be so enormous that the task of extracting useful information from them would be as formidable as their original acquisition.

One therefore resorts to probability statements about the many-particle system as a whole. The positions of the electrons are expressed as the distribution of the electron density in the substance. The velocities of the electrons are expressed in terms of the distribution of their momenta, where the momentum of each particle is the particle's mass times its velocity. Since one electron is indistinguishable from another, this pair of distributions actually comprises all the information that one can in principle gather on the system. It enables one to predict the atomic condition that determines all the macroscopic properties of materials: the probability of finding electrons at a certain point with a certain momentum. Understandably, then, physics places a high premium on any method that can determine something about the density distribution or the momentum distribution of the electrons and the atomic nuclei in solid matter. The positron method can do both.

Positrons are obtained from preparations of sodium chloride—ordinary table salt—in which some of the usual sodium-23 atoms have been replaced by atoms of the radioactive isotope sodium 22 produced in a cyclotron. Sodium 22 emits positrons in the course of its decay. There are other sources of positrons, but sodium 22 has the advantage that with each positron it simultaneously emits a gamma ray that has an energy of 1.3 MeV. The gamma ray can be detected and can trigger an electronic clock to signify the birth of the positron.

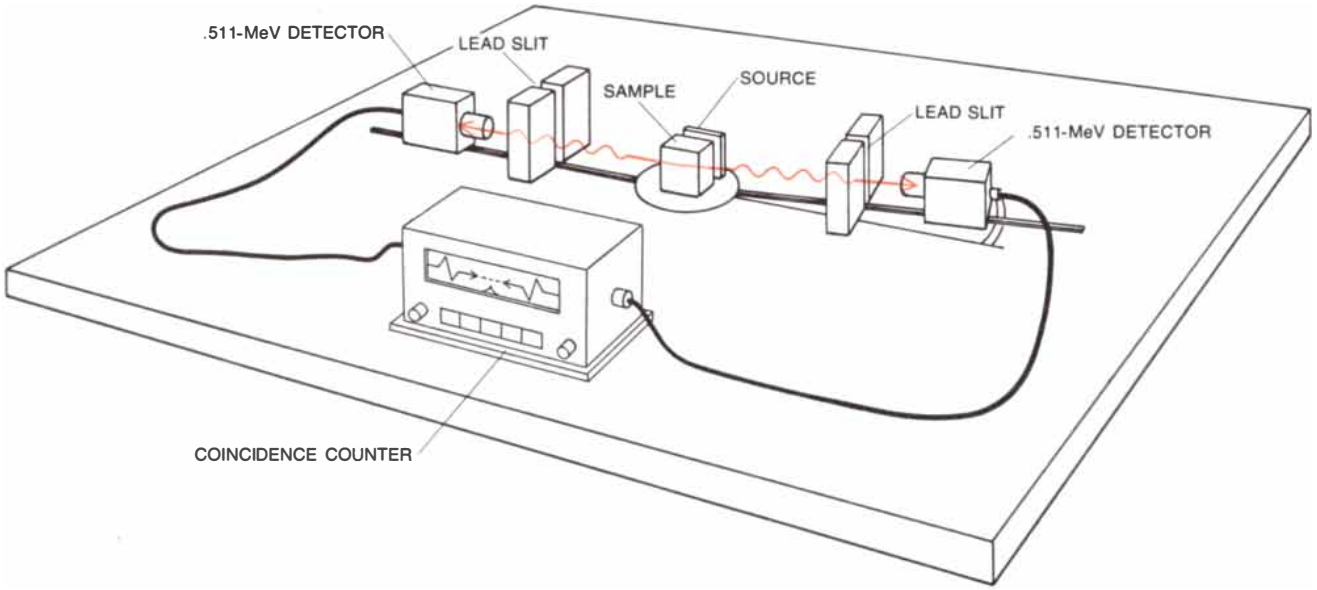


POSITRONS LIVE LONGER IN DEFECTS, as is shown by this chart of the number of annihilations in copper perfected by annealing (*black*) compared with the number of annihilations in copper fatigued by repeated flexing (*color*). Number of annihilations is plotted as a function of the time delay between the reception of the 1.3-MeV birth signal of positron and the reception of a .511-MeV gamma ray photon signaling its death. Slope of the declining curves is proportional to annihilation rate and inversely proportional to the lifetime.

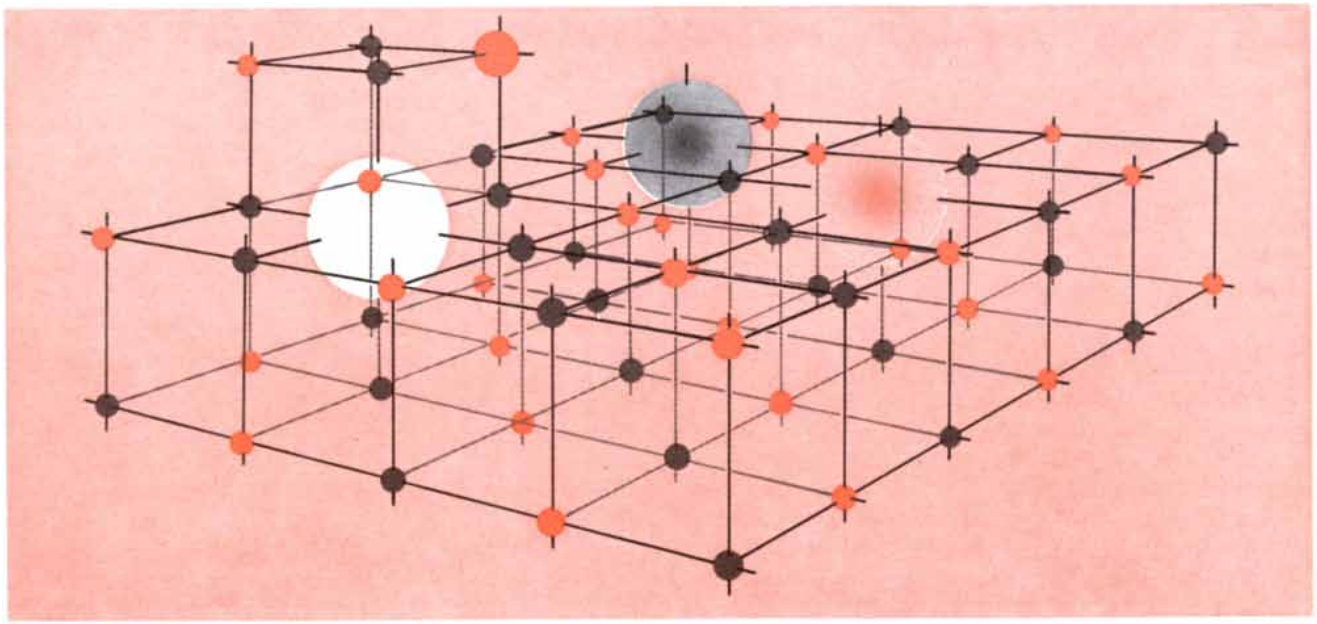
When positrons are injected into a solid, they come to rest very quickly. The rate of their subsequent annihilation is proportional to the electron density at the site of the annihilation. That rate is determined by measuring the mean lifetime of the positrons, that is, the average value of many measurements of the time interval between the moment when a 1.3-MeV gamma ray signifying the birth of a positron is detected, starting the clock, and the moment when one of the .511-MeV gamma rays from the annihilation of the positron is detected, stopping the clock. Positrons typically live a few hundred picoseconds, roughly the time it takes for light to travel one centimeter. (One picosecond is 10^{-12} second.) The clock must be able to resolve differences of a few picoseconds between the mean lifetimes of positrons that are annihilated in different electronic milieus. Positrons nonetheless live between 10,000 and a million times longer than the time it takes for electrons or atoms to complete one of their regular motions in a solid. In the context of those motions positrons are stable particles that during their lifetime obey the laws of the quantum theory of atoms, a circumstance that makes it possible to extract information from the characteristics of their annihilation about the steady-state bulk properties of the material into which they are injected.

The measurement of the lifetime of the positrons thus determines the elec-

tron density of the material. The momentum distribution of the electrons can be determined by observing the two .511-MeV gamma rays that result from the annihilation ending the positron's life. The gamma rays emerge exactly 180 degrees apart in the frame of reference where the annihilating positron-electron pair is at rest. Since the electrons inside the sample are moving, however, their momenta must be carried away by the gamma rays in order to conserve the total momentum of the system. The result is that in the laboratory frame of reference the gamma rays actually emerge at an angle that deviates slightly from 180 degrees. The amount of the deviation can be measured by having two detectors register the arrival of the gamma rays in coincidence, and arranging the detectors so that they can be swung independently around a central pivot holding the sample [*see upper illustration below*]. Even though the devi-



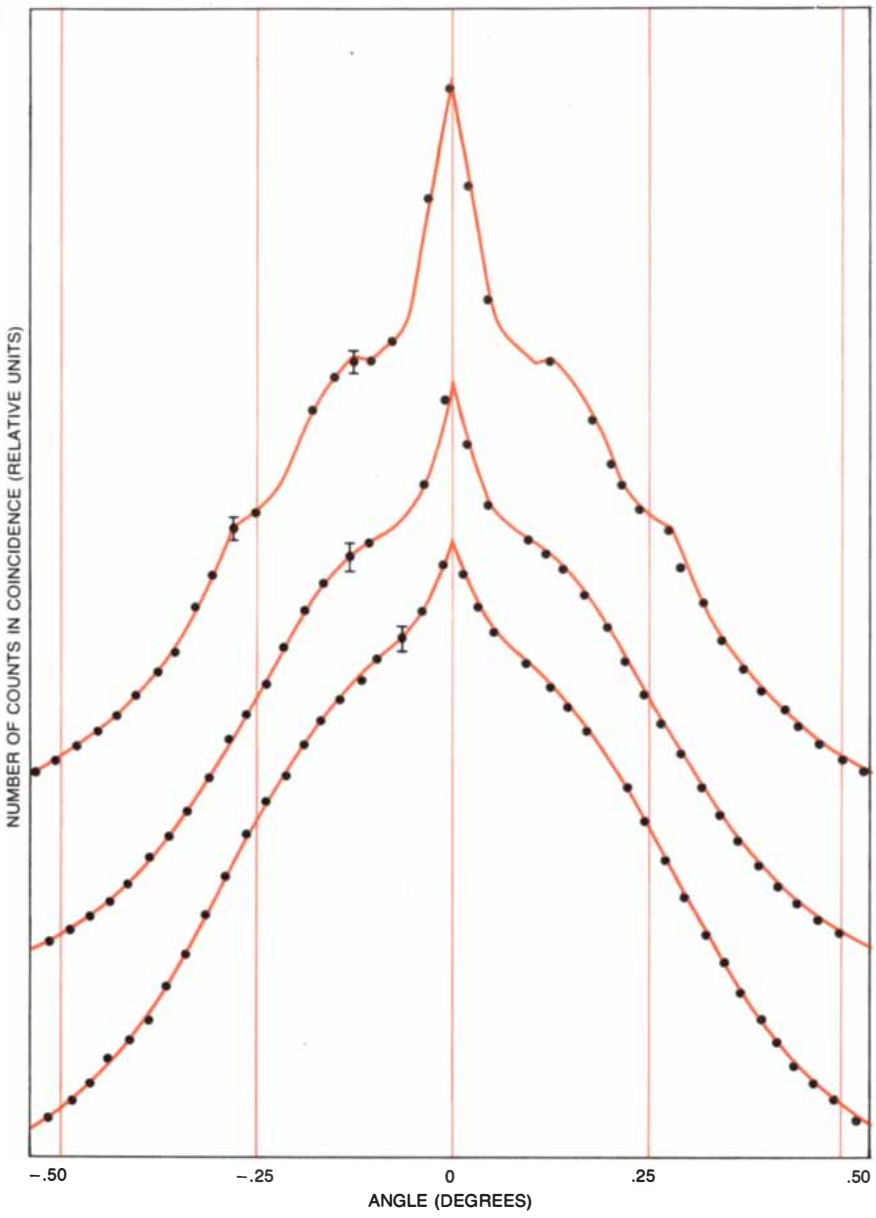
ANGLE BETWEEN TWO GAMMA RAYS emerging from the annihilation is measured by two scintillation counters connected in coincidence and swung on tracks around a central pivot holding the sample. Angle between two gamma ray photons is 180 degrees in a frame of reference where both positron and electron are at rest. Electrons move within the sample, however, and their momentum must be carried away by the gamma rays. Excess momentum causes gamma rays to emerge at an angle that deviates from 180 degrees.



ANNIHILATION CENTERS ARE FORMED when a positron is trapped by a vacancy in a crystal. The crystal shown is sodium chloride (ordinary table salt), which has a cubic lattice of positive sodium ions (*small dots in dark color*) alternating with negative chlorine ions (*dark gray dots*). If there is an impurity in the lattice, such as a doubly charged positive ion of calcium (*large dot in dark color*), crystal compensates for the extra positive charge by creating a vacancy (*white*) at the site of a sodium ion. Such a vacancy acts as a negative center that traps a positron (*dot in light color*). There is a positive analogue to the negative center created when a negative chlorine ion is missing; it can capture an electron (*light gray dot*) and form color center. Positrons trapped in annihilation centers encounter fewer electrons, and those electrons have lower momenta than the ones with which positrons annihilate in bulk of crystal.

ation from 180 degrees is very slight, on the order of .1 degree, refined techniques of detection and computation have been developed for measuring the angle precisely. The mean angular deviation indicates the average momenta of the electrons encountered by the positrons within the sample. Moreover, the shape of the angular deviation, that is, the scatter of the angular values around the average deviation, reveals details of the electronic structure in the substance where the positrons are annihilated. Beginning with the pioneering work of Alec T. Stewart of Queen's University at Kingston in Ontario and of Stephan Berko of Brandeis University, the positron method has told us a great deal about the electronic structure of crystals.

The positron method can also yield directly the kinetic energy of the electrons within the sample, as has been demonstrated by Innes K. MacKenzie of the University of Guelph in Ontario. Electrons in motion have a kinetic energy that is conserved in the annihilation process. That energy betrays its presence as slight variations, on the order of .001 percent, from the exact value of .511 MeV for the gamma rays that emerge from the annihilation. The variations in the energy are essentially a measurement of the Doppler shift of the gamma rays that result from positrons being annihilated in encounters with moving electrons.

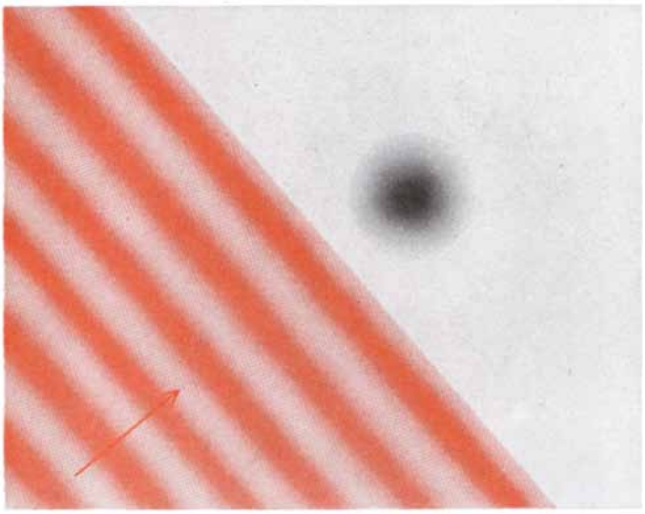


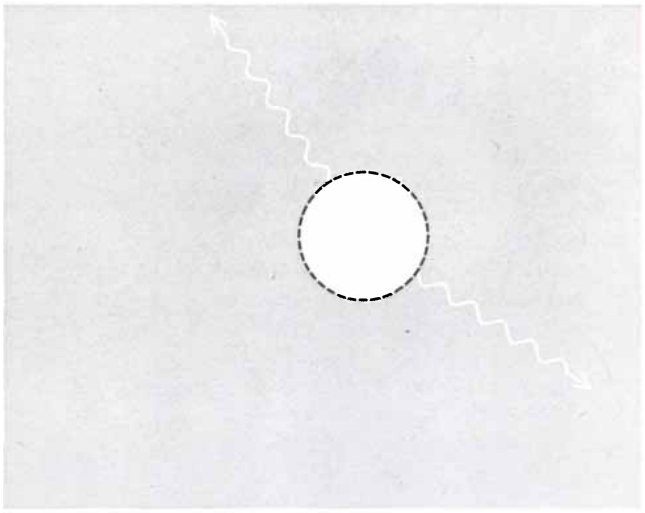
STRUCTURE OF THE ANGULAR DEVIATION can be more clearly resolved in pure crystals than in impure ones because the momenta of all the electrons and the positrons are well defined. In pure synthetic quartz (*top curve*) a pronounced structure appears at the angles that correspond to the periodicity of the momenta of annihilating positron-electron pairs. Quartz from natural sources in Madagascar (*middle curve*) and Brazil (*bottom curve*) is less perfect, and structure becomes washed out with increasing number of imperfections.

How is the behavior of positrons in a crystal affected by defects in the crystal? This is particularly well illustrated in an idealized way with ionic crystals such as sodium chloride. Sodium chloride has a simple cubic lattice structure in which positively charged sodium ions regularly alternate with negatively charged chlorine ions. If a negative chlorine atom is missing, the empty site in the lattice acts as a positive center and can trap an electron. The bound electron can absorb light, and hence it forms a color center that changes the color of the crystal. If a positive sodium ion is missing, the lattice vacancy acts as a negative center that can bind a positron and form what is termed an annihilation center.

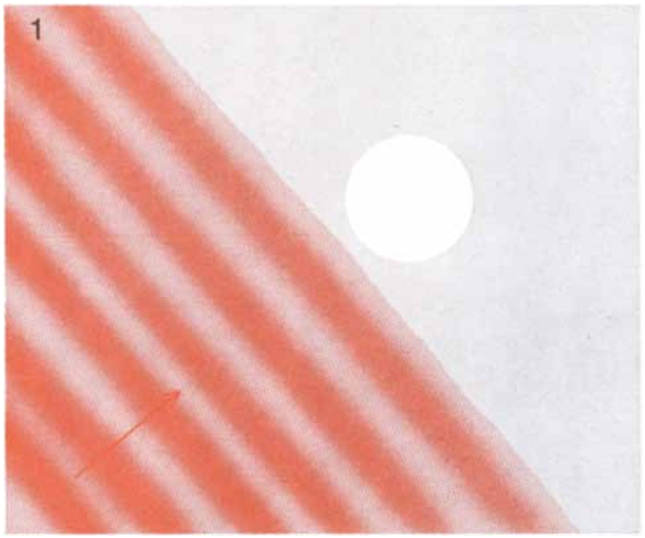
When positrons are deposited in the crystal, they move about randomly as prescribed by the quantum theory of particles in regular lattices. During their lifetime of some 200 to 500 picoseconds in the bulk of the crystal, they sweep through a region some 10 to 100 nanometers across. (One nanometer is 10^{-6} millimeter.) If a positron comes near a defect that can trap it, it stays in that location and forms an annihilation center. Positrons forming annihilation centers encounter fewer electrons than their counterparts in the bulk of the perfect crystal, and so they live longer. If the lifetime of the annihilation center is to a sufficient degree longer than the lifetime of the positrons in the bulk of the crystal, the difference can be detected experimentally. Moreover, the momenta of the electrons that the positrons encounter in an annihilation center are less than the momenta of the electrons that the positrons probe in the perfect atomic arrays in the bulk of the crystal; therefore the angular deviation from 180 degrees of the gamma rays coming from annihilation centers is less than the deviation of the gamma rays originating with annihilations in the bulk of the crystal.

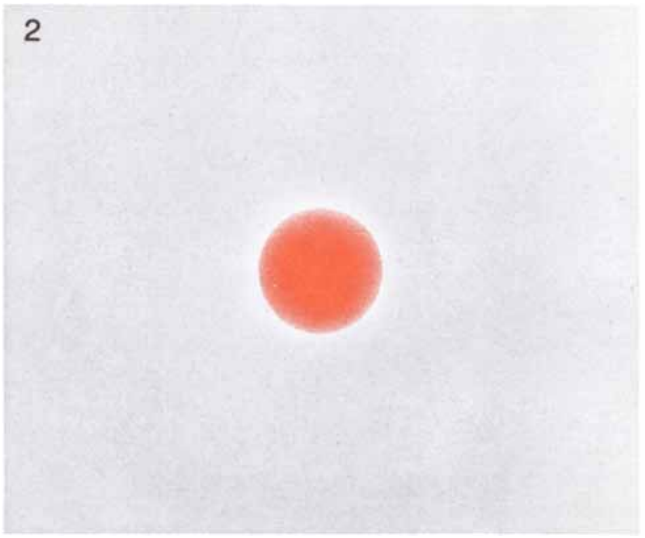
The probability that an annihilation center will be formed is proportional to the concentration of defects in the crystal. A measurable fraction of the positrons can be trapped in defects if only a few atoms in a million are missing from the lattice. If the concentration

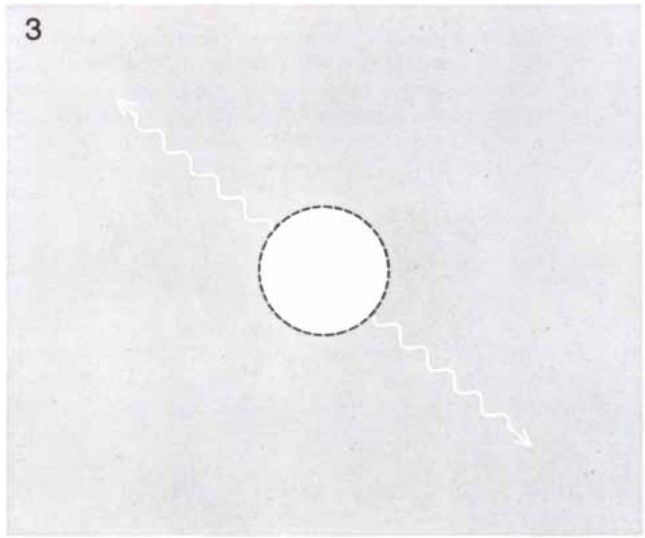


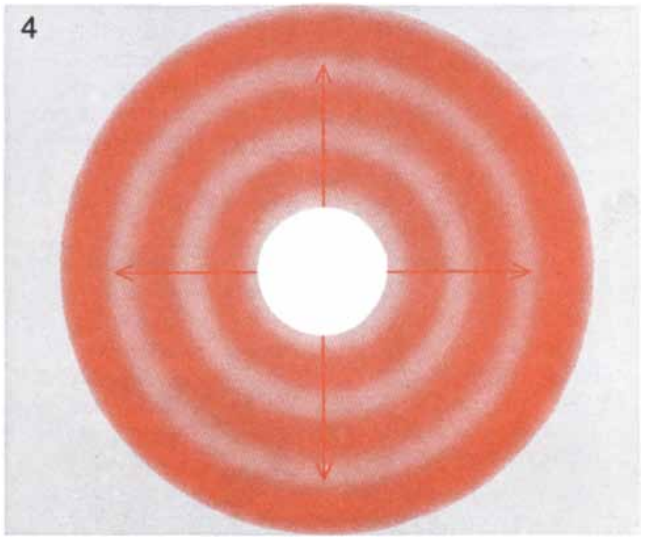


POSITRON PROPAGATES as a wave toward an electron in the bulk of the crystal (*left*). When the two particles annihilate each other (*right*), the resulting gamma rays carry away the momentum of the electron and emerge at an angle different from 180 degrees.









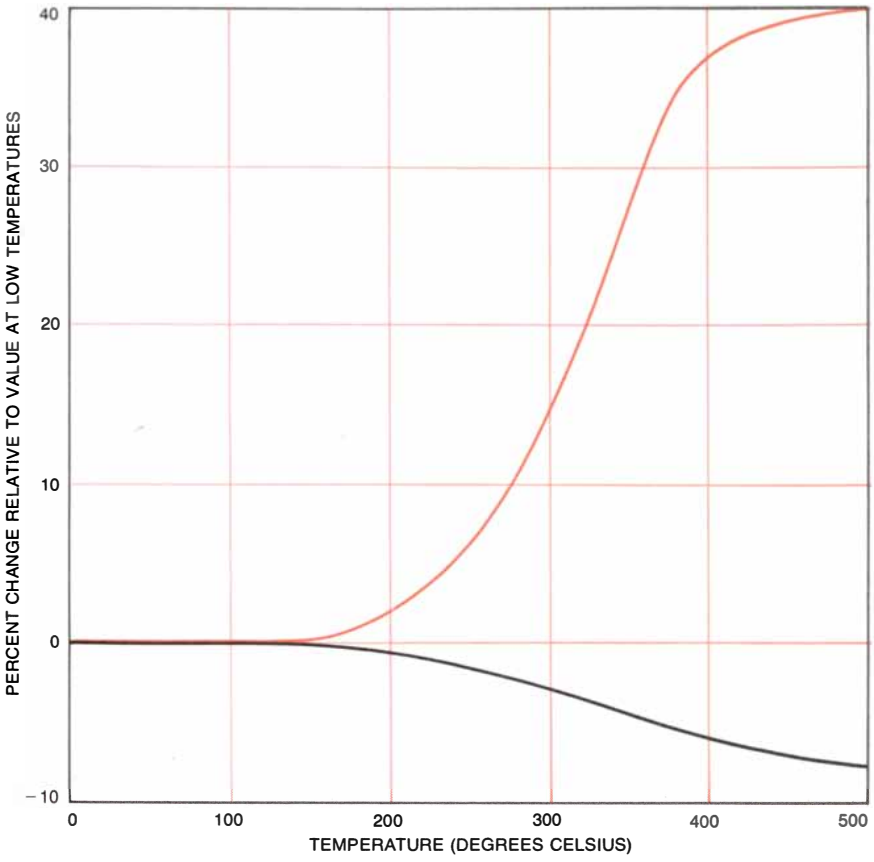
POSITRON PROPAGATES TOWARD A VACANCY (*1*), where it is trapped to form an annihilation center (*2*). The electrons (*gray*) that the trapped positron (*color*) encounters at the edge of the vacancy have lower momenta than the ones it would encounter in the bulk of the crystal. When positron is annihilated, resulting gamma rays (*3*) emerge at an angle of nearly 180 degrees. Only rarely does positron escape from an annihilation center before encountering an electron; if it does, it leaves as spherical wave (*4*).

of defects becomes high, with say one atom missing in every 10,000 lattice sites, practically all the positrons form annihilation centers before they encounter an electron. Positrons are thus an extraordinarily sensitive probe for nearly perfect solids, in which other methods of detecting defects fail.

The changes in the annihilation characteristics of positrons provide a genuinely new method for the study of ionic crystals because positrons respond to defects that are complementary to the defects that can be probed with electrons. One can deliberately create a certain number of positive-ion vacancies in a crystal by dissolving in it the doubly charged positive ions of an element such as calcium. Each doubly charged calcium ion replaces a singly charged sodium ion. In order to remain electrically neutral the crystal compensates for the second charge of the calcium ion by developing a positive-ion vacancy. When positrons enter the crystal, they can lodge in the vacancies. Experiments that Hsi-Fong Waung, Alfredo Dupasquier, Gottfried Dürr and I have conducted at New York University have established that the mean lifetime of the positrons increases systematically with the concentration of vacancies if more than a few sodium atoms in every 100,000 are missing in the crystal.

Positrons can serve as a microscopic probe to illuminate processes that produce complex defects. For example, if an ionic crystal is irradiated with X rays or subatomic particles, the irradiation gives rise to clusters of defects that affect the properties of the crystal in many ways. Two years ago I visited Robert Paulin at the French National Institute of Nuclear Sciences and Techniques, where we studied defects that were induced in sodium chloride by bombarding it with X rays. We observed that when positrons are injected into the sample, the angular deviation of the annihilation gamma rays changes as one raises the temperature of the crystal. In this way we were able to follow the migration of the X-ray-induced defects in the crystal until at a certain temperature the crystal regained its original structure. The observations support the conjecture that the mobility of the atoms in ionic crystals is linked to the migration of positive-ion vacancies—the very defects probed by the positron method.

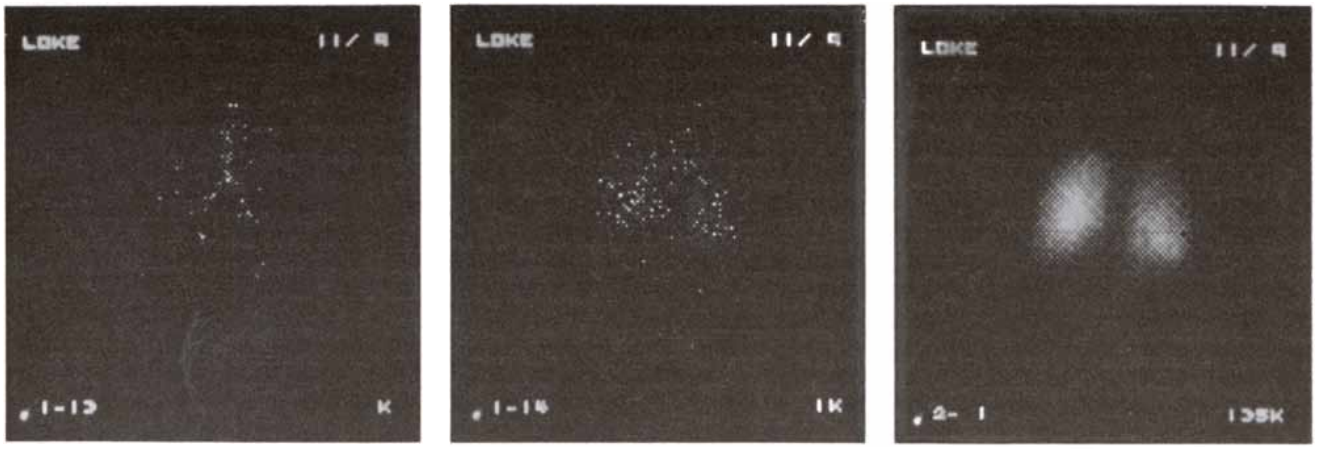
Processes that give rise to clusters of defects in crystals have attracted much interest because they bear directly on certain questions of applied science. For example, one needs to know to what extent certain ionic solids can survive as claddings for the walls of fusion reactors, the devices in which it is hoped that energy will be released from controlled thermonuclear reactions. As another example, abandoned mines in salt deposits, isolated from water for geologically long times, should be suitable repositories for radioactive waste materials, but salt crystals that are heavily damaged by radiation might explode or flow under pressure. If they did, the salt layers in the mines irradiated by such wastes could yield under the weight of the overlying formations, allowing water to seep into the mine and radioactivity to leak out of it.



MINIMUM ENERGY (E_v) that a crystal must acquire from the thermal motions of the atoms in it before a vacancy can be spontaneously created can be measured by the method of electron-positron annihilation. As the temperature of the metal is increased the mean positive lifetime (*color*) lengthens and the angular deviation between the gamma rays (*black*) narrows. Curves shown are for aluminum. Both yield same value of .67 electron volt for E_v.

Metallurgy has also profited from the positron method. In 1964 the Russian physicist I. Ya. Dekhtyar observed that the cold-working of iron-nickel alloys has the effect of narrowing the width of the angular deviation in the annihilation gamma rays. Berko subsequently confirmed the observation with aluminum and attributed it to the fact that positrons were being trapped in the defects known as dislocations introduced during the working of the metals. The field sprang to life and attracted wide interest outside the small community of positron physicists when MacKenzie and his colleagues reported three years later that the lifetime of positrons in a metal increases when the metal is heated to moderate temperatures of only a few hundred degrees Celsius. Conventional methods for detecting defects in a metal work only at temperatures near the melting point of the metal. The mean positron lifetimes change when there is less than one vacancy per million atoms in the crystal, a concentration rarely before accessible to experimental scrutiny. Subsequent measurements of the angular deviations of the gamma rays indicated that the width of the deviations narrowed in just the way that the theory of the formation of annihilation centers predicts from the changes in the mean positron lifetime.

Calculations have shown that the propagation of the positrons toward the



POSITRON SCINTIGRAPHS are images constructed by a computer from the signals from an array of scintillation detectors connected in coincidence. The detectors register the positions at which pairs of gamma rays originate from the annihilation of positrons and electrons. The scintigraphs are of a patient who has inhaled nitrogen gas containing the positron-emitting isotope nitrogen 13. Picture at left shows how the gas travels down the windpipe and into the two bronchi that branch below it. A tenth of a second later the gas has reached the alveoli (air sacs), shown in the middle photograph. Picture at right shows the lungs one second after inhalation. The positron camera was constructed by C. A. Burnham and Gordon L. Brownell of the Massachusetts General Hospital.

vacancies is not strongly affected by temperature and that once the positrons are trapped they are too tightly bound to escape from the annihilation center before they are annihilated by an electron. Therefore the changes in their annihilation characteristics are linked directly to the concentration of the vacancies and to the way this concentration varies with temperature. The central quantity governing the formation of vacancies with increase in temperature is the energy the crystal must acquire from the thermal motions of its atoms before a vacancy can spontaneously arise. That energy, denoted E_v, has tenaciously resisted reliable calculation by theory. It has to be determined by experiment.

Conventional methods of determining E_v are limited to metals with a simple crystal structure and are effective only when there is a high concentration of vacancies. At such concentrations, however, multiple vacancies interfere with the measurement. No such limitations apply to the positron method. It has already yielded a list of accurate new values for E_v in a large number of metals. In the few cases where one can compare the values obtained by the positron method with the values obtained by conventional methods one finds complete agreement. Preparations are now under way to measure the value of E_v in metals with very high melting points, such as niobium, that are candidates for the metallurgy of the future and for which no such information is available.

There are still other applications of the positron method. One such application is research into the electronic properties of the surfaces of solids, which are quite different from the electronic properties deep within the solid. Such research has been stimulated by the development of microelectronics and the need for efficient catalysts in the control of air pollution. Paulin first demonstrated with fine grains of powder that in insulators positrons and electrons can both escape through the surface. In metals, however, they are trapped in the surface region and the characteristics of their subsequent annihilation tell us about the electronic milieu at the surface. The positron method has also been applied to the study of voids in metals that form after long exposure to radiation in nuclear reactors, a study that has important implications for nuclear engineering. The voids can change the mechanical properties of the material and can also become filled with radioactive gases. The voids can migrate along temperature gradients toward the surface of the metal and thus bubble radioactive wastes into the atmosphere.

The positron method has also found its place in biomedical research. Energetic beams of mesons can be used to treat internal tumors by virtue of their ability to penetrate deep without damaging the intervening tissue. At the end of their range they cause atomic nuclei to fission, producing the desired therapeutic effect. Some of the fission fragments emit positrons. One can precisely locate the site of the reaction by detecting the gamma rays from the annihilation of the positrons with banks of detectors paired in coincidence. The signals from the detectors can control the accelerator generating the mesons and can in this way direct the end point of the beam to the tumor and lock it into position.

With positron "scintigraphs" one can follow the functions of living organisms. A scintigraph is an image constructed by a computer from the signals of many scintillation detectors arranged to detect the annihilation gamma rays in coincidence. For example, a patient can inhale nitrogen containing the positron-emitting isotope nitrogen 13. The scintigraphs clearly show how the gas enters the windpipe, passes through the bronchi and finally reaches the alveoli (air sacs) in the lung. Such pictures contribute to the study of the pulmonary function and may soon reveal details in the pattern of the flow of blood through the brain. Experiments are in preparation at the Lawrence Berkeley Laboratory of the University of California in which minuscule amounts of the positron-emitter oxygen 15 will be injected deep into the brain. Oxygen 15 has a half-life of only two minutes, but that should be long enough for it to be incorporated into the blood and for the blood flow to be followed with scintigraphs. Asymmetries in the flow pattern could well give early warning of malignant anomalies inside the cranium.

Thus the positron, which began by demonstrating the symmetry laws of elementary particles, has grown into a unique probe for exploring the microscopic properties of matter. The positron method is steadily being applied to new studies in materials science and medicine, and it should become an even more valuable tool as it matures.

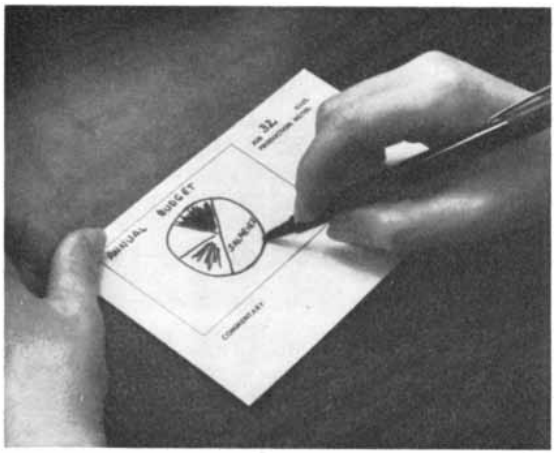
How to give a better-than-offhand talk without being a showman

1. Think about:

What does my audience have in common?
Why are they coming?
How much do they know already?
How much more do they want to know?
What attitudes will they bring with them?
What do I want to change or accomplish with my message?

2. "Multimedia" presentations can be written and produced for you by professionals. There are occasions where more modest efforts of your own might be more appropriate. Make sure this is one such.

3. Get a pack of 4″ x 6″ index cards.



4. Draw a large box in the upper left-hand corner of a card. In that box draw a crude sketch of what comes into your mind when you concentrate on one of the principal points you want to make. It may be a chart, clipping, symbol, diagram, or a photo of a person, place, or thing. Underneath state the point in as few words as needed to cue yourself to the thought.

5. Do a similar card about the thought that leads into the thought you have just expressed. Now do one about the thought that follows the first one. Keep going like that.

6. When you run out of ideas to tack on ahead or behind, think of the important points that haven't fallen into sequence yet. Make cards for them. **Always work up the sketch before the words.** (If lively words flow out of you too easily to work that way, probably anything you'd say on any subject would fascinate any audience. In that case you hardly need any of this advice.)

7. Arrange the cards on a table in an order that makes sense.

8. Now get critical. Is the development of the ideas too plodding? Would some other scheme of arranging the cards liven up the beginning and the end? Which cards should be tossed out? Where are you skipping too fast? Where are you trying to pack too much into a single card? Make out the additional cards you need.

9. Now get critical about your stack of cards from the standpoint of practicality. Some of your sketches would take too much time and art talent to turn into presentable slides. Substitute images easier to obtain from internal sources or the public domain. If you (and perhaps your secretary) are on your own for this, the KODAK EKTAGRAPHIC Visualmaker kit can much simplify your slide-making problems, both in copying extant material and in snapping originals.

10. Run through the talk. Make believe your sketches are already slides on the screen. Speak from the cues you've written under the sketches. (When you give the actual talk, the slides themselves may suffice as cues. Then you wouldn't be reading at all. Why assemble an audience just to hear you read?)

11. Decide whether you have too much or too little material. Discard or add cards accordingly.

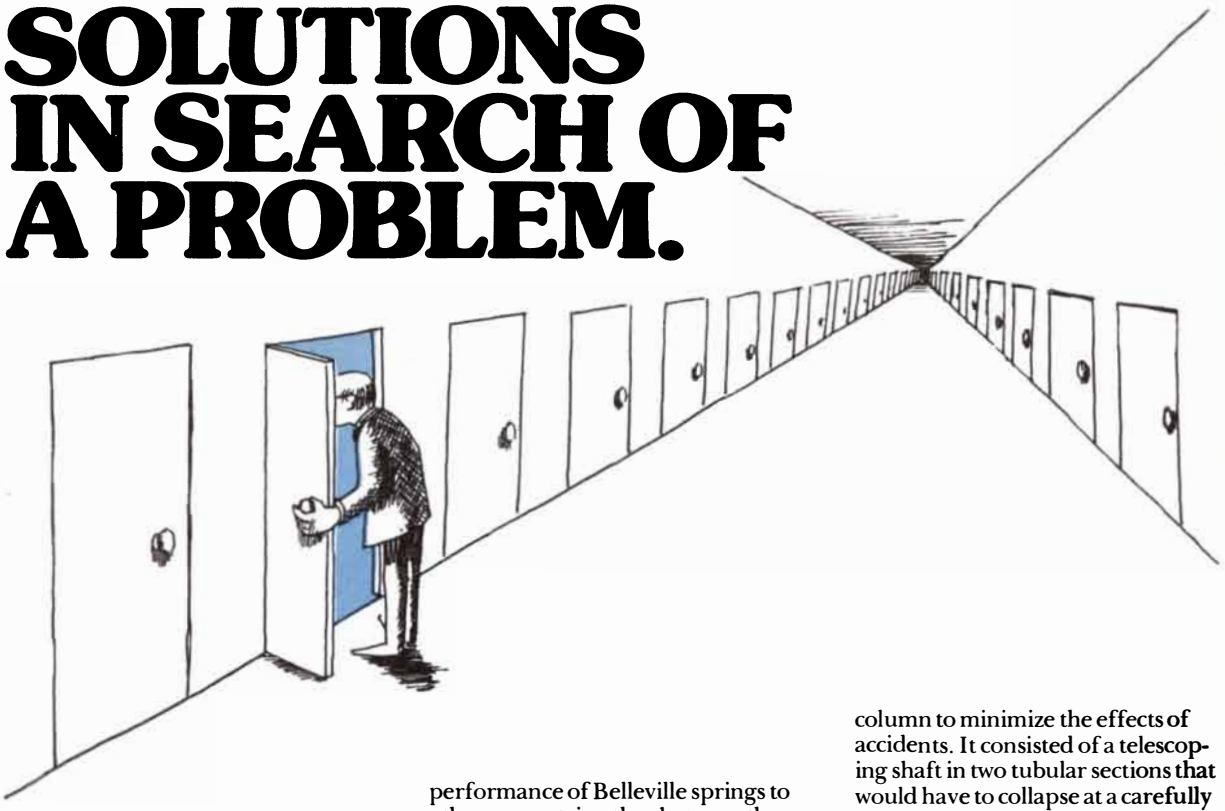
12. Now you are ready to prepare your slides. Worst of the pitfalls is type or other detail that the important but shy people in the back row can't quite make out on the screen. To avoid this and other mistakes, send for the free Kodak publication "Slides with a Purpose." Address request to Dept. 55X, Kodak, Rochester, N.Y. 14650. We'll throw in literature about the KODAK Visualmaker and other handy products.

13. You're on!

14. You're great!

Kodak

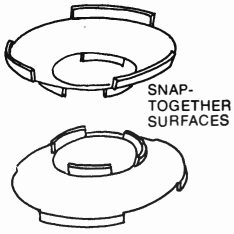
SOLUTIONS IN SEARCH OF A PROBLEM.



In a rational world, first comes the problem then the solution. Right? Right. . . but not always. Sometimes the solution comes first and the problem has to be searched out. A 100-year-old idea, for example, can be adapted to break through technological limitations imposed by older engineering materials. Which brings us to our story.

Merci, M. Belleville

Back in 1866 one Julien Belleville patented a coned metal disc spring that we now call, to no one's surprise, a Belleville spring. It has a convex saucer shape and a center hole. Under load it flattens and returns to its original shape when the pressure is off, even as you and I.



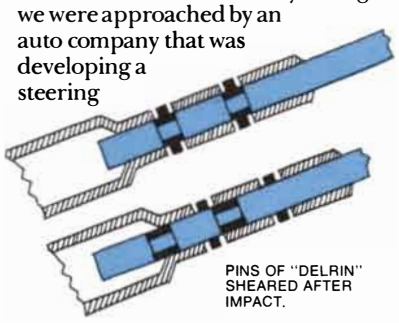
SNAP-TOGETHER SURFACES

So, we've been working on Belleville springs made of an engineering plastic – Delrin® acetal resin. We've designed springs of "Delrin", molded them as individual discs and as large sheets with *snap-together surfaces*. And we've even extended the performance of Belleville springs to other geometries, thanks to another pretty good invention, the computer.

We've got a lot of answers – for possible use in door closer springs, springs in switches, relays and push buttons, furniture innersprings and, from one of our engineers who thinks young, pogo sticks. Maybe you have a new or even old idea that can expand in function and usefulness when it's matched with our engineering plastics technology.

Born to fail

That isn't the name of a soap opera. Rather it describes an unusual technique: planned failure. We've always known that "Delrin", because of its reproducible properties, could be designed to fail at a predetermined stress. Then several years ago we were approached by an auto company that was developing a steering column to minimize the effects of accidents. It consisted of a telescoping shaft in two tubular sections that would have to collapse at a carefully predetermined stress.



PINS OF "DELRIN" SHEARED AFTER IMPACT.

Together with the company we worked out the technology of injection molding "Delrin" into the tubular sections as part of the mold. The design depended on "Delrin" shear pins and bearing collars.

And so we have another apt illustration of a solution looking for a problem and in this case finding it.

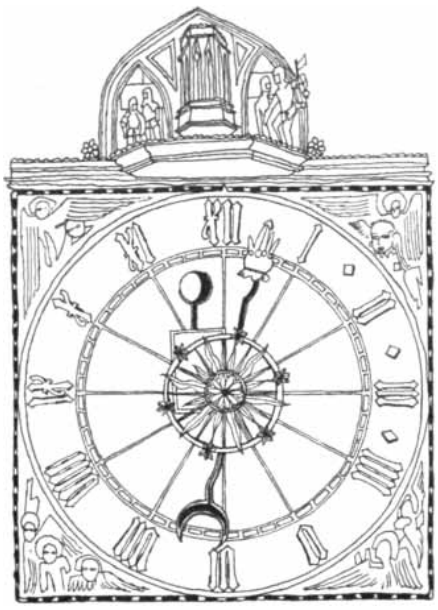
Invitation to a Dialog

Surprising breaks with tradition can happen when you look at your problem and our engineering plastics technology with a new, unrestricted curiosity. If our expertise in the properties and uses of engineering plastics may be of help, you can continue this dialog by either writing Dick Johannes, Du Pont Co., Plastics Department, Rm. D-13064, Wilmington, DE 19898 on your company letterhead, or calling Dick at (302) 774-5826.

DU PONT

REG U S PAT OFF

SCIENCE AND THE CITIZEN



What Knowledge Is to Be Sought?

When a scientist seeks public financing for his research, the appropriateness of his request is normally judged by juries of his peers: the review panels for the National Institutes of Health, the National Science Foundation and other Federal granting agencies. The 94th Congress has given numerous signs that it might like to make such judgments for itself. Senator William Proxmire, Democrat of Wisconsin, has issued several criticisms of the National Science Foundation for supporting what he judges to be "wasteful" or frivolous research projects. The House of Representatives, in passing a bill authorizing $755.4 million for the NSF for fiscal year 1976, adopted an amendment that would give Congress veto power over any grant by the foundation. A few hours earlier the House had narrowly defeated an amendment providing for congressional review of the foundation's funding of curriculum development. The specific target was a behavioral-science course for elementary schools. "Man: A Course of Study," created by the nonprofit Education Development Center, Inc., on a grant from the NSF, has drawn attack from some parents as being offensive to their "traditional values" and representing an attempt by the Federal Government to subvert the local control of education.

Coincidentally the Organisation for Economic Co-operation and Development, an international body with headquarters in Paris, offered an outsider's view of what has brought about the change in the climate of public support for science in the U.S. from openhanded enthusiasm to a generalized suspicion of science and a demand for emphasis on applied research. Social unrest, economic turbulence, environmental concerns and recoil from war in Vietnam all contributed, according to the OECD, to inducing the new mood.

The ferment was most plainly apparent in the House debate on the authorization bill for the National Science Foundation. Representative John B. Conlan, Republican of Arizona, began the attack with his amendment on curriculum research. He said the foundation had been using the taxpayers' money "to aggressively promote and market outrageous social-studies curriculum material." Although the amendment was defeated, it created an atmosphere favorable for the next one, which was proposed by Representative Robert E. Bauman, Republican of Maryland. "In the last year," he said, "dozens of newspapers and columnists have had a field day discussing various Government spending projects which strike ordinary people as, at best, ridiculous.... If your office is like mine, whenever one of these articles appears, you get a large volume of letters asking why the Government spends money on such projects and why we congressmen do not do something about it."

The Senate adopted an authorization bill that proposed more money for the National Science Foundation and did not include the Bauman amendment. Senator Edward M. Kennedy, Democrat of Massachusetts, who is chairman of the subcommittee that prepared the bill, said his group and the parent Committee on Labor and Public Welfare unanimously regarded the amendment as "not only unworkable but contrary to the principles which have brought this nation to its leadership position in scientific research."

It appeared likely that the amendment would be dropped by the conference committee reconciling the two versions of the bill. Nonetheless, H. Guyford Stever, the director of the National Science Foundation, called the amendment "a signal which all scientists should heed."

Can Do

Since the end of World War II, Canada has worked virtually alone among nations in an effort to develop a commercial nuclear power reactor that has heavy water (deuterium oxide) as a moderator and is fueled with unenriched uranium. One of the reasons for selecting this approach is that the CANDU (Canadian deuterium-uranium) system offers the efficient and economical use of uranium as a fuel. According to E. Critoph of Atomic Energy of Canada Limited, the performance of the 2,128-megawatt Pickering Nuclear Generating Station, the world's largest when it was completed in 1973, has shown that the Canadian-designed reactor is highly reliable and the most efficient in terms of fueling cost of any proved reactor system.

Critoph described the Pickering station, which consists of four identical 532-megawatt reactors, at a recent energy symposium in Washington sponsored by the American Physical Society. In 1974, he said, in spite of the fact that by then the Pickering station had been superseded in size, it produced more electricity than any other nuclear station: a net of 13,777,000 megawatt-hours. Ontario Hydro, which owns and operates the station, reports that the energy cost was only 7.49 mills (.749 cent) per kilowatt-hour. The energy cost was 20 percent lower than that for a coal-fired generating station consisting of four 500-megawatt units that had been built by Ontario Hydro at about the same time.

Four new reactors are being built at the Pickering site, which will double the station's generating capacity by 1982. In addition Ontario Hydro is building the Bruce Nuclear Generating Station, with four 750-megawatt reactors, on the shore of Lake Huron. Construction is well advanced: the first reactor is due to start up in 1976. Ontario Hydro plans to double the size of the Bruce station in the 1980's and has selected a site near Bowmanville on Lake Ontario for building another 3,000-megawatt nuclear power station that is scheduled for completion in 1982.

Hidden Charm

After a decade of searching without success for a particle exhibiting the property called charm, physicists have suddenly been confronted with at least three promising candidates. So far the evidence for charm remains inconclusive, and it may well turn out that there

is a better way to account for the observations, but for the first time charm appears to be less bizarre than alternative explanations.

Charm is an abstract property, postulated to explain the suppression of certain particle interactions. It is associated with the hadrons, the class of particles that feel the strong, or nuclear, force and are thought to be constructed of the fundamental entities called quarks. All the hadrons known could until recently be assembled from just three quarks; the charm hypothesis requires that there be a fourth. Particles containing this fourth, charmed quark may have been created recently in interactions of hadrons with high-energy neutrinos.

One of the candidates for charm has been detected in a bubble chamber at the Brookhaven National Laboratory by Nicholas P. Samios and a group of eight other physicists. The distinctive feature of the event is the production of a single particle that bears another abstract property: strangeness. The event photographed in the bubble chamber is believed to represent the collision of a neutrino that has an energy of 13.5 GeV (billion electron volts) with a proton; the products of the collision are a muon, four pions and a neutral lambda particle, which has a strangeness of −1 unit. The direct creation in this interaction of just one particle with negative strangeness would violate a well-established rule of hadron physics, but the apparent violation can be eliminated if the lambda particle is assumed to be the decay product of a charmed particle. The quark that bears the strangeness quantum number and the charmed quark are strongly coupled, so that charmed particles can readily give rise to strange ones. If this interpretation is correct, the particle being observed is a charmed baryon (baryons are hadrons made up of three quarks). It would have a mass of about 2.4 GeV, almost exactly the value predicted for a charmed baryon by Sheldon Lee Glashow and his colleagues at Harvard University. The principal reason for caution in evaluating the Brookhaven observation is that only one such event has been recorded. Alternative interpretations are possible, and although each separately is quite improbable, none can be eliminated until more data have been collected. At the Fermi National Accelerator Laboratory (Fermilab) a group of investigators under the direction of Frank A. Nezrick has employed a more powerful neutrino beam to search for events similar to the one discovered by Samios and his colleagues. If the rate of production can be reliably estimated from the single event at Brookhaven, about a dozen of the charmed particles should have been detected in the Fermilab experiment; none has been seen. Nezrick emphasizes, however, that his results are preliminary.

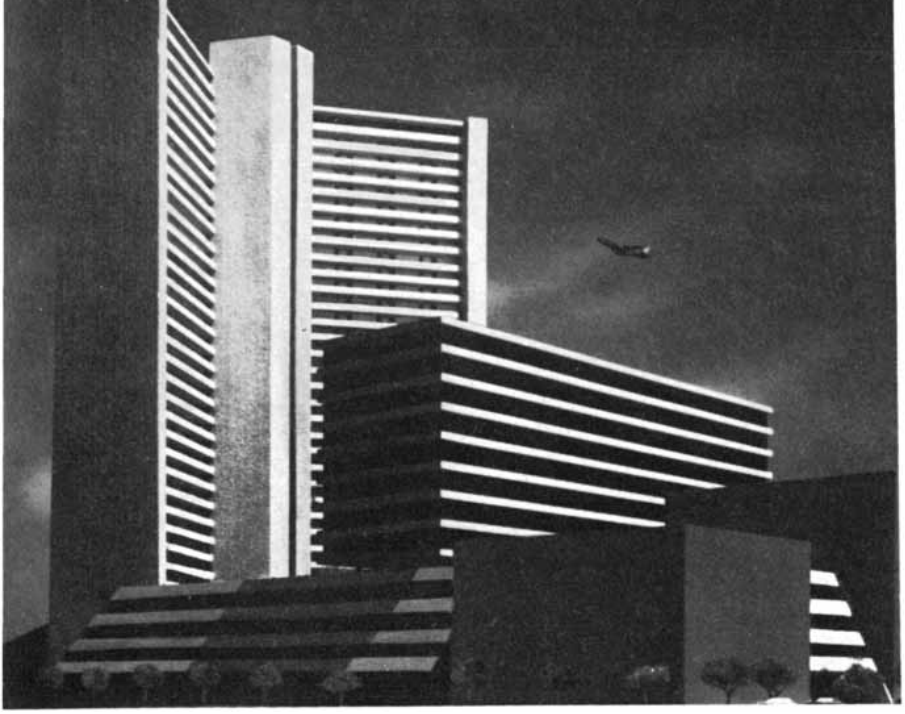
An interaction in some ways similar to the Brookhaven event may signal the discovery of a different charmed particle by Paul Musset and his colleagues at the European Organization for Nuclear Research (CERN). The CERN reaction also involved the collision of neutrinos with protons, but the strange particle produced was a neutral kaon rather than a lambda. More than 40 events have been recorded, so that there is ample material for statistical analysis; no conclusions are possible, however, because of an ambiguity in the interpretation of the data. The kaon can have a strangeness of either +1 unit or −1 unit, and the two forms cannot be distinguished in the CERN events. If the kaons have positive strangeness, the interaction is "normal" and there is no need to invoke charm in order to explain it; only if the kaons have negative strangeness is a charmed particle likely to be involved. If that should be the case, the kaons are probably the decay products of a charmed meson (a hadron made up of a quark and an antiquark).

Another curious result in neutrino-proton interactions has been observed at Fermilab by a group of workers from three universities headed by David B. Cline, Alfred K. Mann and Carlo Rubbia. They have found that in a small proportion of the events the final state consists of a neutrino or an antineutrino and two muons, all of which are leptons, the class of particles lighter than the hadrons. One of the muons can be produced directly by the transformation of the incident neutrino, but the other muon and neutrino must be the decay products of some unknown particle, which has been designated Y. The mass of the Y particle has been estimated to be greater than 2 GeV, far beyond the range of known lepton masses. It is therefore probable that the particle is a hadron; hadrons, however, can decay into leptons only by the weak interaction, a much slower process than decay by the strong or electromagnetic interactions. Some mechanism must suppress those faster modes of decay; one possibility is that the Y particle carries a property that must be conserved in strong and electromagnetic interactions but that can be changed in weak ones. Cline eschews speculation, but theorists (including Glashow) have proposed that the property could be charm.

The notion of charm has also been proposed to explain the new particles (named psi or *J*) discovered last year at the Stanford Linear Accelerator Center (SLAC) and at Brookhaven (see "Electron-Positron Annihilation and the New Particles," by Sidney D. Drell; SCIENTIFIC AMERICAN, June, 1975). The most widely held hypothesis concerning these particles is that they are mesons consisting of a charmed quark and a charmed antiquark. A complete set of quantum numbers, defining such properties as spin and parity, has now been determined for the psi particles; they are in accord with the predictions of the charm model. Several other predictions, however, have not been confirmed. In particular a large spectrum of energy states of the particles is expected, and transitions between states should be accompanied by the emission of gamma rays. Richard Hofstadter and his colleagues at SLAC have undertaken a search for the gamma rays, but none has yet been detected.

Scaffold of the Double Helix

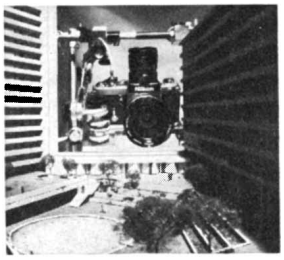
For all of the past 25 years' remarkable advances in cell biology, little has been learned about the nature and functioning of chromatin: the genetic material of eukaryotic cells, that is, all cells other than bacteria and blue-green algae. The genetic and biochemical investigations that identified DNA as the substance of the genes, deciphered the genetic code and showed how the code is replicated and translated into proteins were conducted largely with bacteria and their genetic material: filaments of DNA. In the cells of higher organisms the situation is more complex. The DNA is intimately associated with two kinds of protein (and a small amount of RNA) in chromatin, a highly organized nucleoprotein whose form varies with its function; it is "extended" when it is actively synthesizing new cell material, "condensed" when it is not synthesizing and highly condensed (into chromosomes) when the cell is dividing. The two kinds of protein, histones and nonhistones, play major roles in turning specific genes on and off in different cells and at different times. The histones, it has become clear, are also largely responsible for establishing and maintaining the structure of chromatin. For some time it was assumed that the double helix of DNA was somehow wrapped in histone molecules. In the past two years enough evidence has been accumulated from a variety of experiments to suggest that the scheme should be turned inside out; the histone molecules may be aggregated into a



rchitects Langdon & Wilson. A I A

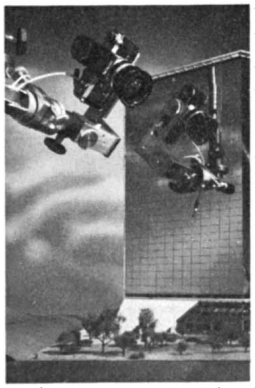
This building is 2 feet high

It's a model, of course. But, a model so well detailed and so superbly photographed that you can almost hear the hum of the elevators. If you were a prospective tenant, you could readily visualize its effect on the prestige of your business.



In achieving this exceptional realism, architectural photographer Jack Laxer has demonstrated for us not only his expertise but also his innovative and highly imaginative approach to this demanding application. When he recently acquired his new Nikon F2 cameras, he went to extraordinary lengths in exploiting their capabilities.

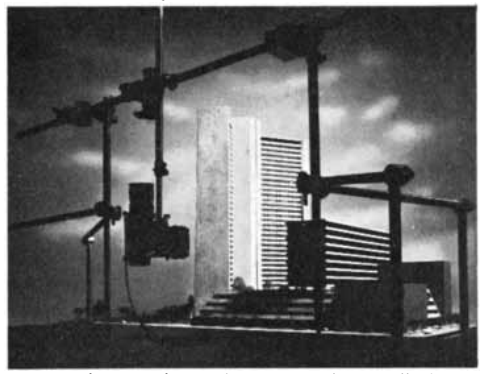
He designed ingenious support systems that allow totally flexible camera movement in all three directions and provides precise adjustments for about 800 different camera positions. It was a sizable investment in time, effort and money—but the unique combination of cameras, customized support system and talent has already paid off. The architects and builders, who are Laxer's clients, now can count on greater realism in photography for their promotional brochures. In fact, his Nikon photos are so accurate and so detailed that they enable clients to make design changes before the actual construction begins.



Apart from its legendary precision and reliability, the Nikon F2 offers special advantages for this type of work. Some are built in, such as the viewing system, the most accurate among 35mm cameras, and the multi-exposure provision that lets you add a background or put clouds in the sky.

Others are available in the Nikon System: The perspective-correcting PC-Nikkor lens, actually designed for architectural work. The 6X Magnifying Finder providing, as Laxer puts it, "an image virtually equal to 6x9 inches." The unique Nikon Bellow Attachment IV whose front standard movements permit corrective closeup photography. And, two of the nineteen interchangeable finder screens have grid patterns that simplify alignment of vertical and horizontal shapes in critical compositions.

This kind of precision photography has always been considered the exclusive domain of large, heavy studio view cameras. Yet, here it is done professionally with a 35mm camera that is best known for its speed and agility in handling split-second news, sports and similar action.



But, that is what Nikon versatility is all about. To help you make superior photographs of whatever interests you, whether a model just inches from the camera or a solar eclipse across millions of miles of space. To inspire you to look for new visions, new ways to picture your world and to test the capabilities of a camera whose limits have yet to be reached. And, to help you put it all together, there's the Nikon School of Photography, traveling to all parts of the U.S.A.

Your Nikon dealer will be glad to demonstrate the Nikon F2 and its system. Ask also about the Nikon School. Or write for folio 19, Nikon Inc., Garden City, New York 11530. Subsidiary/Ehrenreich Photo-Optical Industries, Inc. EPI (In Canada: Anglo-Photo, Ltd., P.Q.)

beaded central scaffolding with the DNA wrapped around them.

The structural role of the histones was first indicated by the fact that there is just about as much histone by weight in chromatin as there is DNA and that this equivalence holds true over a wide range of organisms; in other words, a certain amount of DNA goes with a certain amount of histone. There are five types of histones, separable on the basis of their weight and their amino acid composition. That composition is remarkably similar in a wide variety of organisms; such invariance, unusual for even functionally similar proteins in different organisms, again pointed to a structural role. Given the molecular weights of the various substances, there are about two molecules of the histones designated H2b1, H2b2, H3 and H4 and one molecule of the histone H1 for every 200 base pairs, or subunits, of DNA. How are the histones organized?

Last year Roger D. Kornberg, a Harvard University Junior Fellow working at the Medical Research Council Laboratory of Molecular Biology in Cambridge, and Jean O. Thomas of the University of Cambridge were able to extract the individual histones from chromatin and see how they reassociated. They found that two molecules each of the histones H3 and H4 form a four-molecule group, or tetramer, and that H2b1 and H2b2 form some other kind of small polymer. When they combined the tetramers with H2b1 and H2b2 and DNA, a complex was formed that yielded a specific X-ray-diffraction pattern that had previously been found to be characteristic of purified chromatin from cells: a repeating structure with intervals of about 100 angstroms. In an article in *Science* Kornberg proposed that the four histones and DNA constitute a repeating structure and that H1 is somehow added on. He pointed out that such a repeating unit would be accompanied by some 200 base pairs of DNA. That number coincides with the results of a large number of experiments in which chromatin is cleaved by nucleases, enzymes that digest DNA. In both plant and animal cells (and even in the animal-cell virus SV40, as Paul M. Wassarman and M. L. DePamphilis of the Harvard Medical School reported recently) the chromatin is most readily cleaved into pieces containing about 200 DNA base pairs.

Kornberg hypothesized that the H3-H4 tetramer may form the core of the repeating structure. That possibility is supported by the globular nature of the tetramer and by findings that H3 and H4 are the histones whose amino acid sequence is most strictly conserved across species and also are the most resistant to extraction from chromatin. Histones H2b1 and H2b2 would then act somehow as spacers for the tetramers. Most of the DNA of the repeating unit "would follow some path on the tetramer," with the rest of it connecting tetramers. A chromatin fiber would consist of "tightly packed DNA and associated protein alternating with more extended DNA and associated protein, rather like beads on a string." The 200-base-pair cleavage, then, would represent cleavage of the relatively more vulnerable connecting strands between tetramers.

Kornberg's proposal was supported by electron micrographs of chromatin made by Ada L. Olins and Donald E. Olins of the Oak Ridge National Laboratory and by Jack D. Griffith of the Stanford University School of Medicine, which showed alternating beadlike thick regions and thin regions. So far, however, the only clearly established dimension for chromatin appears to be the 200-base-pair DNA repeat. The exact grouping and structure of the histone molecules remain to be confirmed by more precise biochemical and X-ray analysis.

Caruso Lives

In the film *The Conversation* an expert in electronic eavesdropping (played by Gene Hackman) exhibits his skill by extracting the words of two conspirators that had been drowned out in the original tape recording by the blaring sounds of a group of street musicians. In the film the two scrambled signals, one representing speech and the other music, are unscrambled by analogue techniques that make use of filters either to remove or to emphasize particular frequencies of sound. Such techniques are widely employed commercially to eliminate scratches and imperfections from old recordings of famous performers.

In a recent issue of *Proceedings of the IEEE* Thomas G. Stockham, Jr., Thomas M. Cannon and Robert B. Ingebretsen of the University of Utah describe how digital techniques can be used for "blind deconvolution," defined as the problem of unscrambling two signals that have become "convolved" when both are unknown. Signals become convolved when sounds are reverberated or resonated or, in the case of photographs, when images are blurred. Stockham, who was one of the six experts on the panel convened to examine the famous 18½-minute gap in one of the Watergate tapes, originally became interested in the problem of blind deconvolution as it is presented in seismic recordings and in blurred photographic images.

In order to test various mathematical approaches to the convolution problem, Stockham and his colleagues experimented with early phonograph records made by Enrico Caruso in the pre-electronic era when acoustic energy was gathered by a large horn and simply fed into a sound box mechanically linked to a stylus that cut a groove in a wax master record. In the restoration of old records, contrary to a widespread impression, surface noise is not the most serious defect. The worst problem is the resonant or reverberant characteristics imparted to the primary signal by the primitive recording horns that were used to collect acoustic energy. The recording-horn contribution varied from recording to recording; if it had not, it could be extracted simply by analyzing a small number of records.

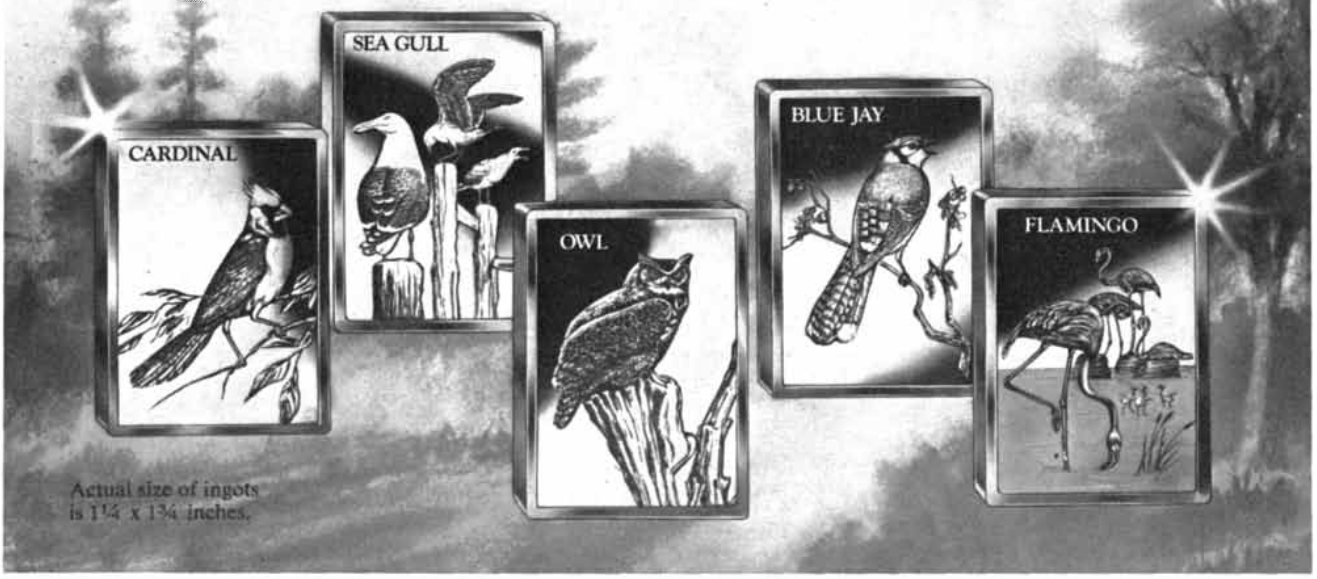
The problem was overcome in two steps. In the first step a single recording (in one example a 1907 recording of the aria "Vesti la Giubba" in Leoncavallo's *I Pagliacci*) was chopped into a succession of half-second intervals with a 50 percent overlap from one interval to the next. Each interval thus represents a slightly different musical passage, perturbed by the same recording mechanism. A special averaging process yields a mathematical expression in which the recorded waveform is represented by the sum of two terms: a voice spectrum and information revealing the reverberations of the recording system. The second stage in the restoration process is to remove the first of the two terms so that the second is revealed. That is done by subjecting a modern recording of the same musical selection to the same dissection process (in this case "Vesti la Giubba" sung by Jussi Bjoerling). In a modern recording the response of the recording system is so "flat" (that is, the response of the system to all frequencies is so even) that the averaging process yields an expression in which there is no second term. The first term (the voice spectrum alone) can then be subtracted from the right-hand side of the expression that was derived for the old recording, leaving the second term (the reverberations of the recording system) standing alone.

A total of about three hours on a computer is needed to deconvolve an old-style recording and remove the unwanted contribution of the recording horn. The digital version of the recording is then converted to the familiar analogue form on a phonograph disc.

The Hamilton Mint Announces a Great Buy for 7,500 Collectors and Investors

The Birds of America

A LIMITED EDITION OF PROOF FINISH PURE SILVER INGOTS



The 12 most popular Birds of America in a new series of ingots. Issued in a single limited edition of only 7,500.

Now rediscover one of America's great heritages, her native birds, in a magnificent new series, "The Birds of America". These beautiful art ingots will be minted in .999 pure silver and each will be a work of incomparable beauty.

A TREASURY OF FINE ART

Twelve flawless, gleaming proof finish ingots will be issued one a month for 12 months, and the collection will consist of a wide panorama of American birdlife from the majestic American Bald Eagle, to the delicate Cardinal and Jay, to the statuesque Great Horned Owl and the graceful Gull. All will be ideal to display individually or as a group by every collector who has an appreciation of beautiful art.

A SINGLE LIMITED MINTING

Issued in a strictly limited edition, we anticipate that this single minting will be one of the most important and highly coveted collections ever issued. Each ingot will measure 1¼ x 1¾ inches and contain 480 grains (one troy ounce) of .999 fine silver, the finest and purest available. Thus, the complete collection will contain *a full troy pound* (5,760 grains) of pure silver. The total number of sets to be thus issued, will be limited to only 7,500 in solid silver **(the smallest limit for a series that we have ever set).** Only one "proof" set will be allowed per subscriber and the dies will be destroyed once the edition limit is reached, thus fully protecting the integrity of this issue.

YOUR PERSONAL SERIAL NUMBER

Your ingots will be minted expressly for you and carry your own matching serial number on each ingot along with The Hamilton Mint Hallmark. In addition, you will receive a Certificate of Authenticity attesting to the limited edition status and precious metal content of your collection.

INVESTMENT POTENTIAL AND ORIGINAL PRICE GUARANTEE

Under the terms of this offer, The Hamilton Mint is officially committed to deliver to you, at the original issue price of $13.75 each, the complete set of 12 bird ingots over a 12-month issue period, no matter how high the price of gold or silver may climb. This is a particularly valuable guarantee, especially when economists predict a continued rise in precious metals in the years ahead.

Please remember that this edition is **limited to only 7,500** sets in pure silver, so act promptly to be sure that you may be included within the edition limits.

A SPECIAL BONUS OF WATERCOLOR PRINTS AND PRESENTATION CASE

All subscribers to the set will receive, with the compliments of The Hamilton Mint, a portfolio of beautiful bird prints, ready for framing. These original watercolor bird paintings by Aleta Brunettin show each bird in its natural habitat, and have been expressly commissioned by The Hamilton Mint for the sole purpose of accompanying "The Birds of America" Collection. Only enough prints will be produced for this purpose, thus making the prints themselves an extremely limited edition.

Subscribers will also receive, free of charge, a deluxe presentation case to house the entire "Birds of America" Collection.

Deluxe Heritage Version: 24 Kt. Gold on Silver
You may also order these fine art ingots in 24 Kt. Gold on .999 Fine Silver. The 2,500 sets thus offered will be extravagantly beautiful and very rare. They will also be individually serially numbered and hallmarked. Each will cost $18.75.

OFFICIAL ORDER FORM SMA-775

The Birds of America

LIMIT: ONE PROOF SET PER SUBSCRIBER

THE HAMILTON MINT
40 E. University Drive, Arlington Heights, Ill. 60004

Please accept my application for a complete Limited Edition Proof Set of "The Birds of America" Collection. I understand that I will receive my first ingot soon after my application is validated and accepted and thereafter receive an invoice once a month for the prepayment of the next ingot in the series. Enclosed is my check or money order for $_____ or charge my order as indicated below.

AN OPPORTUNITY TO BUY THIS SERIES AND SAVE

Please enroll me in series and:

☐ Send me first ingot in .999 Fine Silver for only $13.75 (plus 75¢ for postage and handling).

☐ Send me first ingot in 24 Kt. Gold layered on .999 Fine Silver for only $18.75 (plus 75¢ for postage and handling).

Charge to my: ☐ Master Charge* ☐ BankAmericard

Acct. No._____ Exp. Date_____

*If using Master Charge, also indicate the four numbers above your name here_____

Name_____
(please print)

Address_____

City_____ State_____ Zip_____

Signature_____
(must be signed to be valid)

I WANT JUST SINGLE INGOT: I understand that I can order just one ingot in series, Great Horned Owl, but then I do not get the savings, ingot will not be serially numbered and no future ingots will be reserved for me.

☐ Single ingot @ $14.95 in .999 Fine Silver.
☐ Single ingot @ $19.95 in 24 Kt. Gold on Silver.

(Please add 75¢ for postage and handling per ingot)

Application subject to acceptance by The Hamilton Mint.

Illinois and Louisiana residents add 5% sales tax.
Out of Continental U.S., add $1.50 for certification & registration.

THE HAMILTON MINT IS THE WORLD'S SECOND LARGEST PRIVATE MINT.

THE JOURNAL BEARING

In the oldest and still the commonest bearing a shaft (the journal) rotates in a sleeve (the bearing). Achieving low friction, long life and low cost in such bearings calls for sophisticated design

by John C. Bierlein

Modern machinery, no matter how complex or sophisticated, still runs mostly on direct descendants of the simplest and oldest kind of bearing: the journal bearing. A journal bearing consists essentially of two parts: an inner member, the journal, and an outer member, the bearing. The journal is a portion of a shaft or a shaftlike structure; it transfers the load to the bearing, which is in a support structure. In the journal bearing the part that moves is generally the journal (as with the shaft of a motor), sometimes the bearing (for example the wheel of a Conestoga wagon) and sometimes both parts (such as a connecting rod that joins a piston and the crankshaft in an automobile engine). Since many motors, engines and other machines incorporate journal bearings (sometimes designated plain bearings, sleeve bearings, fluid-film bearings or bushings), the annual production of journal bearings is in the billions.

The earliest evidence of bearings of any kind dates back some 5,000 years. The bearings were made of such materials as stone, wood, shell and bone and served in a few basic applications. An example is the lower pivot of a door. The door was hung on a vertical pole that was held at the top by a loop of hide and that turned at the bottom in a recess of stone or wood. At a later date pivots of this type were lined with copper or bronze. If any attempt was made to reduce friction or noise with a lubricant, the evidence is lacking.

The origins of the journal bearing were traced in the 1830's by W. Bridges Adams of England, who was exploring the ancestry of wheels and axles. He concluded that the wheel was a descendant of the long logs that served in ancient times as rollers for moving heavy objects. In due course, he said, it was recognized that two shorter rollers joined by a beam would be better than long logs. The axle and the short rollers, which developed into wheels, revolved as a unit and were held in place by notches in the frame that held the load or by pieces of wood attached to the frame. (Carts of this type were still in use in Portugal in the 19th century.)

Unless the two wheels were approximately the same size the cart would not have traveled easily in a straight line. Adams therefore reasoned that it was not long before the cartmaker contrived to have each wheel revolve independently around its own center on a journal bearing. This arrangement of axle, wheel and bearing had become standard by the time of the ancient Greeks and Romans.

The origin of lubrication for bearings is more obscure. The oldest example of a lubricated bearing is in a chariot, dating from about 3,400 years ago, that was found in an Egyptian tomb. A lubricant remaining on the axle was analyzed at the Cairo Museum and judged from its melting point of 49.5 degrees Celsius to be beef or mutton tallow. Some 1,000 years after the time of the chariot Herodotus wrote of petroleum refining, and 500 years after that Pliny the Elder compiled a list of vegetable oils. Both events suggest that the ancients recognized the need for lubricants of various kinds—animal, vegetable and mineral—for different purposes.

The Operation of Bearings

Different kinds of load call for different kinds of bearing. A load parallel to the axis of a shaft, such as the load between the propeller and the engine of a ship, is carried by a thrust bearing. A load perpendicular or radial to the axis of a shaft, as exemplified by a wheel turning on an axle, is carried by a journal bearing. If a combination of the two forces exists, a special bearing (an angular bearing, a taper bearing or a pivot bearing) is employed.

A particular form of journal bearing is the rolling-element (ball or roller) bearing. Here the load is transferred through balls or rollers. Rolling-element bearings also are used for thrust forces and combinations of forces. Because rolling-element bearings incorporate balls or rollers and are used to deal with multiple forces, they are generally considered separately from the standard journal bearing. The standard journal bearing is superior to the rolling-element bearing for many purposes because it can carry heavier loads, operate at higher speeds and is less expensive. Usually, however, it must be lubricated. A rolling-element bearing will generally continue to function even if the lubricant disappears. This advantage often dictates the application of rolling-element bearings. Here, however, I shall concentrate on the journal bearing, which serves to carry radial loads, whether the load is rotating or oscillating.

Fundamental to the design and operation of bearings is the study of lubrication, friction and wear. (This field of study is known, more commonly in Britain than in the U.S., as tribology, a term derived from the Greek for rubbing.) The first recorded study of friction and wear was made by Leonardo da Vinci in the 15th century [see "Leonardo on Bearings and Gears," by Ladislao Reti; SCIENTIFIC AMERICAN, February, 1971]. The classic laws of static and dynamic friction were established during the 17th, 18th and 19th centuries, notably by the successive French physicists Guillaume Amontons, Charles Augustin de Coulomb and A. J. Morin.

A discussion by Robert H. Thurston of friction in journal bearings appeared in *Scientific American* in 1873. Thurston

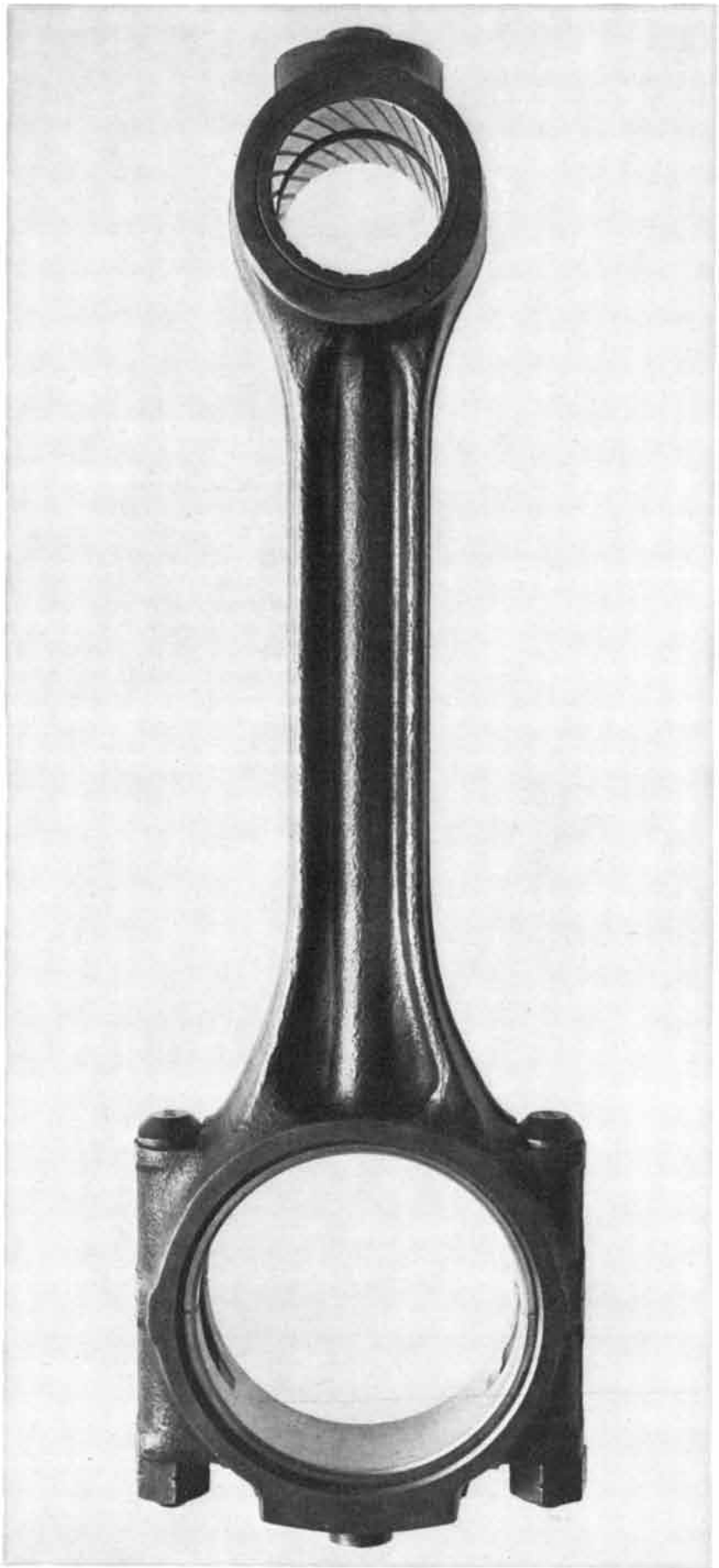
demonstrated that at high loads friction is proportional to the load and independent of the area of contact between journal and bearing. Friction at low loads, he reported, is a function of the viscosity of the lubricant and is approximately proportional to the area of contact.

The earliest lubricants—natural materials such as lard, tallow and vegetable oils—oxidized freely, and as they decomposed, fatty acids formed and adhered to the metal surfaces. In this way journal bearings received adequate boundary lubrication for the loads and speeds of the time. The mineral oils now employed are specially compounded and bearings are carefully designed; as a result bearings can function under heavy loads and at high speeds. In certain modern applications the lubricant is water or air.

During the 19th century the railroads were responsible for significant advances in the lubrication of journal bearings and for improvements in other aspects of journal-bearing technology. Their interest arose from the fact that at the time the load on every railroad-car wheel was carried by a journal bearing. The bearing was usually on only the upper side of the journal, embracing about 120 degrees of its circumference.

Initially the bearing was lubricated through holes in the top. The lubricant was a grease, which would soften when the bearing got warm and would flow down the holes and into the clearance between the journal and the bearing. The problem was that dirt would also get into the holes and make its way into the clearance, giving rise to problems of friction and wear beyond those associated with the movement of the journal in the bearing.

A major advance in lubrication was to have the bottom of the journal turn in a bath of lubricant. If the level of the lubricant got a little low, transfer rolls would apply it to the underside of the journal [*see illustration on next page*]. With underside lubrication dirt could settle to the bottom of the oil bath, where



PAIR OF JOURNAL BEARINGS installed in a connecting rod for a diesel engine are shown in this photograph. The rod connects a piston of the engine with the crankshaft. The piston pin is the journal in the top bearing, which is like the bearing portrayed on the cover of this issue. The bearing is cross-grooved to improve the lubrication. The crankshaft is the journal in the bottom bearing, which does not require cross grooves because the lubricant pressure generated during operation provides sufficient lubrication.

it would not impede the operation of the bearing. The oil bath added a complication, however, since it required the installation of a seal between the axle and the axle box to prevent the loss of oil.

Another major advance in journal-bearing technology was a replaceable bearing made of brass. Such a bearing was patented by Adams in 1847. Six years later he noted that some of the railroads employing the brass bearing were making it with an inside diameter somewhat larger than the outside diameter of the journal. The success of these bearings was one of the earliest indications of the phenomenon of hydrodynamic lubrication, which means movement of the lubricant (and thus continuous lubrication of the bearing and the journal) brought about by movement of the journal. Recognition of the importance of this phenomenon, however, awaited a thorough analysis of the lubrication of bearings, which was made a few years later.

The Analysis of Bearings

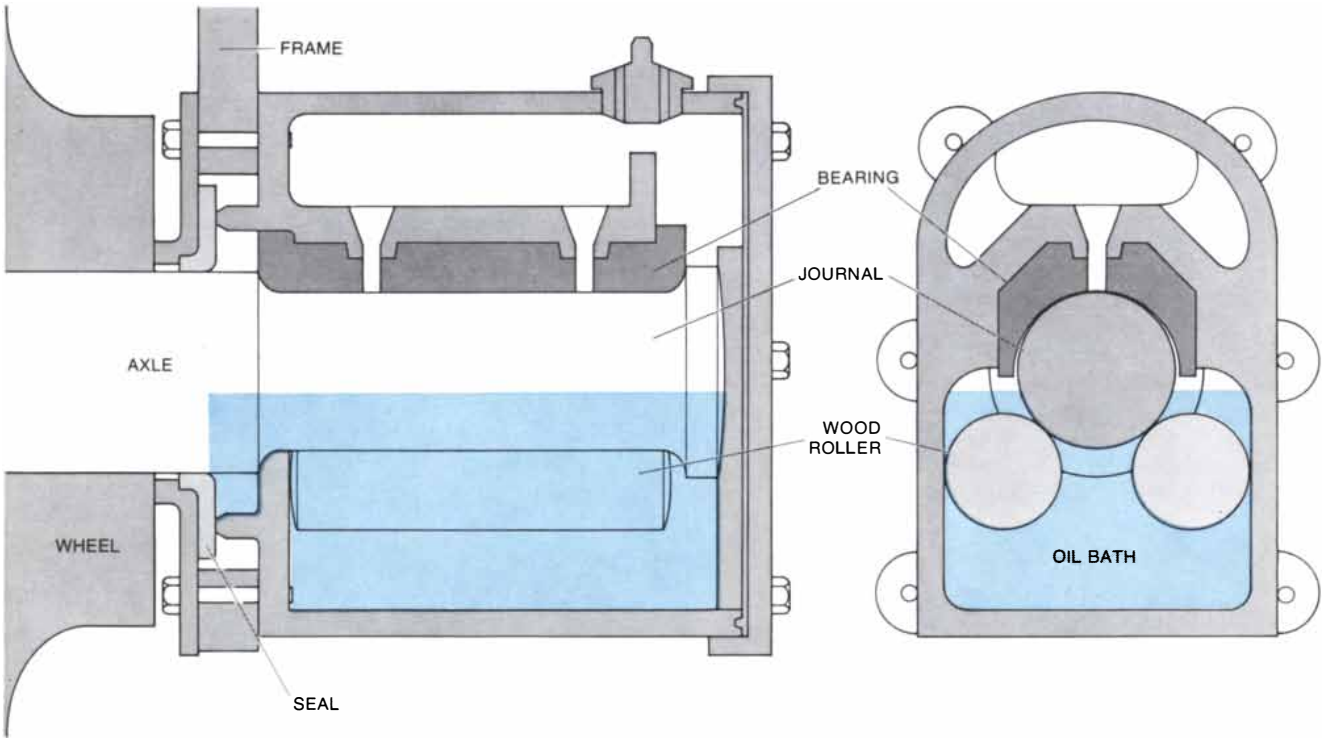
The first extensive laboratory analysis of journal bearings and lubrication was carried out in England in 1883 by Beauchamp Tower, who had been hired for the task by the Institution of Mechanical Engineers. Part of Tower's work involved experiments with a journal bearing that was lubricated from the bottom. He observed that when the journal turned, oil rose in the top hole and flowed out of it. He plugged the hole, first with a cork and then with a piece of wood. When the journal moved, the plug was slowly forced out by the oil.

Tower next put a pressure gauge in the hole. It gave a reading twice as high as the mean journal pressure (a function of the load on the journal and the area of the bearing). Tower checked the pressure in several locations and concluded that the bearing was floating on a film of oil.

Soon afterward Osborne Reynolds, one of the pioneers of hydrodynamics, analyzed Tower's experimental results and established the first comprehensive theory of lubrication. In 1886 he presented to the Institution of Mechanical Engineers the theory of hydrodynamic lubrication. According to this theory, the lubricant pressure observed when a journal is moving is generated by the motion itself, because the viscous lubricant is forced through a constriction, namely the wedge-shaped passage between the surface of the bearing and the journal. As a result the surface of the journal and the surface of the bearing are separated by a film of lubricant when the journal is turning rapidly enough to pressurize the wedge with lubricant.

The work by Reynolds also established several other important principles. One is the relation between the viscosity of the lubricant and the clearance between the journal and the bearing. Another is that the lubricant adheres to both the moving surface and the stationary one.

From this work it has been possible to derive a picture of the states of lubrication in a journal bearing. When the journal is at rest, there is a certain amount of lubrication because of the tendency of the lubricant to adhere to both surfaces. This condition is termed boundary lubrication. The condition persists as the journal begins to turn. Eventually the journal turns fast enough to generate hydrodynamic lubrication. Between boundary lubrication and hydrodynamic lubrication is a condition termed mixed lubrication because it has features of both the boundary and hydrodynamic conditions. Modern engineers employ what they call the ZN/P curve to describe the changes in friction that occur as a journal picks up speed; Z is the viscosity of the lubricant, N the number of revolutions of the journal per second and P the pressure in terms of load per unit area of the bearing [*see bottom illustra-*



RAILROAD JOURNAL BEARING patented by W. Bridges Adams in 1847 represented a significant advance in bearing technology. The bearing is depicted at left as seen looking down the length of the car and at right as seen from the end of the axle. The axle is the journal. The bearing is above the journal, since the load on it comes from the weight of the car. This replaceable bearing, made of brass, was an innovation introduced by Adams. The journal turned in a bath of oil (*color*). If the oil level got a little low, wood transfer rolls that floated on the oil maintained the lubrication of the journal. The oil holes on the top were a carry-over of an older design.

Ford announces the new Pinto MPG and Mustang II MPG

34 mpg highway

$ Pinto: $2,769

Official test results certified by the U.S. Government Environmental Protection Agency, which rates all cars using a common dynamometer testing procedure. Competitive fuel economy results based on EPA Buyer's Guide. Mileage may vary depending on your driving habits. **City mileage 23 mpg.**

See boxes on following pages for MPG equipment.

Base sticker price of Pinto MPG 2-Door Sedan excluding title, taxes, destination and dealer prep. Price comparisons based on sticker prices. Because destination charges are extra on all cars, and dealer prep is extra on all cars except GM cars and Toyota, the price difference may vary in some areas.

That's better than VW Beetle.
That's less than VW Rabbit.
That's better than Toyota Corona.
That's better than Datsun 710.
That's the same as Audi Fox.
That's better than Opel 1900.
That's better than Mazda 808.
That's better than a Monza Towne Coupe.
That's better than many others.

$224 less than VW Beetle.
$555 less than VW Rabbit.
$904 less than Toyota Corona.
$694 less than Datsun 710.
$2,075 less than Audi Fox.
$870 less than Opel 1900.
$222 less than Mazda 808.
$795 less than Monza Towne Coupe.
And less than many others.

Take your pick of Ford's six new high-mileage models on the next 3 pages.

New Ford Pinto MPG

34mpg. $2,769

Base sticker price, excluding title, taxes, destination and dealer prep.

Official U.S. Government Environmental Protection Agency tests. 34mpg highway, 23mpg city.

The country's best-selling sub-compact economy car line now has a new model with higher mileage at a lower price than the leading foreign car.

Pinto MPG 2-Door Sedan with optional WSW tires ($30)

New Pinto MPG with automatic transmission, 30mpg highway, 21mpg city, in official U.S. Government Environmental Protection Agency tests. It's the lowest sticker-priced 30mpg automatic you can buy.

See MPG equipment on next page.

MPG These three letters can change your mind about looking to the imports for good mileage. When you see them on our newest version of America's best-selling sub-compact, you'll know you're looking at a car that beats even the VW Beetle in price as well as EPA test mileage.

You can buy the new Pinto MPG now—at no increase in price—and get the same kind of standard equipment that makes the regular Pinto so popular: rack and pinion steering, 4-speed manual with floor-mounted stick, overhead cam 2.3-liter 4-cylinder engine with solid state ignition, front disc brakes, and more. Best of all, Pinto MPG comes with the same type of Lifeguard Design Safety Features found in our full-size Fords including side door beams, protective bumpers, and many others. (If there's one thing more important than better mileage, it's peace of mind when you're driving.)

Pinto MPG High mileage with wide choice of models.

Pinto MPG 3-Door Runabout (above and below), with optional Exterior Decor Group, deluxe bumper group and WSW tires.

MPG The new Pinto MPG is available in three models, including the popular 3-Door Runabout. That handy rear door flips up and the rear seat flips down to give you a five-foot-long carpeted load floor. The Runabout also includes such standard Pinto features as all-vinyl front bucket seats, and a mini-console. You can have the flexibility of the Runabout—and EPA highway test economy of 34 miles a gallon (city 23 mpg)—if you see your Ford Dealer now and order a new Pinto MPG.

Pinto MPG Wagon with Squire Option and WSW tires.

Priced lower than any foreign wagon. 34mpg highway, official U.S. Government Environmental Protection Agency tests.

MPG This symbol is your tip-off to a wagon with something you don't usually buy a wagon for: good mileage. The Pinto MPG Wagon comes with all the standard Pinto equipment—and is EPA rated at 34 miles to a gallon on the highway, 23 in the city. And you can get it at no increase in price. Pinto Wagon already outsells every other wagon in this country—big or small. If you want good mileage in a wagon, see your Ford Dealer fast.

MPG equipment:
Pinto MPG comes with a 2.3-liter 2V 4-cylinder engine, 4-speed manual (or optional automatic transmission, see box on facing page), a 3.18 axle ratio and catalytic converter. This is the equipment that produced these high results in the EPA tests.

Pinto MPG Wagon production started June 9.

Left, Mustang II MPG Hardtop • Right, Mustang II MPG 2+2

Better mileage than any small luxury car. Lower priced than any foreign competitor.

MPG This insignia tells you America's best-selling small luxury car now gives you the luxury of increased mileage. Official EPA highway tests got 34 miles to a gallon (23 mpg city) with a 4-speed manual transmission. Even with automatic transmission, the new Mustang II MPG got better mileage than many imports. Along with good mileage you get the standard Mustang II equipment: tachometer, rack and pinion steering, steel-belted radials, floor-mounted 4-speed stick—and the same low Mustang II price. Order now for quick delivery.

Mustang II MPG Ghia. Opera windows, vinyl roof and bodyside moldings are just some of the elegant touches that are standard on Ghia and make it such a distinctive small luxury car.

Mustang II MPG $3,529*

$2,748 less than Datsun 280Z
$158 less than Toyota Celica
$209 less than Opel Manta
$974 less than VW Dasher
$1,413 less than VW Scirocco

And $319 less than Monza Towne Coupe with its 5-speed option, which is required to achieve its best mileage results.

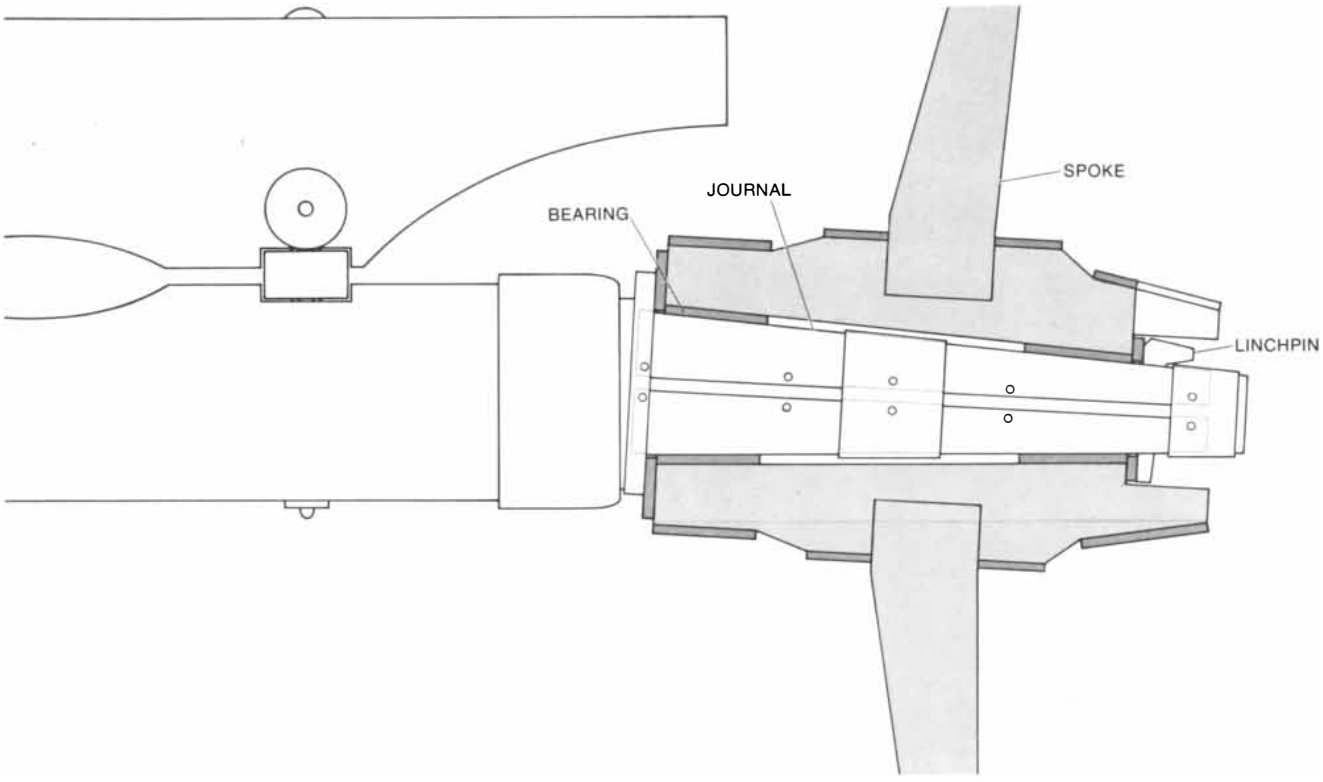
MPG equipment:
Mustang II MPG equipped with 2.3-liter 2V 4-cylinder engine, 4-speed manual (or optional automatic) transmission, a 3.18 axle ratio and catalytic converter.

*Base sticker price of Mustang II MPG Hardtop excluding title, taxes, destination and dealer prep. Price comparisons based on sticker prices. Because destination charges are extra on all cars, and dealer prep is extra on all cars except GM and Toyota cars, the price difference may vary in some areas.

Look close and compare. Ford means value. And your local Ford Dealer can show you.

FORD

FORD DIVISION Ford



CONESTOGA WAGON provides an example of bearings turning in a fixed journal. The commoner arrangement is for the journal to turn in a fixed bearing. The axle was made of wood, but the wrought-iron bearings in the wheel hub turned on wrought-iron strips that formed the journal and were affixed to the axle with metal bands. The wagon wheel was dished for rigidity and resistance to lateral thrust loads, which were caused by tilting of the wagon on uneven surfaces. To carry the thrust loads the bearings were made conical. The lower surface of the axle cone was horizontal so that the wheel could carry the wagon's weight and run true on the axle. As a result the axis of the wheel was inclined downward so that the lowest spoke would be vertical. Lubricant was pine tar and lard.

tion on page 60]. Viscosity is manifested by the resistance of a fluid to flow. (Fluidity is the reciprocal of viscosity.) Viscosity is one of the most important factors in the lubrication of bearings. When the fluid is moving, its mass also becomes important; accordingly engineers calculate kinematic viscosity, which is the fluid's viscosity divided by the fluid's density.

The Evolution of Materials

Until about 1800 bearings were made of materials that were readily available: stone, bone, wood, copper, iron and leather. The notion of creating a material for its suitability as a bearing is relatively new. It resulted in part from the recognition, based on experience, that a journal and a bearing should rarely be made of the same material; when they were, performance could be poor and wear could be severe. Another factor in the evolution of new materials was the application of the concept of making alloys to combine the best properties of various materials.

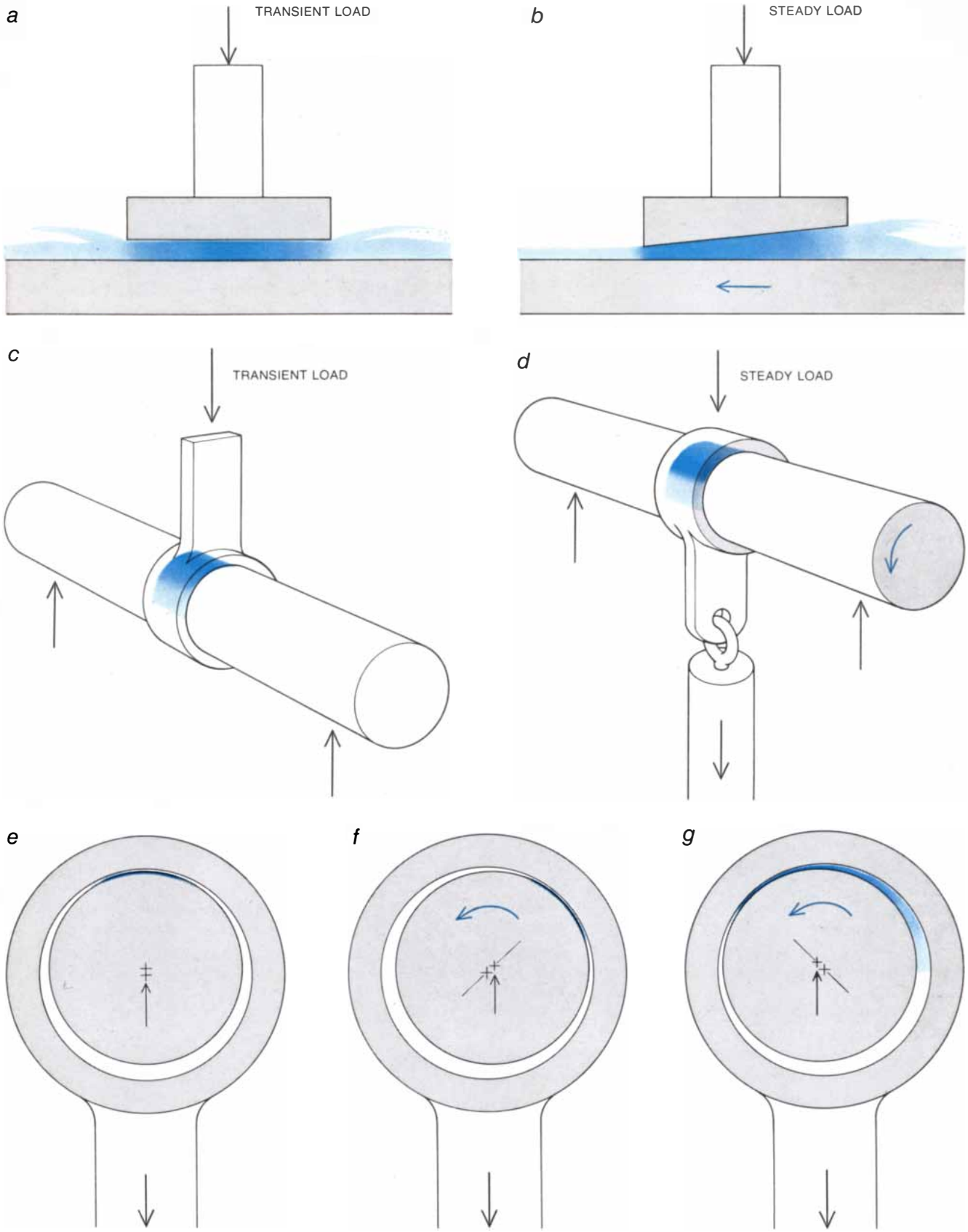
The first American patent on a special bearing material was issued in 1839 to Isaac Babbitt of Massachusetts. His scheme was to put a soft metal lining (such as a composite nominally of 89 percent tin, 9 percent antimony and 2 percent copper) in a harder and stronger shell. Such a lining is known today as a babbitt. The modern babbitt is usually an alloy based on lead and is often called white metal.

The concept of hardening lead by adding antimony was put into practice in the U.S. about 30 years after Babbitt's invention. A typical hardened antimonial lead lining was placed in a shell of hard bronze. In 1870 Alexander Dick of England developed a copper-lead alloy (nominally 80 percent copper, 10 percent lead and 10 percent tin) that served for years as the standard material for the journal bearings of railroad cars. Toward the end of the century C. B. Dudley of the Pennsylvania Railroad improved that alloy by increasing the content of lead to 15 percent. It was not easy: lead and copper are mutually insoluble. Alloys of lead and bronze are difficult to make for the same reason, and so it was not until 1900 that alloys with a lead content of 20 percent could be made. Lead improves the antifriction properties of a bearing but is too soft to serve by itself. The trick therefore is to achieve a high content of lead without impairing the overall strength of the bearing.

In the early part of the 20th century numerous materials were developed for bearings. Bronze figured in many of them. Alloys of aluminum, zinc, magnesium and cadmium also appeared. Today the available materials range from soft rubber to hard ceramics.

Desirable Properties

It goes almost without saying that a bearing should minimize friction between moving parts; indeed, minimizing friction is second only to the bearing's function of enabling one part to move and to transfer a load in relation to another part. A measure of the success of modern bearings in dealing with friction is seen in the typical automotive engine. Total friction usually consumes less than 25 percent of the power of a V-8 engine at 3,000 revolutions per minute. The sliding of the pistons in their cylinders gives rise to about 60 percent of the total friction, whereas the journal bearings associated with the linkage between the pistons and the crankshaft account for only 12 percent of the total friction. In other



FILM OF LUBRICANT separates the journal and the bearing. It can be generated in several ways. A squeeze film results from direct pressure of one surface against the other (*a*). The fluid's resistance to flow develops the pressure to support the transient load. A hydrodynamic film is generated (*b*) when lubricant is pulled into the interface by movement of one surface with respect to the other. The continuously generated film supports the steady load. In a journal bearing (*c*) a squeeze film supports the load. If the load is reversed, the squeeze film develops on the other side. The phenomenon is repeated every time the load reverses. If the journal is rotated (*d*), a hydrodynamic film is generated. Gradually (*e–g*) the movement of the journal pressurizes the wedge with oil. As the journal continues to turn, the wedge of oil moves the center of the journal and therefore also moves the minimum thickness of oil around the bearing.

words, the bearings consume about 3 percent of the total power of the engine.

Several other properties are required of bearings. One is economic: they must not be inordinately expensive, since a typical machine may incorporate so many of them. The other properties are mechanical; six of them in particular are of concern to the engineer and the designer.

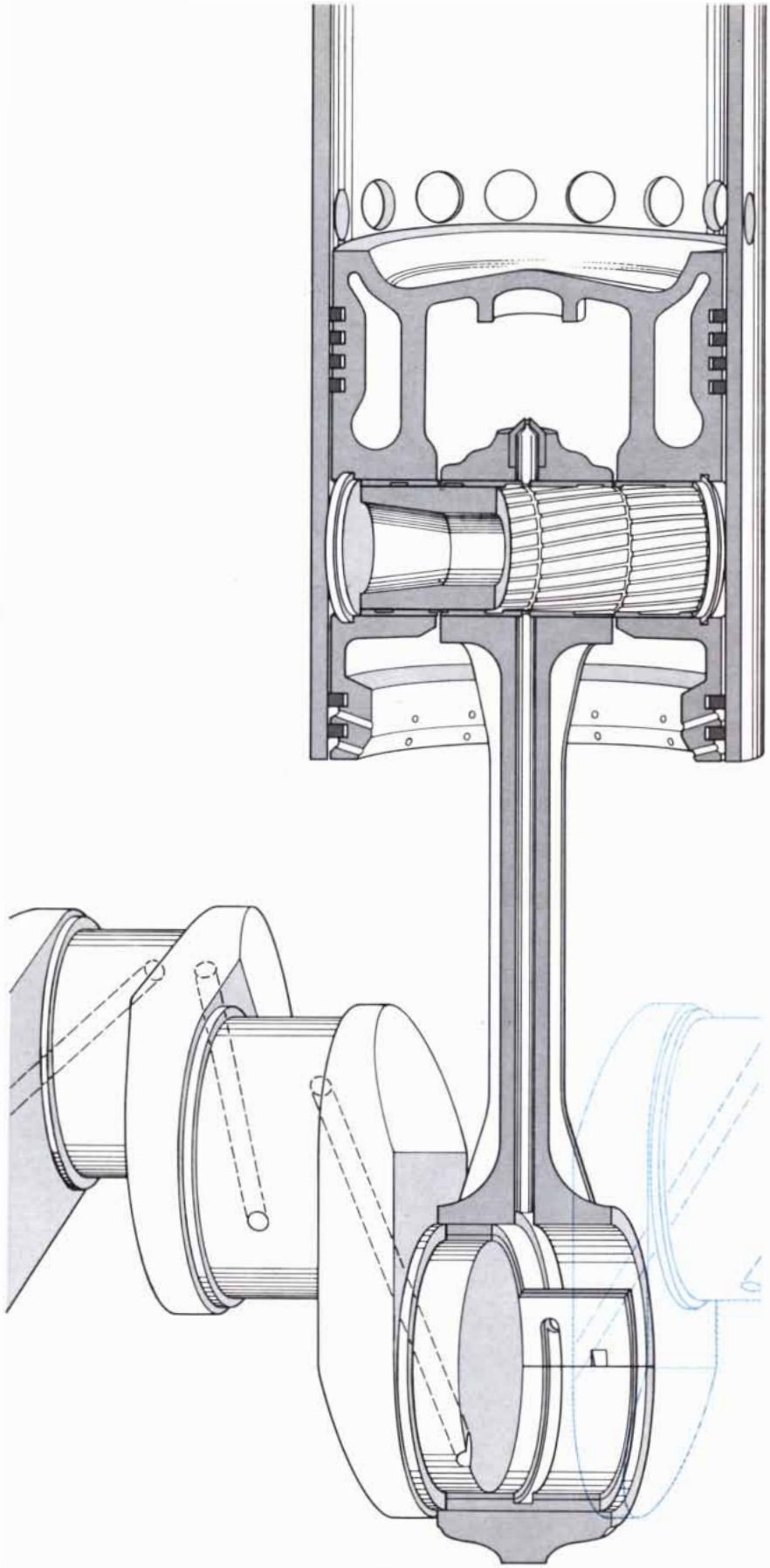
First, a bearing must be resistant to scoring, which arises in part from adhesion. No matter how carefully a journal and a bearing are machined and lubricated, the two surfaces do touch occasionally. At such times they suffer wear, which can cause undesirable changes in the dimensions of the surfaces.

The mechanism of wear can take four forms: adhesion, abrasion, corrosion and fatigue. Adhesive wear is the result of poor resistance to scoring, so that at the points of contact between the two surfaces, where the small peaks on each side tend to adhere to one another, minute bits of metal are ripped away as the journal turns. Abrasive wear arises when the surface of the journal is too rough, so that it abrades the surface of the bearing as sandpaper abrades wood. Hard particles can also abrade both surfaces. Corrosive wear is basically a chemical process, but the mechanical working in the bearing contributes by continually exposing fresh surfaces to the chemical action. Fatigue wear is the removal of material from a surface that has become fatigued.

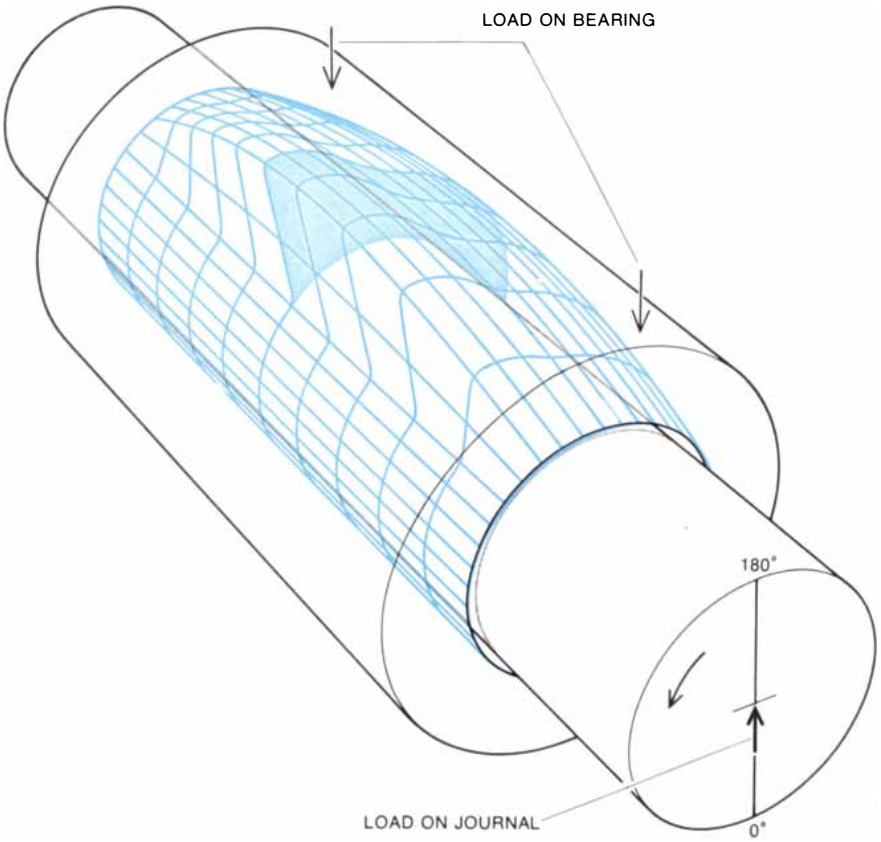
Several mechanical means of testing a bearing for this scoring property and of rating various combinations of materials for their resistance to scoring have been developed. Significant work on this aspect of bearing materials has been done by E. H. Scott and E. R. Booser of the General Electric Company. A good deal of work was also done by my predecessors in the Research Laboratories of the General Motors Corporation, A. E. Roach, C. L. Goodzeit and Richard P. Hunnicutt.

A second desirable property of a journal bearing is conformability. It is needed because a journal and a bearing may be misaligned when they are assembled or may be deflected during operation. The material from which the bearing is made must be capable of undergoing limited wear or of deforming plastically so that it conforms to the journal. Conformability can sometimes be regarded as a composite property including score resistance and strength.

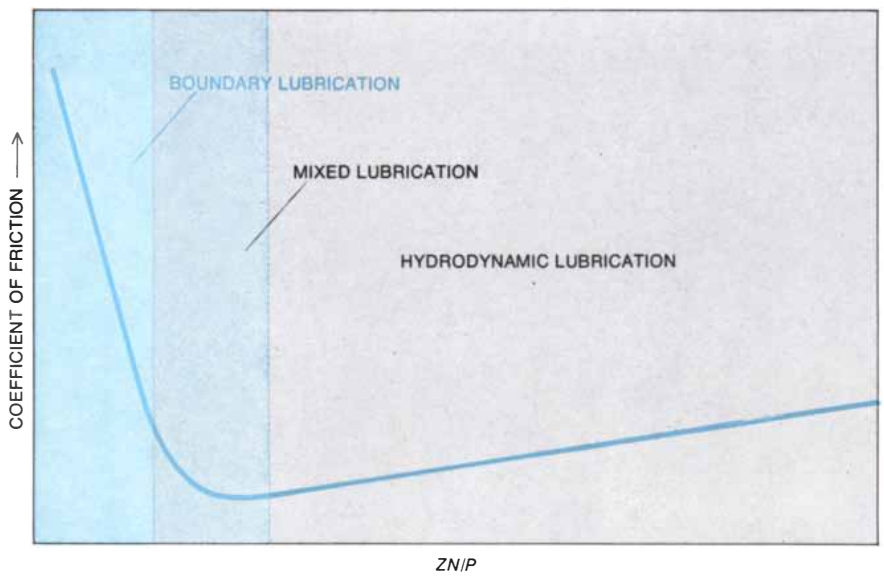
For various reasons, including wear and operation in a dusty environment, foreign particles may enter the clear-



CONNECTING ROD in an engine represents an arrangement in which both the journal and the bearing move. Here a connecting rod is depicted in a diesel engine. The piston pin is the journal in the upper bearing. A segment of the crankshaft is the journal in lower bearing.



DISTRIBUTION OF PRESSURE of the lubricant in a rotating journal bearing is depicted by means of a three-dimensional graph. In the region of highest pressure the film of lubricant is thinnest and carries the bulk of the load, as was observed by Beauchamp Tower in 1883 and analyzed more recently by Donald F. Hays, K. P. Oh, S. M. Rohde and K. H. Huebner of the General Motors Corporation and by John F. Booker of Cornell University.



CONDITIONS OF LUBRICATION in a journal bearing are portrayed in terms of the friction that exists as the journal picks up speed. Friction is high when the journal is moving slowly. The friction results from boundary lubrication, meaning that a small amount of lubricant is present but the journal and the bearing are rubbing on surface layers. Hydrodynamic lubrication exists when a complete film of lubricant separates the journal and the bearing. Mixed lubrication combines features of boundary and hydrodynamic lubrication. The relation of friction and lubrication is reflected by the formula ZN/P, in which Z is viscosity of lubricant, N the number of revolutions of journal per second and P the pressure.

ance between the journal and the bearing. If they remain in that clearance, they will have an abrasive effect. It is therefore desirable for the bearing material to have the property either of enabling the particles to be pushed into the bearing material itself or of allowing the particles to be partially embedded and then washed out a short time later.

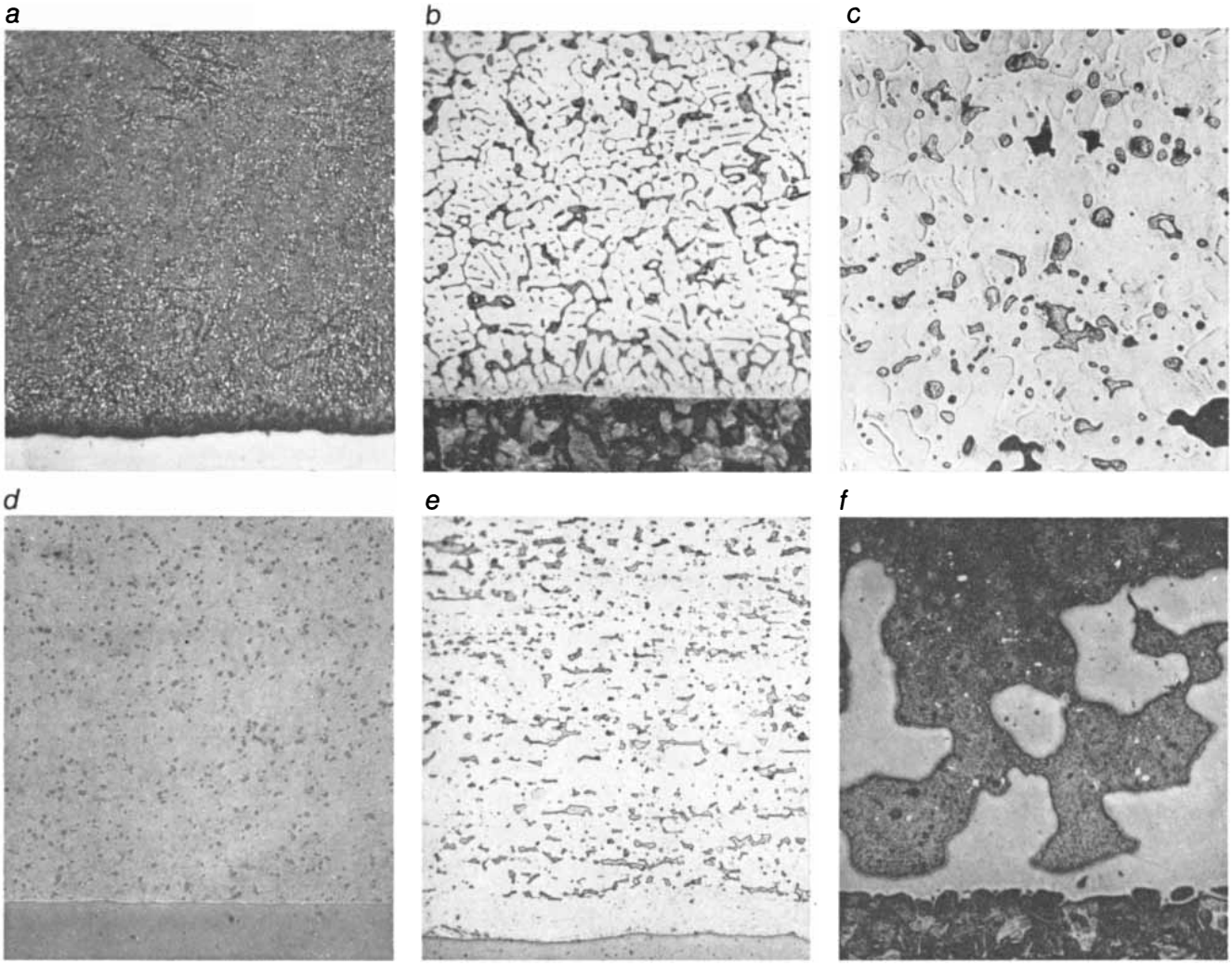
That is the property of embedability. It is essential in a journal bearing because of the number and variety of foreign particles that can enter the bearing. I have mentioned wear and dirt as sources of particles; in addition particles can be residual sand from cast parts such as the engine block and machining chips from holes drilled in the block as channels for the supply of oil. Whatever the source, the particles are of a variety of sizes, shapes and materials.

The Property of Embedability

Embedability was first recognized by S. A. McKee of the National Bureau of Standards in 1927. Much further work on the subject has been done at the General Motors Research Laboratories; embedability was defined by H. C. Mougey in 1936, the first model or hypothesis was developed by Roach in 1951, and I have extended and refined Roach's model. In the classical model the softer white metals easily allowed particles to be embedded, whereas the harder bronzes were less tolerant of particles and so gave rise to a good deal of wear. That wear could be reduced by making the journal harder.

In the course of our work my assistant L. W. Handwerker, Jr., and I developed an apparatus for mechanically testing embedability. The apparatus slowly and steadily injected particles into the clearance between the journal and the bearing. The test could be continued for as long as four hours, compared with a maximum of about five minutes for previous testing procedures, so that we were able to study the mechanism of embedment in greater detail. We monitored the rate of injection of particles and their concentration with a radioactive tracer.

An important factor in embedability is the adhesive resistance of the bearing material. A material with poor resistance to adhesion and scoring will cause the abrasive contaminant to stick to the surface of the bearing. If the bearing material is quite soft, the particle will not present a significant problem, but if the material is at an intermediate level of hardness, the particle will act like a machining tool and will abrade the sur-



BEARING MATERIALS are shown at an enlargement of 200 diameters. Babbitt, or white metal (*a*), is a lead-base alloy that has good antifriction properties but usually must be put on a backing of stronger material. Here the darker material is babbitt and the lighter material at bottom is a steel backing. Cast copper-lead (*b*) has three layers, consisting of steel at the bottom, then a thin overlay of babbitt. Leaded tin-bronze (*c*) is shown here in a bushing, so that it has no steel backing or babbitt overlay. Aluminum-silicon (*d*) has steel at the bottom and then aluminum-silicon. A babbitt overlay on this bearing does not appear. Aluminum-tin (*e*) shows a layer of steel, then a thin layer of pure aluminum, which aids in binding the bearing material to the steel, and finally the aluminum-tin. In a sintered copper-nickel bearing (*f*) the base material is steel. Above it is a layer made from a copper-nickel powder that was sintered onto the steel. Pores in the material were filled with babbitt, which also appears as an overlay at the top. The babbitt is the darker material.

face of the journal. Materials of higher strength and intermediate hardness must therefore have good resistance to adhesion to be successful.

In a four-stroke-cycle automotive engine the journal bearing is dynamically loaded. During the intake stroke the piston moves downward, pulling a mixture of fuel and air into the cylinder above it. In the compression stroke it moves upward, compressing the mixture. In the power stroke the mixture is ignited and the piston is thrust downward. In the exhaust stroke the piston moves upward, pushing the combustion gases out of the cylinder. For every four strokes the load on the bearing is reversed four times, so that it is being borne first by one side of the bearing and then by the other. This reverse loading enables the bearing to clean itself—on the unloaded side—of any partly embedded particles. The bearing's embedability is therefore improved over the level envisioned in the classical theory, which considered only unidirectional loading.

One might wonder what would happen if a bearing became fully embedded with foreign particles, that is, if the abrasive forms a continuous layer in the bearing surface. The performance of the bearing would be impaired, since the surface would tend to take on the characteristics of the abrasive and would effectively become much harder. Although the harder surface might have an improved resistance to adhesion, the abrasive quality of it would increase the wear of the journal. Particles that approximate the size of the bearing clearance will cause the greatest amount of damage. Smaller particles wash out more easily and larger particles have difficulty entering the clearance. Larger particles, however, may break down into smaller ones to cause damage. These problems are unlikely to arise in the lifetime of a typical journal bearing in an automotive engine provided that the oil and the oil filter are changed as recommended.

Other Properties

A fourth property that one looks for in a journal bearing is suitable strength. The four-stroke-cycle automotive engine is again illustrative. Two of the loads on a journal bearing are particularly high: the combustion load, which occurs dur-

ing the power stroke, and the primary inertial loads, which occur at the end of every stroke because the piston's movement in one direction must be stopped so that the direction can be reversed. The yield and fatigue strength of the bearing material must be high enough to support the combustion and inertial loads. What happens in fatigue is that cracks develop in the surface of a bearing, loosening small pieces of metal. These "tiles" increase in number and are gradually washed out, leaving a surface that is inadequate to carry the load. On the other hand, if a bearing material is overdesigned in strength, conformability and embedability can become problems.

Resistance to corrosion is a fifth desirable property in a bearing. A lubricating oil gradually becomes acidic as a result of oxidation. Certain mineral oils develop a complex organic acid that may attack metals in bearings such as lead and cadmium. In modern engines, as a result of improvements in the metallurgy of bearings and in the compounding of oil, corrosion is a minor problem, again provided that the oil is changed at the recommended intervals.

Finally, the thermal characteristics of a bearing are sometimes important. The lubricant carries away most of the heat that is generated as the journal and the bearing interact, but when both the load and the speed are high, the journal and the bearing must also conduct heat, particularly during conditions of boundary lubrication. Engineers calculate the recommended load and speed limits of a material by a *PV* factor, in which *P* is the unit load and *V* is the sliding velocity. If the coefficient of friction is assumed to be constant, the *PV* factor gives an upper limit for the amount of heat that will be generated per unit area of bearing surface. Bearing materials that are low in friction or high in thermal conductivity generally have a higher *PV* factor.

Modern Bearing Materials

The material of a bearing should be softer than the material of the journal, but on rare occasions the designer may modify this principle. A bearing that must carry a high load, for example, is likely to be made with a thin overlay of a soft material (such as babbitt) on a backing material of greater strength. The effective strength of the thin overlay is improved by the proximity of the stronger backing material. An overlay capable of good performance is a mixture of lead, tin and copper codeposited by electroplating.

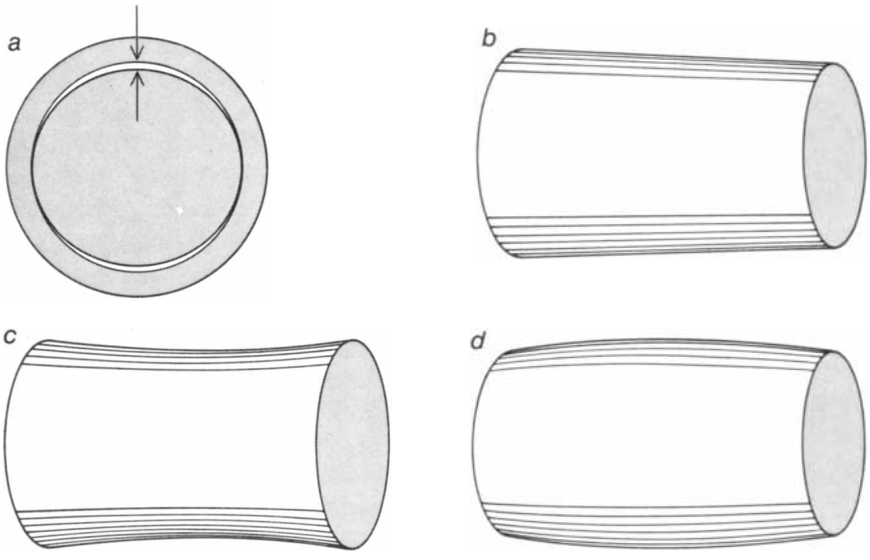
A bearing designed for a low or intermediate load can be made of either a metal or a nonmetallic material. I have mentioned white metals, which also are called babbitts and antifriction metals; they are alloys based on tin, cadmium or lead. The alloys with a lead base are the least expensive. Usually they are hardened with antimony (from 4 to 15 percent of the alloy) and made resistant to corrosion with tin (3 percent or more). White metals have excellent embedability and resistance to scoring. Their loading capacity can be raised by making them thinner.

Alloys of copper and lead also serve as bearing materials. The content of lead ranges from 10 to 50 percent. Such an alloy is either sintered or cast onto a backing that is usually steel. If the dispersion is uniform, a copper-lead alloy will have a fairly high resistance to fatigue. The resistance diminishes as the lead content is increased, but the properties of embedability and resistance to scoring are improved. In the journal bearings of an automotive engine or a diesel engine a copper-lead alloy is likely to be given a thin overlay of babbitt. A concern with copper-lead bearings is corrosion, since the tin in the babbitt tends to diffuse into the copper, increasing the possibility that the lead in the bearing will be corroded by acids in the lubricating oil. One means of dealing with this problem is to install a thin barrier of nickel separating the babbitt from the copper. The tin migrates to the nickel, but if enough of it is put into the babbitt, the nickel will become highly concentrated with tin and slow its diffusion before the babbitt has lost the amount of tin that would make corrosion a problem.

Bronze is another material widely employed in bearings. Bronze is an alloy containing copper and one or more of the metals tin, phosphorus, zinc and lead. Tin-bronze bushings are common in machines that run at intermediate speeds under intermediate loads.

Several kinds of aluminum alloy can serve as a bearing material. One such material is an alloy of aluminum, silicon (from 3 to 5 percent) and cadmium (1 percent). Another is aluminum with tin (from 5 to 7 percent in a low-tin alloy and about 20 percent in a high-tin alloy) plus copper, nickel, silicon or magnesium. On a steel backing these alloys have good resistance to fatigue. Aluminum alloys resist corrosion but are of intermediate quality in embedability and resistance to scoring. Therefore they are often given an overlay of babbitt.

The first aluminum alloy developed in the U.S. for bearings resulted from the work of my predecessor Alfred W. Schluchter some 35 years ago. A recent development initiated by my associate Arnold O. DeHart has babbitt entrained within the aluminum alloy, so that no expensive overlay of babbitt is required. A new automotive bearing alloy that my assistant E. J. Shipek and I have developed is similar in construction to a commercial bearing material but is made with different materials. In both processes a steel back has a porous matrix sintered onto it, and then babbitt is cast into the pores. The standard matrix



SHAPE OF JOURNAL significantly affects the performance of the journal-bearing unit. Owing to wear or poor manufacture a journal can be imperfectly round (*a*), tapered (*b*), shaped like an hourglass (*c*) or like a barrel (*d*). Another problem is lobing, or undulations of the surface, which can disrupt the film of lubricant and impair bearing operation.

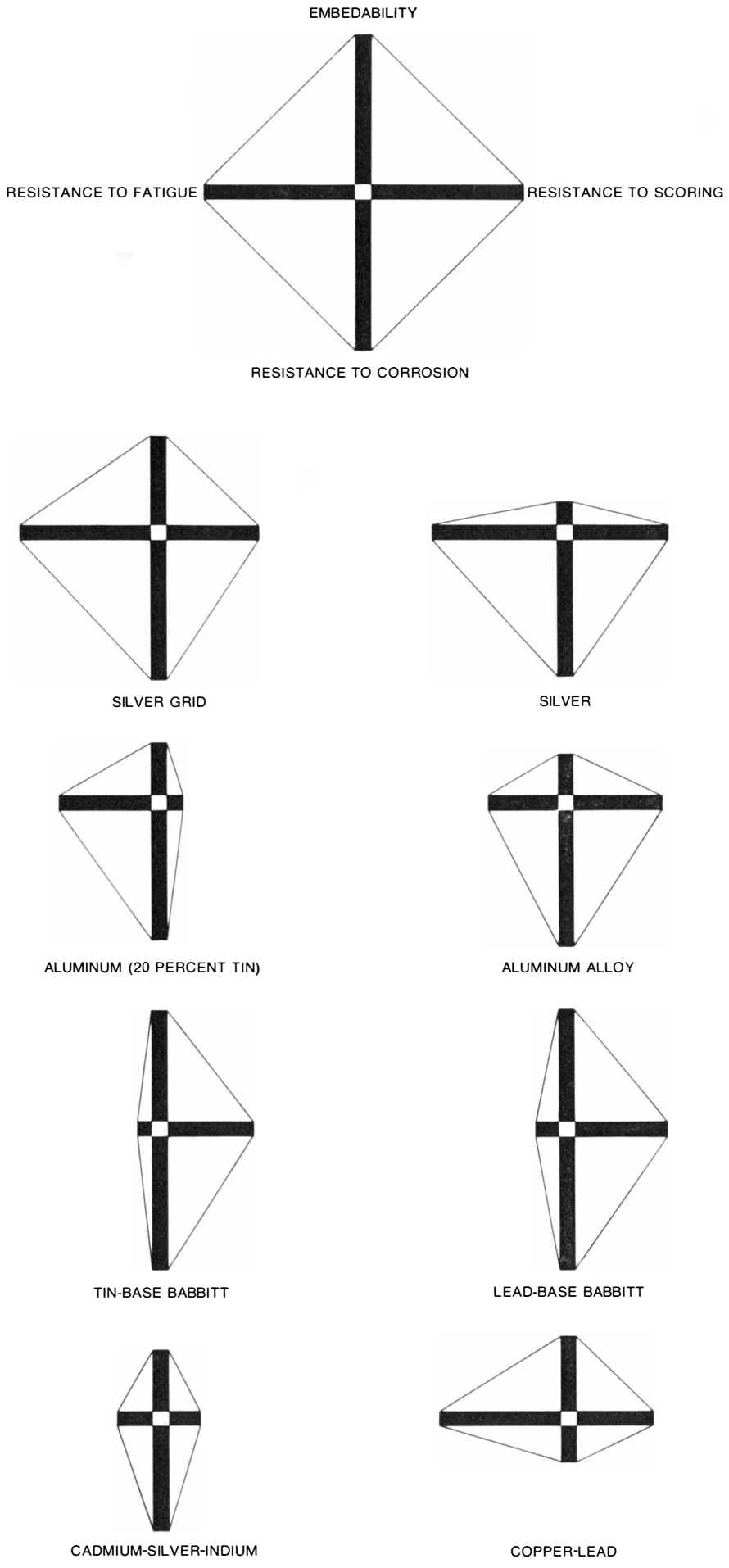
is 85 percent copper and 15 percent nickel. Our new matrix is considerably less expensive than copper and nickel. It can easily be made thinner and can operate without an additional overlay of babbitt, since in score resistance our matrix is much superior to the copper-nickel matrix.

One often hears of prelubricated bearings, which are described as having been "oiled for life." They are usually copper-based materials, sintered into a porous matrix that is filled or impregnated with a lubricant such as oil, graphite or molybdenum disulfide. Bearings of this type serve in small motors, in household appliances and in the front-suspension joint bearings (the ball joints) of automobiles.

Nonmetals such as thermoplastics, thermosets, polytetrafluoroethylene and carbon-graphites are successful bearing materials because of their excellent resistance to scoring and corrosion. Indeed, they resist scoring so well that most of them can function with little or no lubrication. The limiting capabilities for plastic and elastomeric materials are strength and thermal conductivity; they are being improved in these respects by means of fiber reinforcement and metallic filling. Ceramic materials have good compressive strength and can function well in a vacuum and at high temperatures.

Considerations of Design

The design of a journal bearing involves the bearing and the journal as a unit. The primary variables are load and speed. The designer must also take into account the conditions of lubrication. Among the accompanying illustrations [*cover and page 51*] are representations of a journal bearing with grooves. The grooves are needed to provide lubrication in a bearing that changes its direction of movement without undergoing reversal of load. An example is a bearing carrying a piston pin of a diesel engine. The diesel is a two-stroke-cycle engine. In the first stroke the piston is



PROPERTIES OF BEARINGS are set forth in a scheme where diamond shape at top represents an ideal bearing. Embedability, meaning the ability of a bearing surface to embed abrasive foreign particles and thus keep them from harming the journal and the bearing, and three other properties are represented. A silver grid bearing, for example, has good resistance to fatigue and corrosion but is poorer in embedability and resistance to scoring. The lead-base babbitt has good embedability and resistance to scoring.

forced upward, compressing air that has been blown into the cylinder. The load is carried by one half of the piston-pin bearing. At the top of the stroke fuel is injected into the cylinder. The compressed air is so hot that the fuel ignites and the piston is driven downward for the second stroke. The other half of the bearing is never really loaded because the piston is stopped and reversed between the first and second strokes by the pressure of the gas rather than by the bearing. The grooves establish a lubricating film, which is generally boundary lubrication or a mixture of boundary and hydrodynamic lubrication.

In a bearing that undergoes reversal of load, such as a piston-pin bearing of a four-stroke-cycle automotive engine, grooves would be inappropriate. The lubrication here (at the times of reversal) is primarily a film resulting from the squeezing effect of the load on the lubricant. Grooves would diminish the effect. The existence of a squeeze film was demonstrated experimentally in 1944 by Arthur F. Underwood of the General Motors Research Laboratories.

Another factor in design is the relation between the load on the bearing and the speed of operation. The ability of the bearing to carry a load without seizing is high at low speeds, drops to a low level at intermediate speeds and increases at higher speeds. At low speed the requirements introduced by the *PV* factor (unit load times sliding velocity) are low, and boundary lubrication takes place. The load that can be put on a properly designed boundary-lubricated bearing is relatively high.

As the speed increases, the requirements of the *PV* factor are raised. Hence the temperature goes up, the viscosity of the lubricant is reduced and the load capacity of the bearing declines. The trend continues until the speed is high enough to develop a significant amount of hydrodynamic lubrication. The result at this stage is a mixture of boundary and hydrodynamic lubrication. Here the load capacity of the bearing is at a minimum, but the presence of a certain amount of hydrodynamic lubrication contributes to cooling through flow of the lubricant. With a further increase in speed a properly designed hydrodynamically lubricated bearing will display an increasing load capacity.

The shape of the journal surface is important in the design of a bearing. The topography of the journal, meaning the detailed characteristics of the surface, must be considered in relation to the tolerance to be built into the bearing. A journal should have no lobing, a term that refers to small undulations of the surface. Excessive lobing will affect the bearing adversely, so that any lobes present should be small in number and in amplitude. Roughness, indentations, protuberances, holes and unintentional grooving, if they are not limited, will drastically reduce the load capacity of the bearing and increase wear.

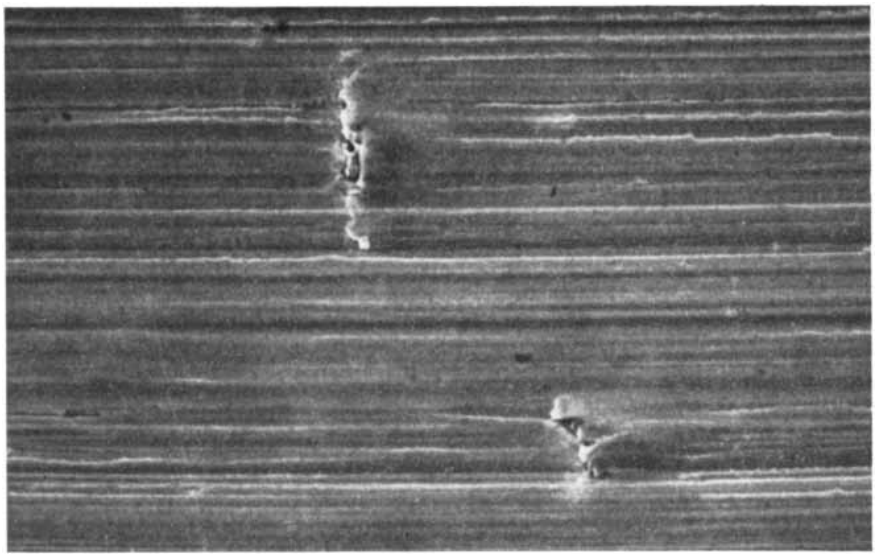
Aberrations in the geometry of a journal, introduced by wear and sometimes by poor manufacturing, can present serious problems. The aberrations include taper, out-of-roundness, an hourglass shape or a barrel shape. They result in an increase in the thickness of the lubricant film in some places and a loss of thickness of the film (sometimes sufficient to cause rubbing) in others.

Frequently a bearing has to be made in two parts in order to fit it around the journal. A crankshaft bearing is an example. Such a bearing is called a split bearing. The profile of the standard split bearing is eccentric, designed to make the lubricant support the load in the regions of high pressure and to increase the clearance near the split line. The eccentric design in the regions of high pressure improves the load capacity, and the increased clearance at the split line enables the bearing to tolerate misalignment. The regions of high pressure develop from the combustion load and the inertial load. A unique design developed by my colleague DeHart in collaboration with Duane H. Harwick of GM Delco Moraine extends the concentric crown area (in order to increase further the load capacity of the bearing) but still retains clearance (eccentricity) at the sides. The bearing is called Conecc, for concentric-eccentric design.

Another major consideration in design is the ratio of the length of the bearing to its diameter (L/D or $L/2R$, where R is the radius of the bearing). The usual ratio is from .5 : 1 to 1 : 1. At higher ratios (when the bearing is longer) a bearing can carry a bigger load, but geometry, alignment and limitation of space become problems. On the other hand, if the diameter of the unit (both the bearing and the journal) is reduced, thereby raising the ratio, the ability of the journal to carry its load without much deflection becomes a factor.

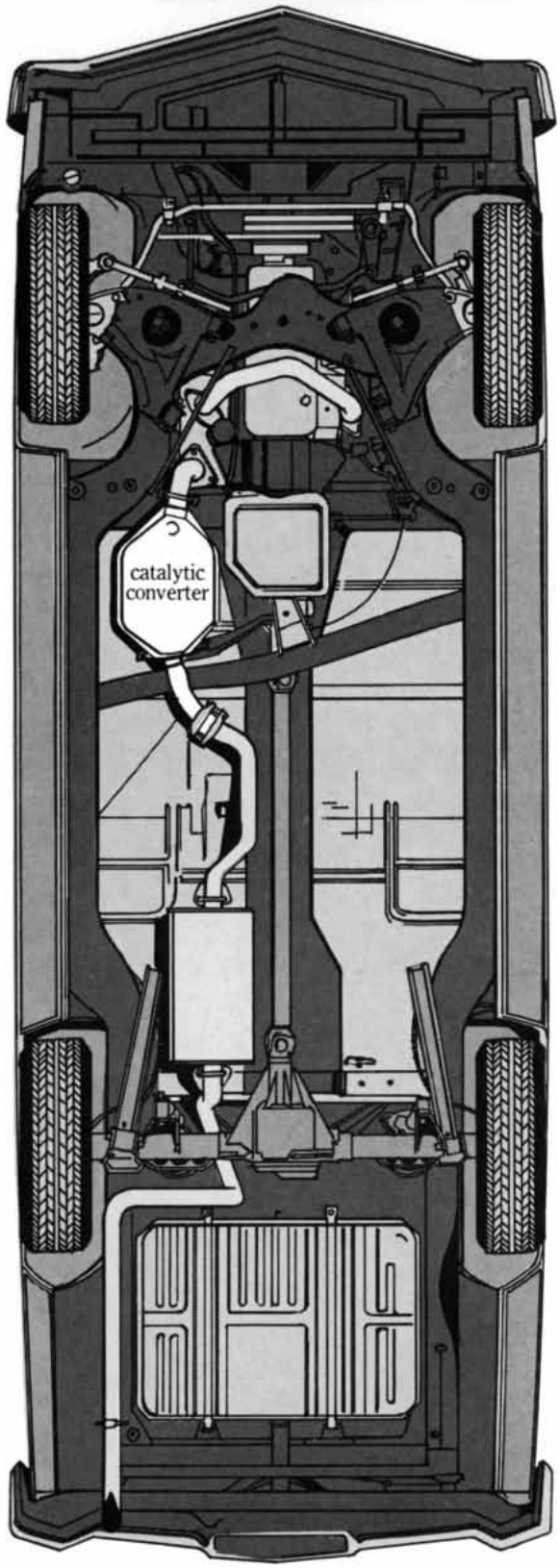
Finally, the designer of a bearing must take vibration into account. Vibration results from such instabilities as a journal or a shaft that is out of balance. Instabilities also develop from certain complex conditions of lubrication, speed, load and design. My colleague Dah-chen Sun has analyzed certain instabilities by means of porous journal bearings in which the lubricant, which is air, passes through the wall of the bearing and into the clearance between the bearing and the journal.

In sum, the technology of journal bearings has changed a great deal over the long period of time in which they have served, by whatever name, in a large variety of machines. Through international efforts in the fields of materials, lubrication and design the technology is still evolving.



SURFACE OF JOURNAL appears in a scanning electron micrograph at an enlargement of 750 diameters. Although the journal, which is part of an automobile crankshaft, has been ground to a high degree of smoothness, it cannot be perfectly smooth. The bearing must be able to come in contact with this surface occasionally without itself becoming scored or worn. The result is achieved by proper attention to lubrication and to the choice of bearing material. This journal is made of nodular iron, and a few nodules of graphite are visible.

Because of the catalytic converter, GM cars use less gasoline.



Primarily because of the catalytic converter, gas mileage on GM cars has been increased by 28% on a sales-weighted average, according to EPA figures.

The converter gives GM car owners the best of both worlds: emissions of carbon monoxide and hydrocarbons are cut by about 50% from the already lowered levels of 1974, and it is possible once more to tune engines for economy, drivability and performance.

Catalytic converters do add to the basic cost of a GM car. Part of that money goes for insulation that keeps the outer skin temperature of the converter in normal operation about the same as that of an ordinary muffler, and far lower than the temperature of the exhaust manifold.

But when you think of the cost, think of the reduction in fuel consumption over the life of that average GM car; and don't forget, the use of unleaded gas lowers maintenance costs by greatly increasing the life of spark plugs, engine oil and exhaust system components.

After more than a billion miles on the road, the GM catalytic converter has become a world standard in pollution control devices. GM has signed contracts to build converters for auto-makers in Europe and Asia, as well as other U.S. manufacturers.

You get the fuel-saving advantages of a catalytic converter as standard equipment on 1975 cars from General Motors, a world leader in automotive pollution control technology.

General Motors

Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, GMC Truck

Catalytic converter, standard equipment on 1975 model GM cars.

You know what Polaroid Land film can do. But you may not know how many ways there are to do it.

At Polaroid, we believe there are a variety of ways instant photography can be put to good use.

So we've developed a variety of films to suit practically any needs you might have.

Fine grain films. Positive/negative films. Luminous color films. Films that produce black and white, continuous-tone transparencies in 2 minutes. High-contrast line films. We even have a super fast film with a speed equivalent to 10,000 A.S.A.

And most recently we've added Polacolor 2– a superb new film, which like SX-70 Land film, has excellent color saturation and rendition, and the most

stable dyes in general photography.

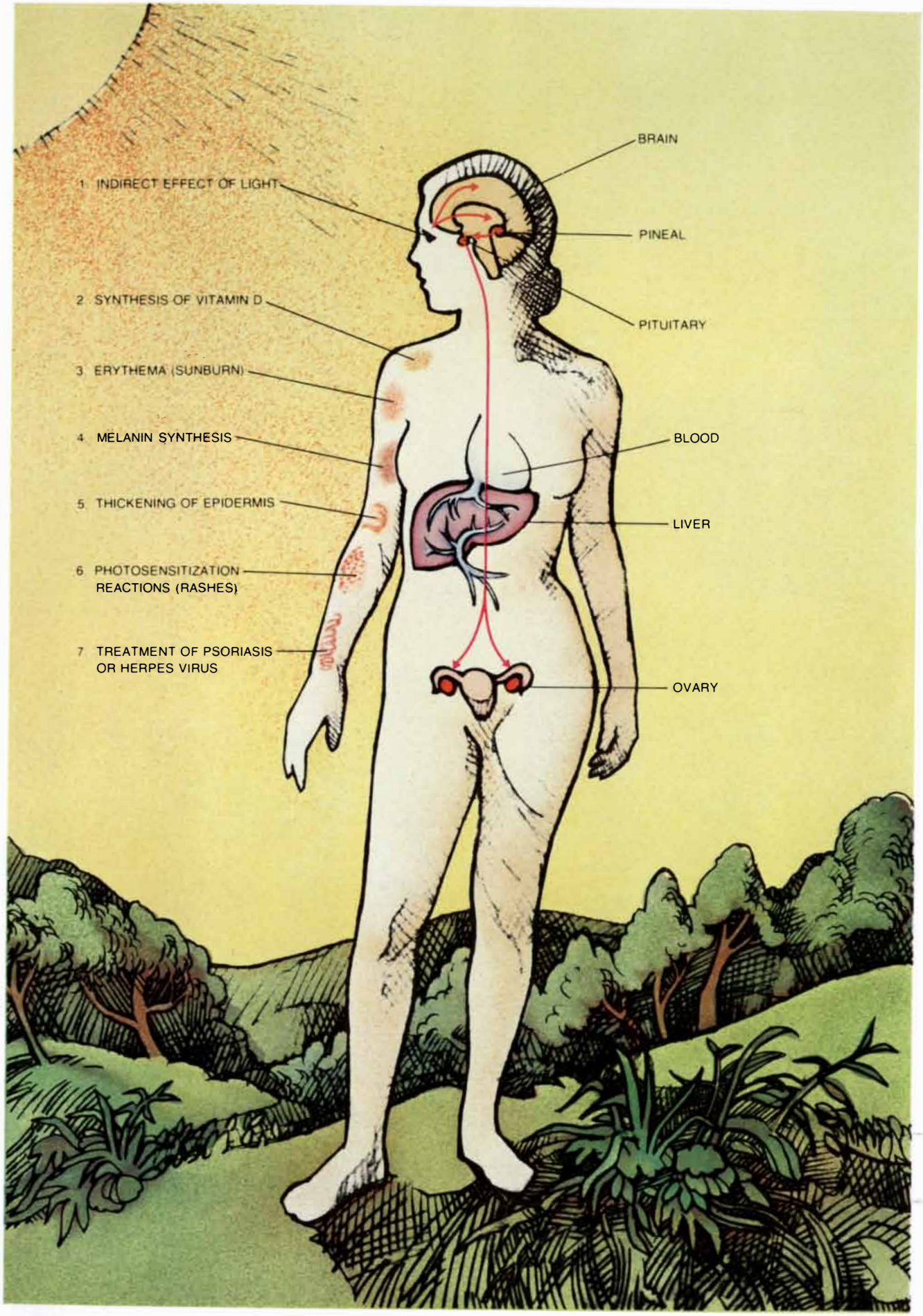
All of these Polaroid self-developing films are now available in sheet, pack or roll formats.

What's more, we have a complete line of cameras for these films, plus adapter backs that allow you to use most of them with your present equipment.

Why not talk to your photo dealer about your own particular needs. Or call us collect: (617) 547-5176.

Chances are there's a Polaroid film that can do the work for you in seconds. Instead of hours.

21 instant films from Polaroid



The Effects of Light on the Human Body

Sunlight tans skin, stimulates the formation of vitamin D and sets biological rhythms. Light is also used in the treatment of disease. Such effects now raise questions about the role of artificial light

by Richard J. Wurtman

Since life evolved under the influence of sunlight, it is not surprising that many animals, including man, have developed a variety of physiological responses to the spectral characteristics of solar radiation and to its daily and seasonal variations. With the coming of summer in the Northern Hemisphere millions of people living in the North Temperate Zone will take the opportunity to darken the shade of their skin, even at the risk of being painfully burned. Coincidentally the sunbathers will replenish their body's store of vitamin D, the vitamin that is essential for the proper metabolism of calcium. Skin-tanning and subcutaneous synthesis of vitamin D from its precursors, however, are only the best-known consequences of exposure to sunlight.

Investigators are slowly uncovering subtler physiological and biochemical responses of the human body to solar radiation or its artificial equivalent. Within the past few years, for example, light has been introduced as the standard method of treatment for neonatal jaundice, a sometimes fatal disease that is common among premature infants. More recently light, in conjunction with a sensitizing drug, has proved highly effective in the treatment of the common skin inflammation psoriasis. It seems safe to predict that other therapeutic uses for light will be found.

At least equally significant for human well-being is the growing evidence that fundamental biochemical and hormonal rhythms of the body are synchronized, directly or indirectly, by the daily cycle of light and dark. For example, my coworkers at the Massachusetts Institute of Technology and I have recently discovered a pronounced daily rhythm in the rate at which normal human subjects excrete melatonin, a hormone synthesized by the pineal organ of the brain. In experimental animals melatonin induces sleep, inhibits ovulation and modifies the secretion of other hormones. In man the amount of the adrenocortical hormone cortisol in the blood varies with a 24-hour rhythm. Although seasonal rhythms associated with changes in the length of the day have not yet been unequivocally demonstrated in human physiology, they are well known in other animals, and it would be surprising if they were absent in man. The findings already in hand suggest that light has an important influence on human health, and that our exposure to artificial light may have harmful effects of which we are not aware.

The wavelengths of radiation whose physiological effects I shall discuss here are essentially those supplied by the sun after its rays have been filtered by the atmosphere, including the tenuous high-altitude layer of ozone, which removes virtually all ultraviolet radiation with a wavelength shorter than 290 nanometers. The solar radiation that reaches the earth's surface consists chiefly of the ultraviolet (from 290 to 380 nanometers), the visible spectrum (from 380 to 770 nanometers) and the near infrared (from 770 to 1,000 nanometers). About 20 percent of the solar energy that reaches the earth has a wavelength longer than 1,000 nanometers.

The visible spectrum of natural sunlight at sea level is about the same as the spectrum of an ideal incandescent source radiating at a temperature of 5,600 degrees Kelvin (degrees Celsius above absolute zero). The solar spectrum is essentially continuous, lacking only certain narrow wavelengths absorbed by elements in the sun's atmosphere, and at midday it has a peak intensity in the blue-green region from 450 to 500 nanometers [*see upper illustration on next page*]. The amount of ultraviolet radiation that penetrates the atmosphere varies markedly with the season: in the northern third of the U.S. the total amount of erythemal (skin-inflaming) radiation that reaches the ground in December is only about a fifteenth of the amount present in June. Otherwise there is little seasonal change in the spectral composition of the sunlight reaching the ground. The actual number of daylight hours, of course, can vary greatly, depending on the season and the distance north or south of the Equator.

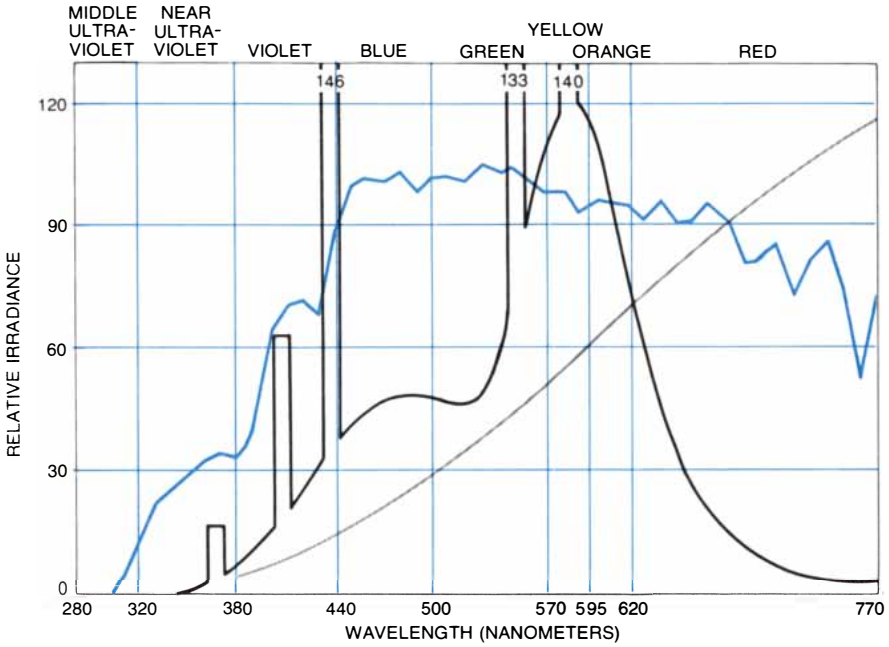
SOME DIRECT AND INDIRECT EFFECTS OF LIGHT on the human body are outlined in the drawing on the opposite page. Indirect effects include the production or entrainment (synchronization) of biological rhythms. Such effects are evidently mediated by photoreceptors in the eye (*1*) and involve the brain and neuroendocrine organs. For example, excretion of melatonin, a hormone produced by the pineal organ, follows a daily rhythm. In animals melatonin synthesis is regulated by light. The hormone, acting on the pituitary, plays a role in the maturation and the cyclic activity of the sex glands. Ultraviolet radiation acts on the skin to synthesize vitamin D (*2*). Erythema, or reddening of the skin (*3*), is caused by ultraviolet wavelengths between 290 and 320 nanometers. In response melanocytes increase their synthesis of melanin (*4*), a pigment that darkens the skin. Simultaneously the epidermis thickens (*5*), offering further protection. In some people the interaction of light with photosensitizers circulating in the blood causes a rash (*6*). In conjunction with selected photosensitizers light can be used to treat psoriasis and other skin disorders (*7*). In infants with neonatal jaundice light is also used therapeutically to lower the amount of bilirubin circulating in the blood until infant's liver is mature enough to excrete the substance. The therapy prevents the bilirubin from concentrating in the brain and destroying brain tissue.

The most familiar type of artificial light is the incandescent lamp, in which the radiant source is a hot filament of tungsten. The incandescent filament in a typical 100-watt lamp has a temperature of only about 2,850 degrees K., so that its radiation is strongly shifted to the red, or long-wavelength, end of the spectrum. Indeed, about 90 percent of the total emission of an incandescent lamp lies in the infrared.

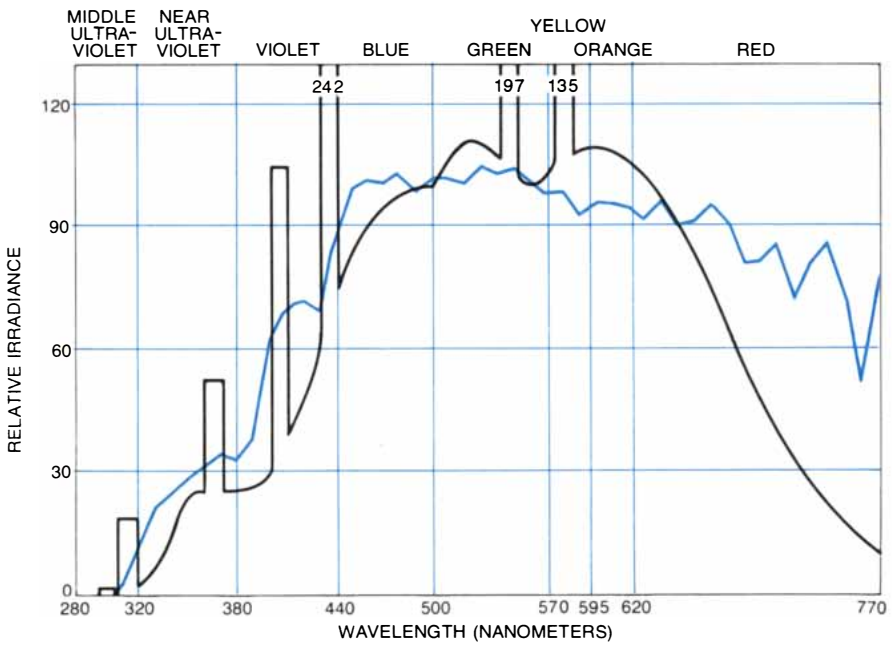
Fluorescent lamps, unlike the sun and incandescent lamps, generate visible light by a nonthermal mechanism. Within the glass tube of a fluorescent lamp ultraviolet photons are generated by a mercury-vapor arc; the inner surface of the tube is coated with phosphors, luminescent compounds that emit visible radiations of characteristic colors when they are bombarded with ultraviolet photons. The standard "cool white" fluorescent lamp has been designed to achieve maximum brightness for a given energy consumption. Brightness, of course, is a subjective phenomenon that depends on the response of the photoreceptive cells in the retina. Since the photoreceptors are most sensitive to yellow-green light of 555 nanometers, most fluorescent lamps are designed to concentrate much of their output in that wavelength region. It is possible, however, to make fluorescent lamps whose spectral output closely matches that of sunlight [*see lower illustration on this page*].

Since fluorescent lamps are the most widely used light source in offices, factories and schools, most people in industrial societies spend many of their waking hours bathed in light whose spectral characteristics differ markedly from those of sunlight. Architects and lighting engineers tend to assume that the only significant role of light is to provide adequate illumination for working and reading. The illumination provided at eye level in artificially lighted rooms is commonly from 50 to 100 footcandles, or less than 10 percent of the light normally available outdoors in the shade of a tree on a sunny day.

The decision that 100 footcandles or less is appropriate for indoor purposes seems to be based on economic and technological considerations rather than on any knowledge of man's biological needs. Fluorescent lamps could provide higher light intensities without excessive heat production, but the cost of the electric power needed for substantially higher light levels would probably be prohibitive. Nevertheless, the total amount of light to which a resident of Boston, say, is exposed in a conventionally lighted indoor environment for 16 hours a day is considerably less than would impinge on him if he spent a single hour each day outdoors. If future studies indicate that significant health benefits (for example better bone mineralization) might accrue from increasing the levels of indoor lighting, our society might, in a period of energy shortages, be faced with hard new choices.



SPECTRUM OF SUN at sea level (*color*) **is compared with the spectra of a typical incandescent lamp (*gray curve*) and of a standard "cool white" fluorescent lamp (*black curve*). The visible spectrum lies between the wavelengths of 380 and 770 nanometers. The peak of the sun's radiant energy falls in the blue-green region between 450 and 500 nanometers. Cool-white fluorescent lamps are notably deficient precisely where the sun's emission is strongest. Incandescent lamps are extremely weak in the entire blue-green half of the visible spectrum.**



BROAD-SPECTRUM FLUORESCENT LAMP known as Vita-Lite (*black curve*) closely approximates spectral characteristics of sunlight (*color*). Wavelengths emitted by fluorescent lamps can be adjusted by selecting phosphors with which inner surface of lamp is coated.

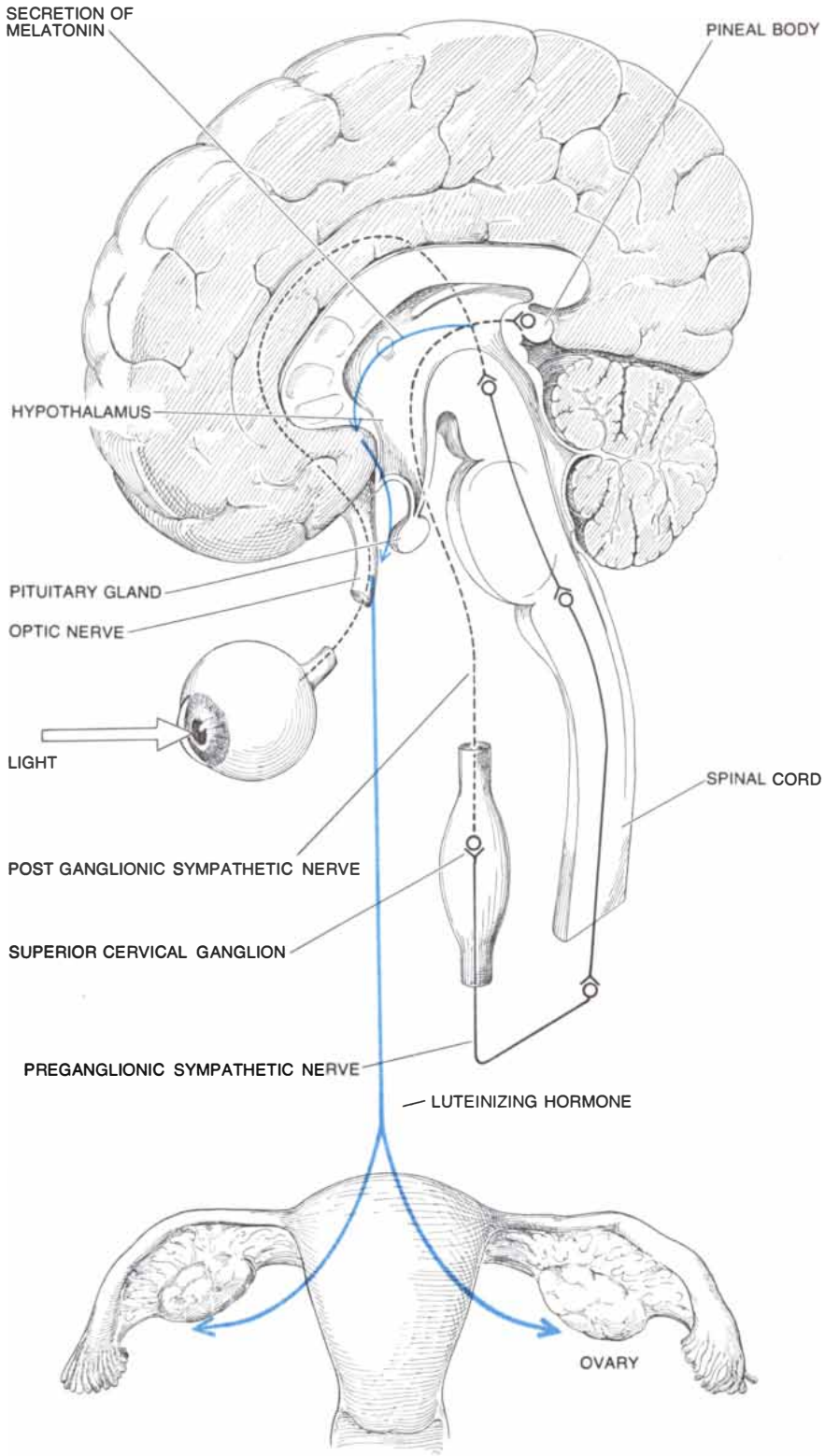
Each of the various effects of light on mammalian tissues can be classified as direct or indirect, depending on whether the immediate cause is a photochemical reaction within the tissue or a neural or neuroendocrine signal generat-

ed by a photoreceptor cell. When the effect is direct, the molecule that changes may or may not be the one that actually absorbs the photon. For example, certain molecules can act as photosensitizers: when they are raised to transient high-energy states by the absorption of radiation, they are able to catalyze the oxidation of numerous other compounds before they return to the ground state. Photosensitizers sometimes present in human tissues include constituents of foods and drugs and of toxins produced in excess by some diseases.

In order to prove that a particular chemical change in a tissue is a direct response to light one must show that light energy of the required wavelength does in fact penetrate the body to reach the affected tissue. In addition the photoenergetic and chemical characteristics of the reaction must be fully specified, first in the test tube, then in experimental animals or human beings, by charting the reaction's "action spectrum" (the relative effectiveness of different spectral bands in producing the reaction) and by identifying all its chemical intermediates and products. Visible light is apparently able to penetrate all mammalian tissues to a considerable depth; it has even been detected within the brain of a living sheep.

Ultraviolet radiation, which is far more energetic than visible wavelengths, penetrates tissues less effectively, so that erythemal radiations barely reach the capillaries in the skin. The identification of action spectra for the effects of light on entire organisms presents major technical problems: few action spectra have been defined for chemical responses in tissues other than the skin and the eyes.

The indirect responses of a tissue to light result not from the absorption of light within the tissue but from the actions of chemical signals liberated by neurons or the actions of chemical messengers (hormones) delivered by circulation of the blood. These signals in turn are ultimately the result of the same process as the one that initiates vision: the activation by light of specialized photoreceptive cells. The photoreceptor transduces the incident-light energy to a neural signal, which is then transmitted over neural, or combined neural-endocrine, pathways to the tissue in which the indirect effect is observed. For example, when young rats are kept continuously under light, photoreceptive cells in their retina release neurotransmitters that activate brain neurons; these neurons in turn transmit signals over complex neuroendocrine pathways that reach the anterior pituitary gland,



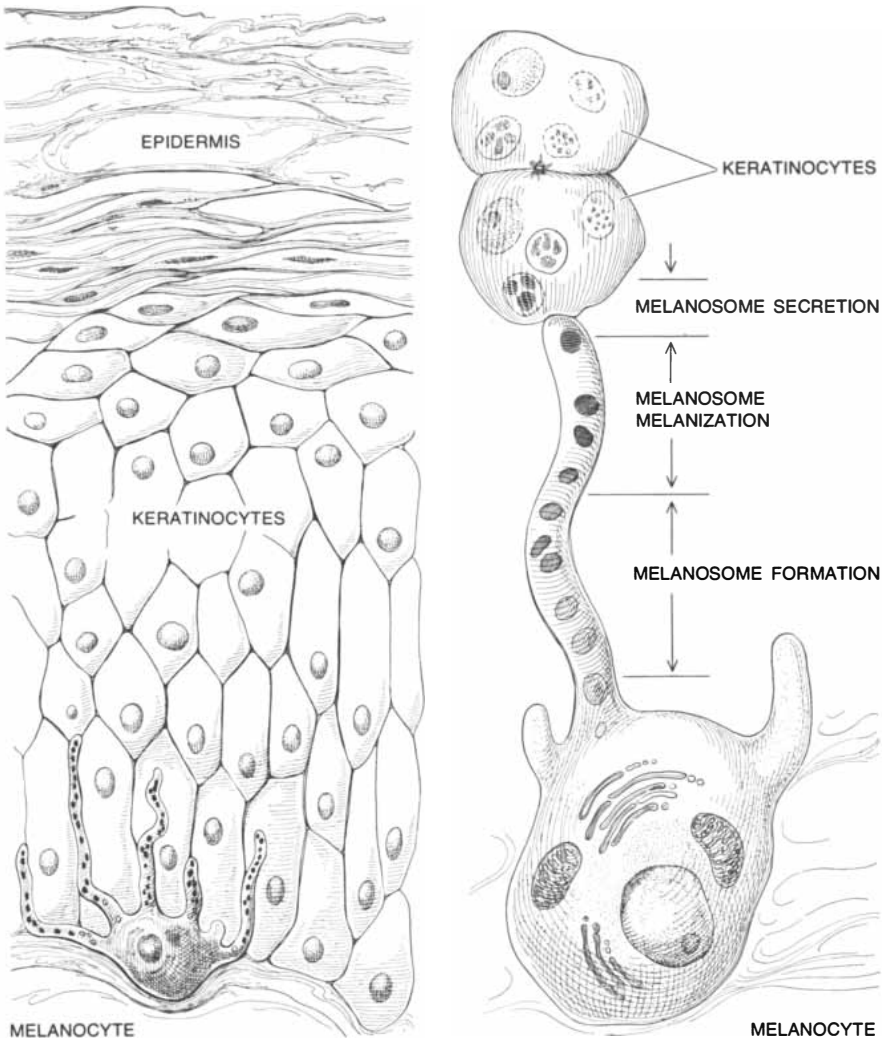
INDIRECT EFFECT OF LIGHT ON OVARIES OF RATS is shown schematically. Light activates receptors in the retina, giving rise to nerve impulses that travel via a chain of synapses through the brain, the brain stem and the spinal cord, ultimately decreasing the activity of neurons running to the superior cervical ganglion (in the neck) and of the sympathetic nerves that reenter the cranium and travel to the pineal organ. There the decrease in activity reduces both the synthesis and the secretion of melatonin. With less melatonin in blood or cerebrospinal fluid, less reaches brain centers (probably in hypothalamus) on which melatonin acts to suppress secretion of luteinizing hormone from anterior pituitary. Thus more hormone is released, facilitating ovarian growth and presumably ovulation.

where they stimulate the secretion of the gonadotropic hormones that accelerate the maturation of the ovaries [*see illustration on preceding page*].

That the ovaries are not responding directly to light can be shown by removing the eyes or the pituitary gland of the rat before exposing it to continuous light. After either procedure light no longer has any influence on ovarian growth or function. Various studies confirm that the effect of light on the ovaries is mediated by photoreceptive cells in the retina. It has not been possible to show, however, which of the photoreceptors in the eye release the neurotransmitters that ultimately affect the pituitary gland.

Natural sunlight acts directly on the cells of the skin and subcutaneous tissues to generate both pathological and protective responses. The most familiar example of a pathological response is sunburn; in susceptible individuals exposed over many years sunlight also causes a particular variety of skin cancer. The chief protective response is tanning. Ultraviolet wavelengths in the narrow band from 290 to 320 nanometers cause the skin to redden within a few hours of exposure. Investigators generally agree that the inflammatory reaction, which may persist for several days, results either from a direct action of ultraviolet photons on small blood vessels or from the release of toxic compounds from damaged epidermal cells. The toxins presumably diffuse into the dermis, where they damage the capillaries and cause reddening, heat, swelling and pain. A number of compounds have been proposed as the offending toxins, including serotonin, histamine and bradykinin. Sunburn is largely an affliction of industrial civilization. If people were to expose themselves to sunlight for one or two hours every day, weather permitting, their skin's reaction to the gradual increase in erythemal solar radiation that occurs during late winter and spring would provide them with a protective layer of pigmentation for withstanding ultraviolet radiation of summer intensities.

Immediately after exposure to sunlight the amount of pigment in the skin increases, and the skin remains darker for a few hours. The immediate darkening probably results from the photooxidation of a colorless melanin precursor and is evidently caused by all the wavelengths in sunlight. After a day or two, when the initial response to sunlight has subsided, melanocytes in the epidermis begin to divide and to increase their synthesis of melanin granules, which are then extruded and taken up into the adjacent keratinocytes, or skin cells [*see illustration at left*]. Concurrently accelerated cell division thickens the ultraviolet-absorbing layers of the epidermis. The skin remains tan for several weeks and offers considerable protection against further tissue damage by sunlight. Eventually the keratinocytes slough off and the tan slowly fades. (In the U.S.S.R. coal miners are given suberythemal doses of ultraviolet light every day on the theory that the radiation provides protection against the development of black-lung disease. The mechanism of the supposed protective effect is not known.)



MECHANISM OF SUN-TANNING is an extension of the mechanism responsible for skin pigmentation. After exposure to the sun melanocytes begin to divide and increase their output of melanin granules, produced in the tiny intracellular bodies called melanosomes. The melanosomes are secreted into the adjacent keratinocytes, or skin cells, where the melanin causes the skin to take on a darker appearance. The tan fades as the keratinocytes slough off.

In addition to causing sunburn and tanning, sunlight or its equivalent initiates photochemical and photosensitization reactions that affect compounds present in the blood, in the fluid space between the cells or in the cells themselves. A number of widely prescribed drugs (such as the tetracyclines) and constituents of foods (such as riboflavins) are potential photosensitizers. When they are activated within the body by light, they may produce transient intermediates that can damage the tissues in sensitive individuals. A typical response is the appearance of a rash on the parts of the body that are exposed to the sun.

In individuals with the congenital disease known as erythropoietic protoporphyria unusually large amounts of porphyrins (a family of photosensitizing chemicals) are released into the bloodstream as a result of a biochemical ab-

normality. The porphyrins absorb visible radiations and give rise to intermediates that are toxic to tissues. Patients with the disease complain at first of a burning sensation in areas of the skin that are exposed to sunlight; reddening and swelling soon follow.

Investigators can easily induce these typical symptoms without serious consequences in patients suffering from mild forms of erythropoietic protoporphyria, so that the disease is one of the few of its kind where the action spectrum for a direct effect of light has been studied in detail. The skin damage is caused by a fairly narrow band of wavelengths in the region of 400 nanometers. This band has also been shown to coincide with one of the absorption peaks of abnormal porphyrins. The symptoms of the disease can be ameliorated by administering photoprotective agents such as carotenoids, which quench the excited states of oxygen produced as intermediates in the photosensitization reactions.

In the past few years physicians have treated several skin diseases by deliberately inducing photosensitization reactions on the surface of the body or within particular tissues. The intent is to cause selective damage to invading organisms (such as the herpes virus), to excessively proliferating cells (as in psoriasis) or to certain types of malignant cells. The activated photosensitizers appear to be capable of inactivating the DNA in the viruses or in the unwanted cells. In treating herpes infections the photosensitizer (usually a dye, neutral red) is applied directly to the skin or to the mucous membrane under the ruptured blister; the area is then exposed to low-intensity white fluorescent light.

The treatment for psoriasis was devised by John A. Parrish, Thomas B. Fitzpatrick and their colleagues at the Massachusetts General Hospital. They administer a special photosensitizer (8-methoxypsoralen, or methoxalen) by mouth and two hours later expose the afflicted skin areas for about 10 minutes to the radiation from special lamps that emit strongly in the long-wave ultraviolet at about 365 nanometers. The sensitizing agent is present in small amounts in carrots, parsley and limes. It is derived commercially from an Egyptian plant (*Ammi majus* Linn.) that was used in ancient times to treat skin ailments. Scores of patients have responded successfully to the new light treatment, which will soon be generally available.

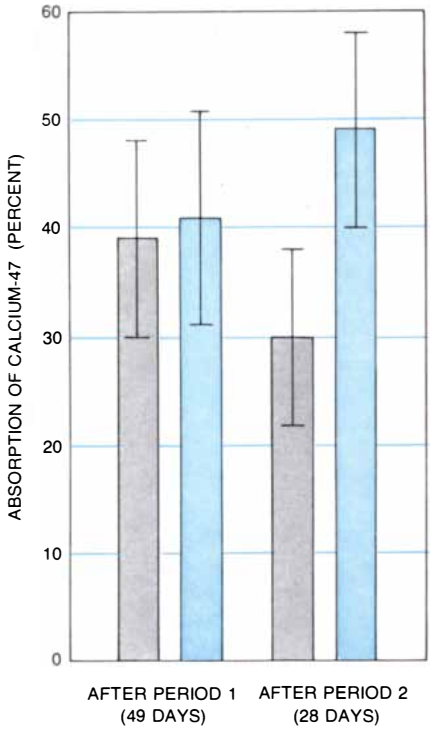
The formation of vitamin D_3, or cholecalciferol, in the skin and subcutaneous tissue is the most important of the beneficial effects known to follow exposure to sunlight. Vitamin D_3 is formed when ultraviolet radiation is absorbed by a precursor, 7-dehydrocholesterol. A related biologically active compound, vitamin D_2, can be obtained by consuming milk and other foods in which ergosterol, a natural plant sterol, has been converted to vitamin D_2 by exposure to ultraviolet radiation. Although vitamin D_2 can cure rickets in children who are deficient in vitamin D_3, it has not been demonstrated that vitamin D_2 is biologically as effective as the vitamin D_3 formed in the skin.

In a population of normal white adults living in St. Louis, studied by John G. Haddad, Jr., and Theodore J. Hahn of the Washington University School of Medicine, some 70 to 90 percent of the vitamin D activity in blood samples was found to be accountable to vitamin D_3 or its derivatives. The investigators concluded that sunlight was vastly more important than food as a source of vitamin D. (Although vitamin D_3 is also found in fish, seafood is not an important source in most diets.) In Britain and several other European countries the fortification of foods with vitamin D_2 has now been sharply curtailed because of evidence that in large amounts vitamin D_2 can be toxic, causing general weakness, kidney damage and elevated blood levels of calcium and cholesterol.

A direct study of the influence of light on the human body's ability to absorb calcium was undertaken a few years ago by Robert Neer and me and our coworkers. The study, conducted among elderly, apparently normal men at the Chelsea Soldiers' Home near Boston, suggests that a lack of adequate exposure to ultraviolet radiation during the long winter months significantly impairs the body's utilization of calcium, even when there is an adequate supply in the diet. The calcium absorption of a control group and an experimental group was followed for 11 consecutive weeks from the onset of winter to mid-March.

During the first period of seven weeks, representing the severest part of the winter, all the subjects agreed to remain indoors during the hours of daylight. Thus both groups were exposed more or less equally to a typical low level of mixed incandescent and fluorescent lighting (from 10 to 50 footcandles). At the end of the seven weeks the men in both groups were found to absorb only about 40 percent of the calcium they ingested. During the next four-week period, from mid-February to mid-March, the lighting was left unchanged for the control subjects, and their ability to absorb calcium fell by about 25 percent. The men in the experimental group, however, were exposed for eight hours per day to 500 footcandles of light from special fluorescent (Vita-Lite) lamps, which simulate the solar spectrum in the visible and near-ultraviolet regions. In contrast with the control subjects' loss of 25 percent of their capacity to absorb calcium, the experimental group exhibited an increase of about 15 percent [*see illustration above*]. The additional amount of ultraviolet radiation received by the experimental subects was actually quite small: roughly equivalent to what they would get during a 15-minute lunchtime walk in the summer.

Our study indicates that a certain amount of ultraviolet radiation, whether it is from the sun or from an artificial



CALCIUM ABSORPTION was increased by a daily eight-hour exposure to broad-spectrum artificial light in a study made by the author and his colleagues at a veterans' home. During the first seven weeks after the beginning of winter, control subjects (*gray bars*) and experimental subjects (*colored bars*) were equally exposed to the same low levels of typical indoor lighting. The bars at the left show their ability to absorb calcium at the end of the initial period. During the next four-week period conditions for the control subjects were unchanged; their ability to absorb calcium fell about 25 percent. The experimental subjects, who were exposed to 500 footcandles of broad-spectrum fluorescent light for eight hours per day for four weeks, showed an average increase of about 15 percent in their calcium absorption.

source, is necessary for adequate calcium metabolism. This hypothesis receives support from a recent study conducted by Jean Aaron of the Mineral Metabolism Unit at the General Infirmary at Leeds in England, who found that undermineralization (osteomalacia) is far more prevalent in autopsy samples collected in England during the winter months than it is in samples collected during the summer. Thus it seems likely that properly designed indoor lighting environments could serve as an important public-health measure to prevent the undermineralization of bones among the elderly and others with limited access to natural sunlight.

Perhaps 25,000 premature American infants were sucessfully treated with light last year as the sole therapy for neonatal jaundice. The rationale for this remarkable treatment is as follows. When red blood cells die, they release hemoglobin, which soon degrades into the yellow compound bilirubin. An increase in the concentration of bilirubin in the blood, due to excessive production of the compound or to failure of the liver to remove it, gives the skin its characteristic jaundiced color.

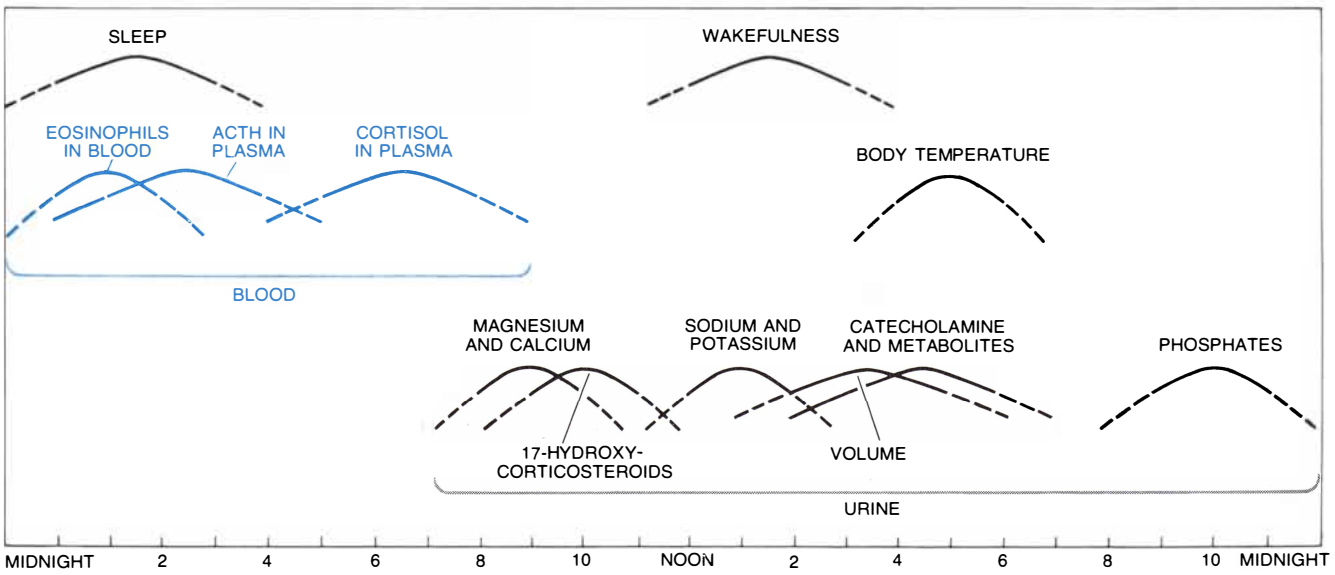
A potentially dangerous form of hyperbilirubinemia afflicts from 15 to 20 percent of premature infants because their liver is physiologically immature; in some cases the amounts of bilirubin released into the bloodstream are also increased as a result of blood-type incompatibility or concurrent infections. In such infants the bilirubin, which is soluble in fat, becomes concentrated in certain parts of the brain, where it can destroy neurons, producing the clinical syndrome kernicterus (yellow nuclei). The toxicity of bilirubin is aggravated by other factors, such as anoxia, acidosis, low body temperature, low blood sugar, low blood protein and infection. The brain damage resulting from kernicterus is often irreversible; it can cause various degrees of motor and mental retardation, leading to cerebral palsy and even death.

All current therapies for neonatal hyperbilirubinemia are based on the hope that if the level of bilirubin in the blood plasma can be kept from reaching between 10 and 15 milligrams per 100 milliliters until the maturing liver is able to remove the offending substance, there will be no brain damage. One widely used therapy involves exchange transfusions, in which jaundiced blood from the infant is completely replaced with normal blood from the donor.

Some years ago it was discovered that bilirubin in solution could be bleached by light and thus destroyed; the nature of the photodecomposition products remains unknown. This observation prompted R. J. Cremer, P. W. Perryman and D. H. Richards, who were then working at the General Hospital at Rockford in England, to see if light might be effective in lowering the plasma bilirubin in infants suffering from hyperbilirubinemia. That possibility was supported by informal observations that newborn infants whose crib had been placed near an open window tended to show less evidence of jaundice than infants whose crib was less exposed to light. Perhaps sunlight was accelerating the destruction of bilirubin. If it was, it should be possible to reproduce the effect with artificial light.

The efficacy of light therapy was fully confirmed in a controlled study conducted by Jerold F. Lucey of the University of Vermont College of Medicine. The treatment consists in exposing jaundiced infants to light for three or four days, or until their liver is able to metabolize bilirubin. Although it was initially assumed that the light converted the bilirubin into nontoxic products that could be excreted, it now turns out that a major fraction of the excreted material is unchanged bilirubin itself. Hence it is at least conceivable that phototherapy has a direct beneficial effect on the liver and the kidneys.

Many questions remain concerning the mechanism of phototherapy for hyperbilirubinemia and the long-term effectiveness of that therapy in protecting infants against brain damage. Blue light is the most effective in decomposing pure solutions of bilirubin. In clinical tests, however, full-spectrum white light in almost any reasonable dosage (continuous, intermittent or in brief strong pulses) has proved effective in lowering plasma-bilirubin levels, regardless of the fraction of the radiant energy that falls in the blue region of the spectrum. Thus the mechanism by which light destroys bilirubin in infants may differ from the simple photochemical reaction that takes place in a test tube. For example, a photosensitization reaction, perhaps mediated by circulating riboflavin, may underlie the desirable effect. Another possibility is that the light may act on the



DAILY RHYTHMS are characteristic of many human physiological functions. Whether these 24-hour rhythms are produced by the daily light-dark cycle or are simply entrained by that cycle remains to be unequivocally established. Each curve represents the typical daily peak for a physiological state or for the levels of particular substances that circulate in the blood or are excreted in the urine.

plasma albumin to which most of the circulating bilirubin is bound. Alternatively the physiologically effective wavelength may be not in the blue region at all but in some other region of the spectrum, still to be identified, that is present in all the white-light sources.

The observation that ordinary sunlight or artificial light sources can drastically alter the plasma level of even one body compound (in this case bilirubin) opens a Pandora's box for the student of human biology. It presents the strong possibility that the plasma or tissue levels of many additional compounds are similarly affected by light. Some such responses must be physiologically advantageous, but some may not be.

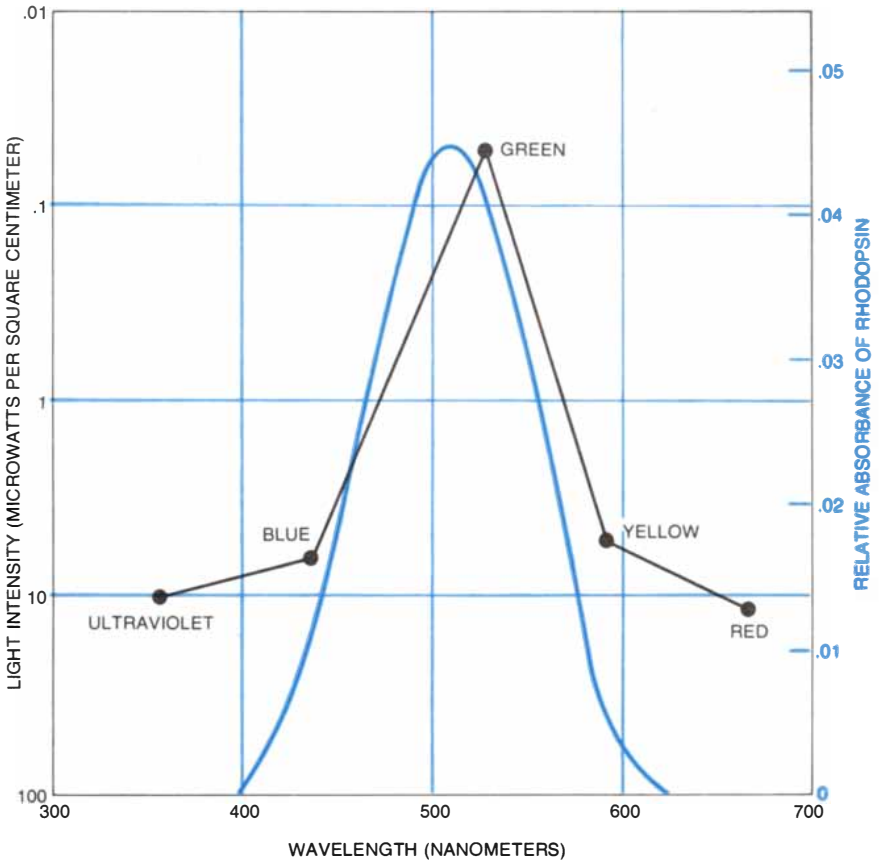
Let us turn now to the indirect effects of light, those associated in one way or another with biological rhythms. The amount of time that all mammals are exposed to light varies with two cycles: the 24-hour cycle of day and night and the annual cycle of changing day length. (Even at the Equator there are small seasonal variations in the light-dark cycle.) These light cycles appear to be associated with many rhythmic changes in mammalian biological functions. Physical activity, sleep, food consumption, water intake, body temperature and the rates at which many glands secrete hormones all vary with periods that approximate 24 hours.

In human beings, for example, the concentration of cortisol, one of the principal hormones produced by the adrenal cortex, varies with a 24-hour rhythm [*see illustration on opposite page*]. The level is at a maximum in the morning hours, soon after waking, and drops to a minimum in the evening. When people reverse their activity cycle, by working at night and sleeping during the day, the plasma-cortisol rhythm takes from five to 10 days to adapt to the new conditions. When the cortisol level is studied in rats, it is found that the rhythm persists in animals that are blinded but not in animals kept under continuous illumination. Blindness in human beings seems to upset the rhythm, so that the times of the daily peaks and valleys are out of phase with the normal pattern and may even vary from day to day.

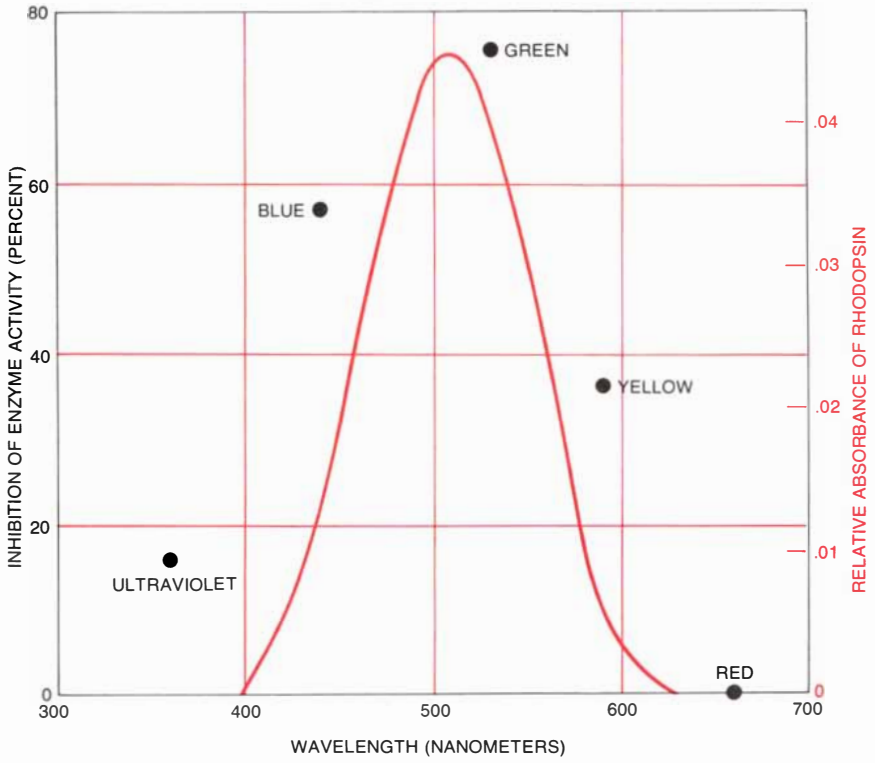
Among the rhythmic functions that can be closely studied in one and the same animal (specifically rhythms in sleep, physical activity and food consumption) it has been shown that in the absence of cyclic exposure to light the rhythms become "circadian" (that is, their periods become approximately 24 hours in length rather than exactly). The fact that such rhythms can "free-run" suggests that they are not simply reflex responses to 24-hour cycles of light or some other environmental component. The factors responsible for the rhythms are not yet known; they might include other cyclic inputs such as the consumption of food or the intake of water, which also free-run in the absence of light. Some investigators are convinced that the rhythms are generated by intrinsic oscillators, commonly called biological clocks.

Little is known about the action spectra or the light intensities needed either to generate or to "entrain" (synchronize) daily rhythms in mammals. There is strong presumptive evidence, however, that in most mammals light exerts its effects indirectly through photoreceptors in the eye. It is not known whether the photoreceptors are the same ones (the rods and the cones) that mediate vision, discharging into nonvisual pathways, or whether they are a distinct family of photoreceptors with their own neural network.

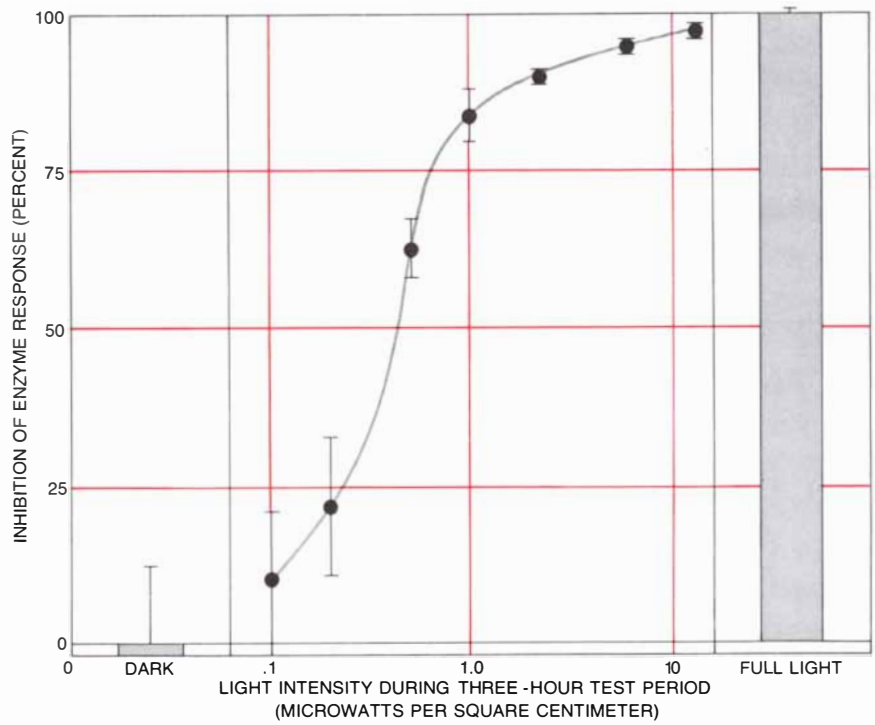
In our laboratory at M.I.T. we have investigated the daily rhythmicity in the body temperature of rats to see what colors of light are most effective in inducing a change in rhythm to a new light-dark cycle and what intensities are needed. The body temperature of rats normally rises by one or two degrees C. at the onset of darkness and falls again at daybreak. We found that green light is the most potent in changing the phase of the temperature cycle and that ultraviolet and red wavelengths are the least potent. The action spectrum plotted from these results closely follows the absorption spectrum for rhodopsin, the photosensitive pigment in the rods of the retina [*see illustration above*]. In separate studies a similar action spectrum, peaking in the green, was found for the wavelengths of light that are most effective in inhibiting the function



BODY-TEMPERATURE RHYTHM IN RATS, which follows a 24-hour cycle, can be altered by shifting the hours of the light-dark cycle. The author and his colleagues have found that green light is much more effective in establishing a new rhythm than radiation of other wavelengths (red, yellow, blue or ultraviolet). Black curve estimates the wavelength and intensity needed to establish a new 24-hour rhythm in half of an experimental group of rats. It closely follows the relative sensitivity of rhodopsin from rat retinas (*curve in color*).



PINEAL ACTIVITY OF RATS can be suppressed by exposing the animals continuously to light. As in the case of the daily temperature rhythm, green light is more effective than light of other spectral colors in suppressing the organ's enzyme activity, as is shown by the labeled dots. The enzyme that is measured is the melatonin-forming enzyme hydroxyindole-O-methyltransferase. Presumably the suppression is mediated by rhodopsin (*curve in color*).



GRADATION IN LIGHT INTENSITY leads to proportional inhibition in the activity of the rat's pineal gland, indicating that light controls synthesis of the pineal hormone, melatonin. Rats that had been kept in constant light for 48 hours were exposed to various light intensities for the next three hours. Their pineals were then analyzed for serotonin-N-acetyltransferase, an enzyme that participates in melatonin synthesis, with the results plotted here.

of the pineal gland of rats [*see top illustration at left*].

Cycles in environmental lighting can interact with biological rhythms in two ways. The light cycle may directly induce the rhythm, in which case either continuous light or darkness should rapidly abolish it, or the cycle may simply entrain the biological rhythm so that all animals of a given species exhibit maximums or minimums at about the same time of day or night. In the latter case the rhythmicity itself may be generated by a cyclic input other than light, either exogenous (for example food intake) or endogenous (a biological clock). If the cycle is simply entrained by light, an environment of continuous light or darkness might not extinguish it. In human beings psychosocial factors are probably of greater importance than light cycles in generating or synchronizing biological rhythms. The biological utility of even so dramatic a rhythm as that of sleep and wakefulness, for example, remains to be discovered.

Annual rhythms in sexual activity, hibernation and migratory behavior are widespread among animals. The rhythms enable members of a species to synchronize their activities with respect to one another and to the exigencies of the environment. For example, sheep ovulate and can be fertilized only in the fall, thus anticipating the spring by many months, when food will be available to the mother for nursing the newborn. In man no annual rhythms have been firmly established, except, of course, those (such as in sun-tanning and vitamin D_3 levels) that are directly correlated with exposure to summer sunlight.

The best-characterized indirect effect of light on any process other than vision is probably the inhibition of melatonin synthesis by the pineal organ of mammals. Although melatonin seems to be the major pineal hormone, its precise role has not yet been established. When melatonin is administered experimentally, it has several effects on the brain: it induces sleep, modifies the electroencephalogram and raises the levels of serotonin, a neurotransmitter. In addition melatonin inhibits ovulation and modifies the secretion of other hormones from such organs as the pituitary, the gonads and the adrenals, probably by acting on neuroendocrine control centers in the brain.

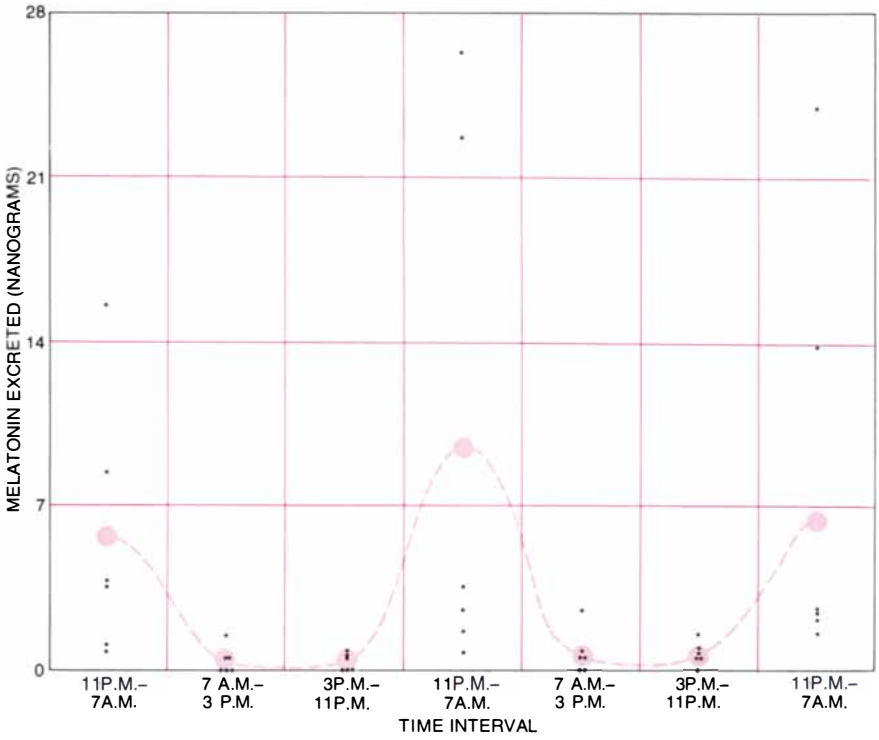
Experiments performed on rats and other small mammals during the past decade provide compelling evidence that the synthesis of melatonin is suppressed by nerve impulses that reach the pineal

over pathways of the sympathetic nervous system. These impulses in turn vary inversely with the amount of visible light impinging on the retina. In rats the pineal function is depressed to half its maximum level when the animals are subjected to an amount of white light only slightly greater than that shed by the full moon on a clear night [*see bottom illustration on opposite page*]. A multisynaptic neuronal system mediates the effects of light on the pineal. The pathway involved, which is apparently unique to mammals, differs from the route taken by the nerve impulses responsible for vision.

Quite recently Harry Lynch, Michael Moskowitz and I have found a daily rhythm in the rate at which normal human subjects excrete melatonin. During the third of the day corresponding to the bedtime hours, 11:00 P.M. to 7:00 A.M., the level of melatonin in the urine is much higher than it is in any other eight-hour period [*see illustration at right*]. It remains to be determined whether the rhythm in melatonin excretion in humans is induced by light or is simply entrained by it.

In some birds and reptiles the pineal responds directly to light, thereby serving as a photoreceptive "third eye" that sends messages about light levels to the brain. In the pineal organ of mammals any trace of a direct response to light is lost. Evidently photoreceptors in the retina mediate the control of the pineal by light. Since, as I have noted, the function of the pineal in rats is influenced most strongly by green light, corresponding to the peak sensitivity of the rod pigment rhodopsin, the retinal photoreceptor would seem to be a rod cell, at least in this species.

Light levels and rhythms influence the maturation and subsequent cyclic activity in the gonads of all mammals and birds examined so far. The particular response of each species to light seems to depend on whether the species is monestrous or polyestrous, that is, on whether it normally ovulates once a year (in the spring or fall) or at regular intervals throughout the year. Examples of polyestrous species are rats (ovulation every four or five days), guinea pigs (every 12 to 14 days) and humans (every 21 to 40 days). The gonadal responses also seem to depend on whether the members of the species are physically active during the daylight hours or during the night. Recently Leona Zacharias and I had the opportunity to examine more than a score of girls and women (members of a diurnally active polyestrous species) who had become blind in the first year of life. We observed that gonadal maturation had in general occurred earlier in this group than in normal girls. In contrast, in rats (a polyestrous species that is active at night) blindness delays maturation, and continuous illumination accelerates the maturation of weanlings with normal vision.

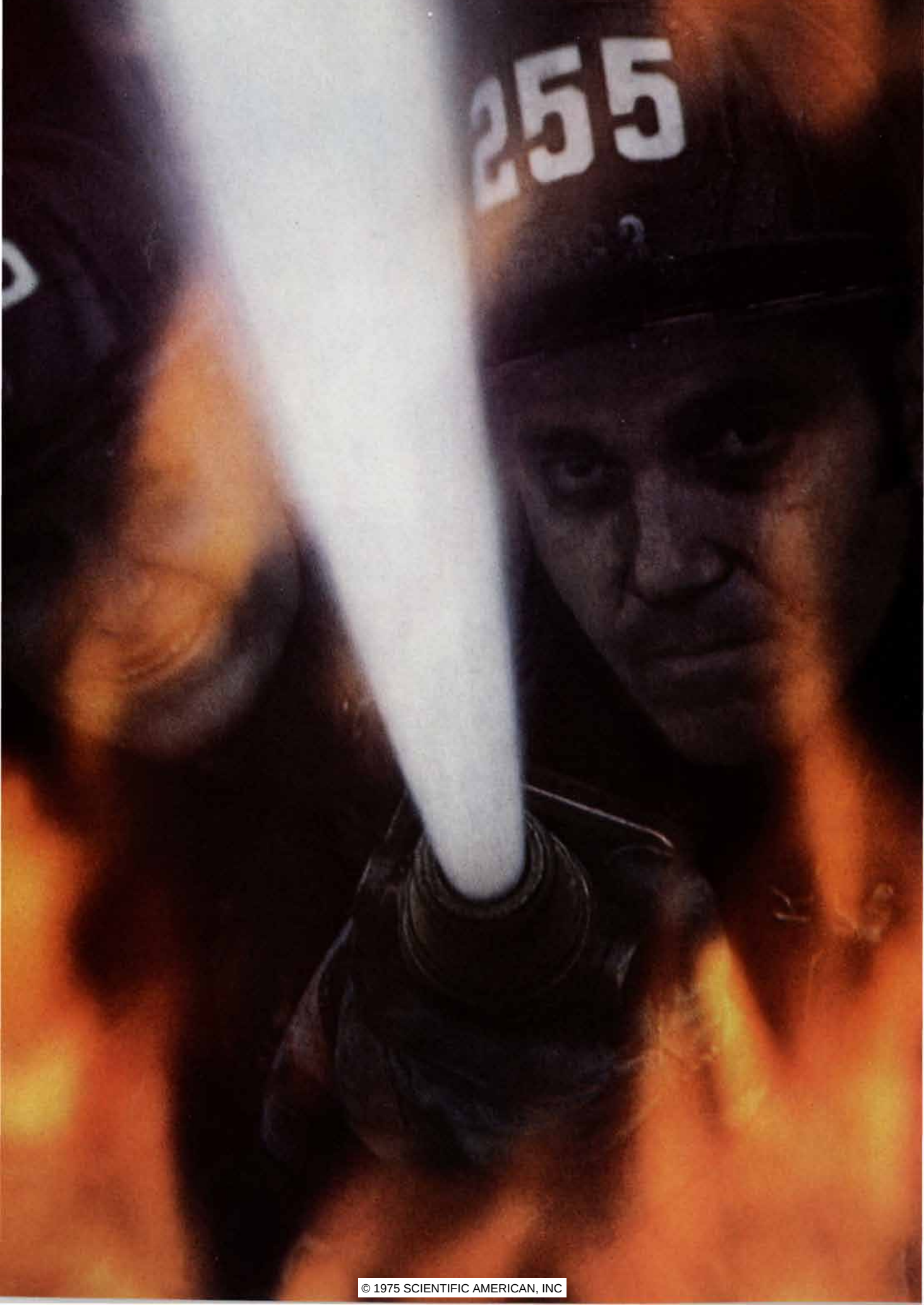


RHYTHM IN MELATONIN SECRETION in human beings has been found by the author and his colleagues. The black dots show the melatonin content of urine samples from six subjects during consecutive eight-hour periods. The colored circles and broken curve correspond to the mean values. High values that were recorded for the 11:00-P.M.-to-7:00-A.M. samples suggest that synthesis of melatonin in man, as in rats, increases with onset of darkness.

The gonads of most birds and of most diurnally active, monestrous animals (the ferret, for instance) mature in the spring, in response to the gradual increase in day length. Ovulation can be accelerated in such animals by exposing them to artificially long days. The annual gonadal activity in domestic sheep, on the other hand, occurs in the fall, in response to the decrease in day length. The mechanisms that cause some species to be monestrous and others polyestrous, or that cause some animals to sleep by day and others by night, are entirely unknown, as are the factors that cause the gonadal responses of various species to light to vary as widely as they do.

The multiple and disparate effects of light I have described support the view that the design of light environments should incorporate considerations of human health as well as visual and aesthetic concerns. We have learned that the chemical constituents of the environment in the form of food, drugs and pollutants must be monitored and regulated by agencies with suitable powers of enforcement. A major part of their responsibility is to see that nothing harmful is put into food or drugs and that nothing essential is left out of food. The food and drug industries, for their part, look to public and private research organizations, including their own laboratories, for intellectual guidance in creating wholesome and beneficial (as well as profitable) products.

In contrast, only minuscule sums have been expended to characterize and exploit the biological effects of light, and very little has been done to protect citizens against potentially harmful or biologically inadequate lighting environments. Both government and industry have been satisfied to allow people who buy electric lamps—first the incandescent ones and now the fluorescent—to serve as the unwitting subjects in a long-term experiment on the effects of artificial lighting environments on human health. We have been lucky, perhaps, in that so far the experiment has had no demonstrably baneful effects. One hopes that this casual attitude will change. Light is potentially too useful an agency of human health not to be more effectively examined and exploited.



Announcing a major development in fire fighting. Faster water.

Fighting a fire is fighting time.

Every minute wasted may cost thousands of dollars in property destroyed. Every second saved may be a life.

And yet, a fire fighter can't always move as fast as he would like.

Because his basic weapon, the 2½-inch fire hose, when full of water, is about as hard and unyielding as a steel pipe. The only way to bend it around a corner is to shut the water off.

And to lug it up a flight of stairs is a test of any man's stamina.

All of this makes it easy to understand why a Union Carbide product called UCAR Rapid Water Additive is revolutionizing fire fighting.

Rapid Water Additive mixes easily with water and makes it flow faster through a hose by reducing friction.

Because the water flows faster, fire fighters can get just as much water as before using a smaller 1¾-inch hose that weighs half as much.

The smaller hose makes a much better weapon. It bends around corners. It can be carried up a stairway fully charged. In short, it helps firemen get water to the fire faster.

At the same time, it makes their job safer and less strenuous.

Rapid Water Additive was invented by Union Carbide and perfected in cooperation with the New York City Fire Department.

It's already helping fire fighters do their job more efficiently in many American cities.

Yours may be one—now or soon.

UNION CARBIDE

Today, something we do will touch your life.

THUNDER

It is the acoustic signal generated by a rapidly expanding channel of heated air. From the information in this signal it is possible to deduce the location, shape and orientation of a lightning flash

by Arthur A. Few

It is obvious today that the ultimate cause of thunder is lightning; it is less obvious just how an electrical discharge in the atmosphere produces the variety of powerful noises heard in a thunderstorm. A lightning flash dissipates a prodigious amount of energy, but how is that energy (or a portion of it) transformed into sound waves? Moreover, a lightning stroke is essentially instantaneous; how does it generate the often protracted sequence of rumbles, claps, booms and other sounds heard in a peal of thunder?

Considering that thunder has been a topic of speculation since antiquity, one might expect that these questions would have been answered long ago. Actually even the most general principles of thunder production were not established until this century, and a theory of thunder that is both comprehensive and detailed has been approached only in the past 10 or 15 years. A number of important questions remain unresolved today.

With my colleagues at Rice University and in cooperation with workers elsewhere I have investigated thunder by recording its "signature": the pattern of sound waves received at a particular location from a single lightning flash. The study of such signatures has revealed much about the nature of thunder itself, and it has led to the formulation of a theory of how features of the lightning discharge are expressed in the acoustic signal. The theory is successful enough for us now to employ thunder as a tool in searching for the source of electricity in clouds.

The relation of lightning to thunder was well established by the end of the 19th century. It remained to be determined what effects of the lightning discharge are responsible for producing the sound. Four principal theories were proposed, and in the first decade of the 20th century they were vigorously debated. (In 1903 *Scientific American* published four contributions on the nature and cause of thunder.)

The first of the theories held that the lightning stroke creates a vacuum and that thunder is produced when the vacuum collapses. Another maintained that water drops in the path of the lightning flash are turned into steam and that the rapid expansion of the steam is accompanied by a loud report. A third proposal suggested that the electrical discharge decomposes water molecules by electrolysis and that the hydrogen and oxygen thereby produced subsequently recombine explosively. Finally, the simplest explanation ascribed thunder to the sudden heating of the air in the path of the lightning flash. Because air has electrical resistance, it is heated by the passage of a current just as a wire is; the expansion of the heated air was considered sufficient to explain the ensuing thunder.

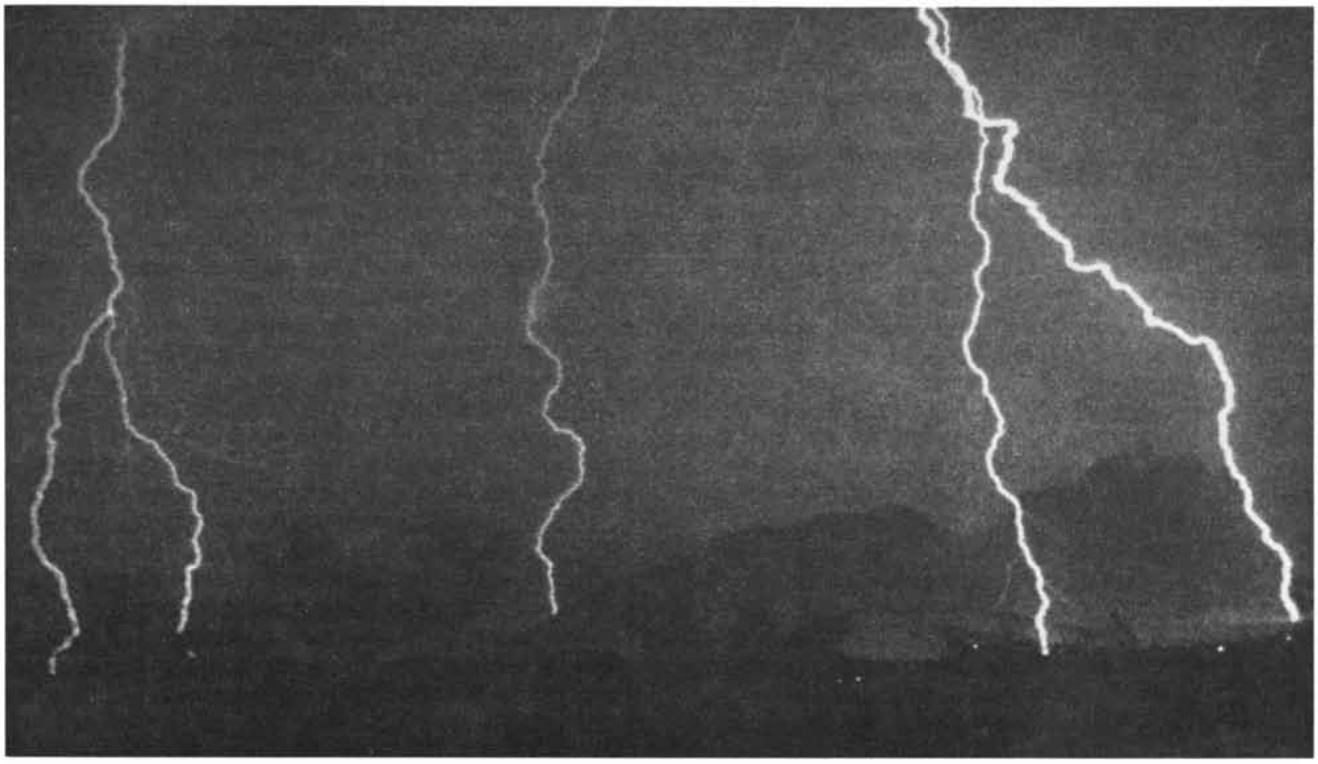
We now know that the last explanation is the correct one. Each surge of current in the lightning flash heats the air in its path, creating a channel of gases at high temperature and pressure. The gases expand into the surrounding air as a shock wave, which after traveling a short distance decays into an acoustic wave.

It is interesting to note that the other theories were not entirely without foundation. For example, a region of diminished air pressure is briefly created in the aftermath of lightning; this partial vacuum is an effect of thunder, however, rather than a cause. Moreover, water drops are certainly evaporated in the lightning channel, and water molecules are decomposed. There is excellent evidence for the decomposition in the optical spectrum of lightning: one of the most prominent features of the spectrum is an emission line of hydrogen. Both of these phenomena, however, are merely collateral effects of lightning; they make no significant contribution to thunder.

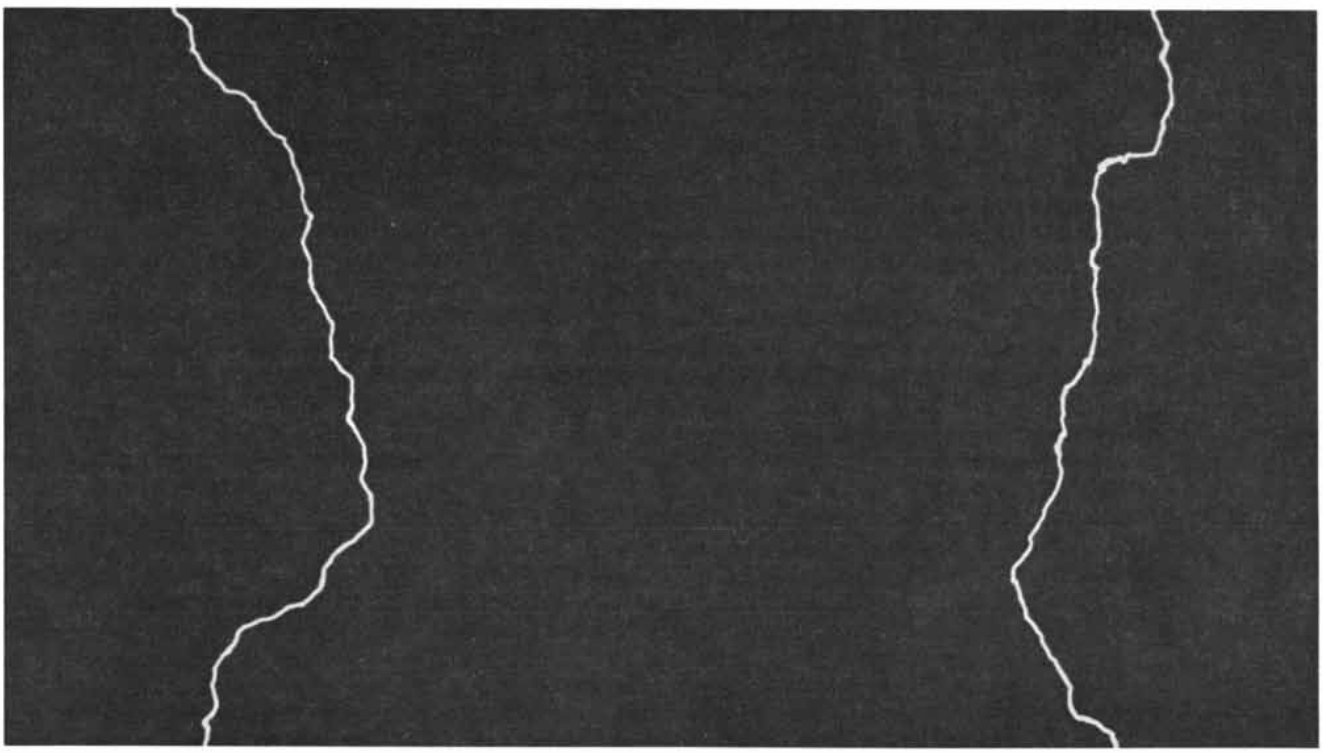
The interest in the nature of thunder that culminated in the theories of the early 20th century declined soon thereafter, and the study of thunder was largely ignored for 50 years. Isolated experiments and observations were made during that period, but they tended more to confuse than to enhance understanding. For example, from measurements of the duration of thunder one can estimate the length of the lightning channel; in many cases such measurements were found to yield lengths much greater than that of the observed lightning flash and in some cases the lengths were greater than the height of the cloud itself. This discrepancy has only recently been explained.

Since 1960 scientific interest in thunder has revived and research into the nature and source of thunder has been renewed. The investigation is greatly assisted by a technology that was unavailable to earlier workers. The fundamental principles are no longer in question; the challenge to research today is to account for the detailed features of the thunder signature.

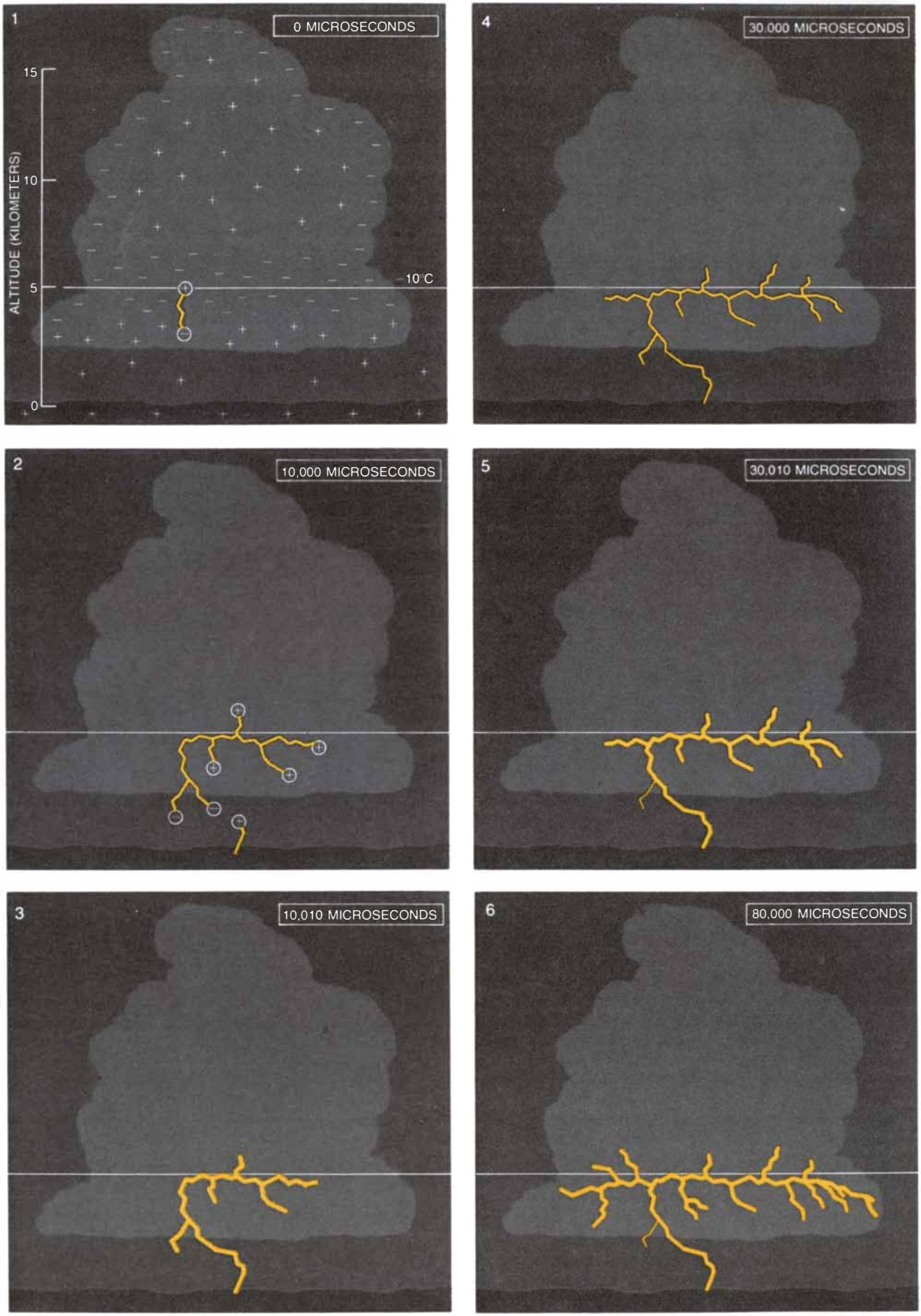
What is heard in a peal of thunder depends in large measure on the characteristics of the particular lightning flash that produced it. Both the temporal sequence of events in the discharge and their arrangement in space must be considered; these two complex factors determine not only the frequency and the amplitude of the radiated acoustic waves but also the order in which the waves



FOUR LIGHTNING FLASHES were photographed near Tucson, Ariz. The prominent kinks and bends are the large-scale, or macrotortuous, elements of the lightning channels; segments between major bends are on the order of 100 meters long. The macrotortuous features determine the overall pattern of claps and rumbles in thunder. The four flashes were not simultaneous but were recorded in a single photograph by making a time exposure lasting for about two minutes. The photograph was made by Henry B. Garrett of Rice University. The forked lightning channel at left was produced by a flash consisting of multiple strokes that followed divergent paths.



SMALLER FEATURES of a lightning channel are discernible in a photograph made with a telephoto lens. The area seen is the lower portion of the forked channel in the photograph at the top of the page. The many small straight segments are on the order of 10 meters long; elements of the channel in this size range are classified as mesotortuous and can be considered point sources of thunder. The microtortuous features (those shorter than about five meters) can rarely be resolved in photographs and have little effect on thunder.



are received at a given observation site.

A lightning flash in most instances begins near the base of a cloud in a region of dense negative charge. This stratum is typically at an altitude of about five kilometers, where the temperature is approximately −10 degrees Celsius. That is the region of the cloud where water droplets freeze, a process that may be connected with the generation of electric charge.

The negatively charged zone of the cloud can be at a potential of as much as 300 million volts with respect to the ground, but even that enormous potential is insufficient to support a spontaneous arc across five kilometers of air. The main discharge can begin only after the channel has been traced by a preliminary low-current discharge called the stepped leader. The stepped leader begins to form when electrons emitted by droplets in the cloud are accelerated by the intense electric field; the electrons collide with molecules in the air, freeing many more electrons and leaving a conductive path of partially ionized air. Such a cascade of accelerated electrons typically progresses only 50 to 100 meters, but with each step a portion of the cloud's charge is transferred downward, and the next step can begin from the tip of the advancing leader.

The course of the stepped leader is highly irregular and in its progress toward the ground it forms numerous branches. Each step is accomplished in less than a microsecond, but there is a pause of about 50 microseconds between steps. As the leader approaches the ground the potential gradient—the voltage per meter—increases, and sparks are emitted from objects and structures on the ground, usually from the highest points first. When one of these sparks meets the downward-propagating leader, a conducting path between the cloud and the ground is completed. Since the potential difference across the path is a few hundred million volts a surge of current immediately follows; this large current is called the first return stroke.

The stepped leader may require 20 milliseconds to create the channel to the ground, but the return stroke is completed in a few tens of microseconds. In some cases that is the end of the lightning flash; more commonly, however, the leader-and-stroke process is repeated in the same channel at intervals of tens to hundreds of milliseconds. The subsequent leaders, called dart leaders, progress faster and more smoothly than the stepped leader because the electrical resistance of the path they follow is lower than that of the surrounding air. As the dart leader progresses toward the ground, intracloud processes extend the channel within the cloud so that additional areas are discharged. The subsequent return strokes, however, are usually less energetic than the first one is. A typical lightning flash has three or four leaders, each followed by a return stroke; one flash has been photographed that had 26 strokes.

Each surge of current in the flash, including the steps in the stepped leader, the dart leaders and the return strokes, heats the gases in the lightning channel and thereby generates an acoustic signal. The amplitude and duration of the signal produced by each current surge depend on the magnitude of the current. The complete thunder signal therefore reflects a complex sequence of events in the lightning flash. Ordinarily it is not possible to distinguish in a recorded thunder signature the acoustic pulses generated by individual leaders and strokes, but the sequence of strokes and the currents they carry nevertheless determine what sound is produced. In a few exceptional recordings the correlation of thunder pulses with lightning strokes is evident.

The spatial arrangement of the lightning flash has perhaps a greater influence on the resulting thunder than the temporal organization. The lightning channel is said to be tortuous: it consists of straight segments separated by sharp bends. The structural elements are classified in three size ranges. The large-scale features of the channel, called the macrotortuous elements, are straight segments at least 100 meters long. Straight segments between five and 100 meters long are classified as mesotortuous, and those shorter than five meters are called microtortuous. The mesotortuous and macrotortuous elements are not, of course, actually straight lines; they are made up of many smaller segments and seem to be straight only when considered at the proper scale. Moreover, the classification of a feature as mesotortuous or microtortuous is not determined by a fixed rule but depends on the energy of the discharge. Five meters is an appropriate minimum length for mesotortuous features in strokes carrying large currents, but smaller elements can be considered mesotortuous when the current is comparatively small, as it is in stepped leaders and dart leaders.

The mesotortuous segments are the primary radiators of the acoustic pulses of thunder. The entire channel can be considered a "string of pearls," each pearl a mesotortuous segment radiating a series of pulses determined by the sequence of pulses in the lightning flash. The macrotortuous segments determine the spatial organization of the individual acoustic radiators and hence have a profound influence on what is heard by the observer.

A pulse of thunder begins in a channel of hot gases at high pressure. Spectroscopic evidence indicates that the temperature in the channel can reach 30,000 degrees C., and the pressure can exceed atmospheric pressure by from 10 to 100 atmospheres. Initially the high-pressure core expands as a shock wave. A shock wave is distinguished from an acoustic wave in that it compresses and heats the medium in which it propagates and thereby increases the speed of sound. Because the speed of sound increases as the temperature rises, the shock wave travels faster than sound does in the same medium. The extent of the compression and heating, and therefore the magnitude of the increase in velocity, depend on the amplitude of the wave. Behind the shock wave the air continues to move outward, and a region of low pressure forms.

The expanding shock wave dissipates its energy in performing work on the surrounding air. When all the energy imparted to the shock wave by the lightning stroke has been expended, the wave "relaxes" and the pressure in the vicinity of the channel returns to normal. (The core of the channel itself, however, re-

LIGHTNING CHANNEL develops from a region of concentrated negative charge near the base of a cloud, an area associated with the freezing of water droplets. The flash begins with the formation of a stepped leader (*1*), which moves downward in steps from 50 to 100 meters long and simultaneously extends streamers horizontally through the charged region. As the leader nears the ground, sparks propagate upward to meet it (*2*); when the pathway is completed, the large current of the first return stroke flows (*3*). Subsequent leaders, called dart leaders (*4*), progress much faster and extend the channel to other parts of the cloud; each is followed by a return stroke (*5*). Much of the length of the completed channel is horizontal and only a small portion of it is visible below the cloud (*6*). Each leader and stroke heats the gases in the lightning channel and contributes to the signal ultimately perceived as thunder.

mains a region of high temperature and low density.) From the ambient pressure and the energy per unit length of the lightning stroke one can calculate the radius at which the shock wave relaxes. The relaxation radius is in turn related to the wavelength of the thunder. Thus the wavelength, or pitch, of thunder is determined by the energy of the lightning stroke and the ambient air pressure in the region where the thunder is generated. The more powerful the stroke or the lower the air pressure, the lower the pitch of the resulting thunder. A typical value is 60 hertz [*see illustration below*].

The relaxation radius serves as the actual measure distinguishing microtortuous elements from mesotortuous ones. Features of the channel that are smaller than the relaxation radius are blurred in the expansion of the shock waves. They have little influence on the form of the resulting acoustic waves and therefore cannot be resolved in data derived from thunder. Microtortuous features are thus defined as those that are too small to be detected in the thunder signature.

The shock wave is not an efficient source of acoustic radiation: less than 1 percent of its energy is transmitted to the acoustic wave. The remaining 99 percent is dissipated in heating the air in the vicinity of the lightning channel. The total energy of the shock wave is very large, however, and even the small fraction of it converted into sound generates an acoustic wave of large amplitude. The result is one of the loudest sounds in nature.

Sound produced by the mesotortuous features of the channel is not radiated with equal power in all directions. The study of large sparks produced in the laboratory has shown that more than 80 percent of the acoustic energy is confined to a zone 30 degrees above and below a plane that bisects the spark perpendicularly. A microphone placed end on to the spark will receive much less sound than one placed broadside to it. This directed quality of the radiated sound is one of the most important determinants of what is heard by the listener on the ground.

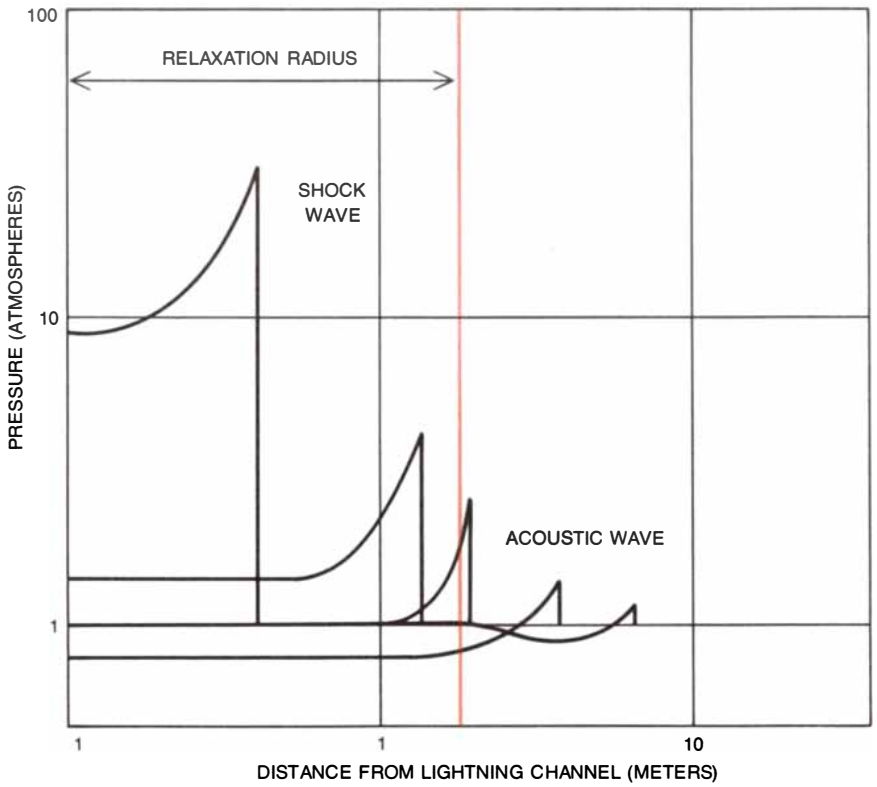
From studies of photographs of lightning it has been found that the average change in direction or orientation between adjacent mesotortuous segments is about 16 degrees. Because this angle is substantially smaller than the zone 60 degrees wide into which each segment radiates most of its energy, a number of mesotortuous segments strung along a single macrotortuous segment will radiate most of their energy in roughly the same direction [*see top illustration on opposite page*]. It is this property that is responsible for the sudden claps and prolonged rumbles of thunder.

Acoustic pulses from all parts of a lightning stroke are, of course, emitted almost simultaneously. The entire acoustic output of a lightning flash is produced in less than a second, in the time required for the complete sequence of leaders and strokes. The thunder persists much longer than that because the lightning channel is long (at least five kilometers long and often considerably more) and signals from its nearest segments reach the listener sooner than those from the farthest extensions. The orientation of the channel, and in particular the orientation of the macrotortuous segments, determine the character of the sounds heard.

If a macrotortuous segment is oriented end on to the observer, the sound received will be of comparatively low amplitude, since most of the acoustic energy is radiated perpendicularly to the segment. Moreover, the wave fronts from each mesotortuous element will reach the listener in succession, beginning with signals emitted by the nearest part of the segment. The result is a protracted rumble or roll of thunder.

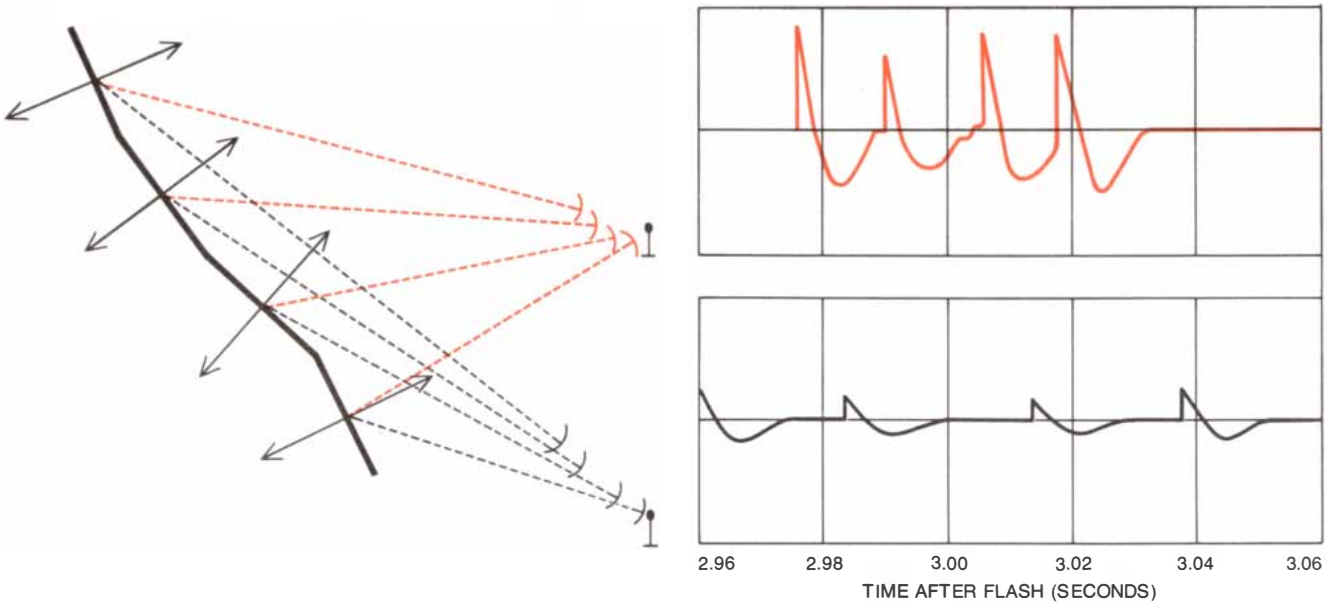
If the macrotortuous segment is broadside to the listener, a much larger portion of the radiated energy will be received. Equally important, the wave fronts from all the mesotortuous elements will arrive almost simultaneously. As a result a brief but intense clap of thunder will be heard [*see bottom illustration on opposite page*].

Thunder produced by a single flash of lightning is commonly perceived as a combination of claps and rumbles because various elements of the lightning channel are oriented differently with respect to the listener. Thunder from the same lightning flash will be perceived differently at different locations, since each location has a unique position and orientation with respect to the lightning channel.

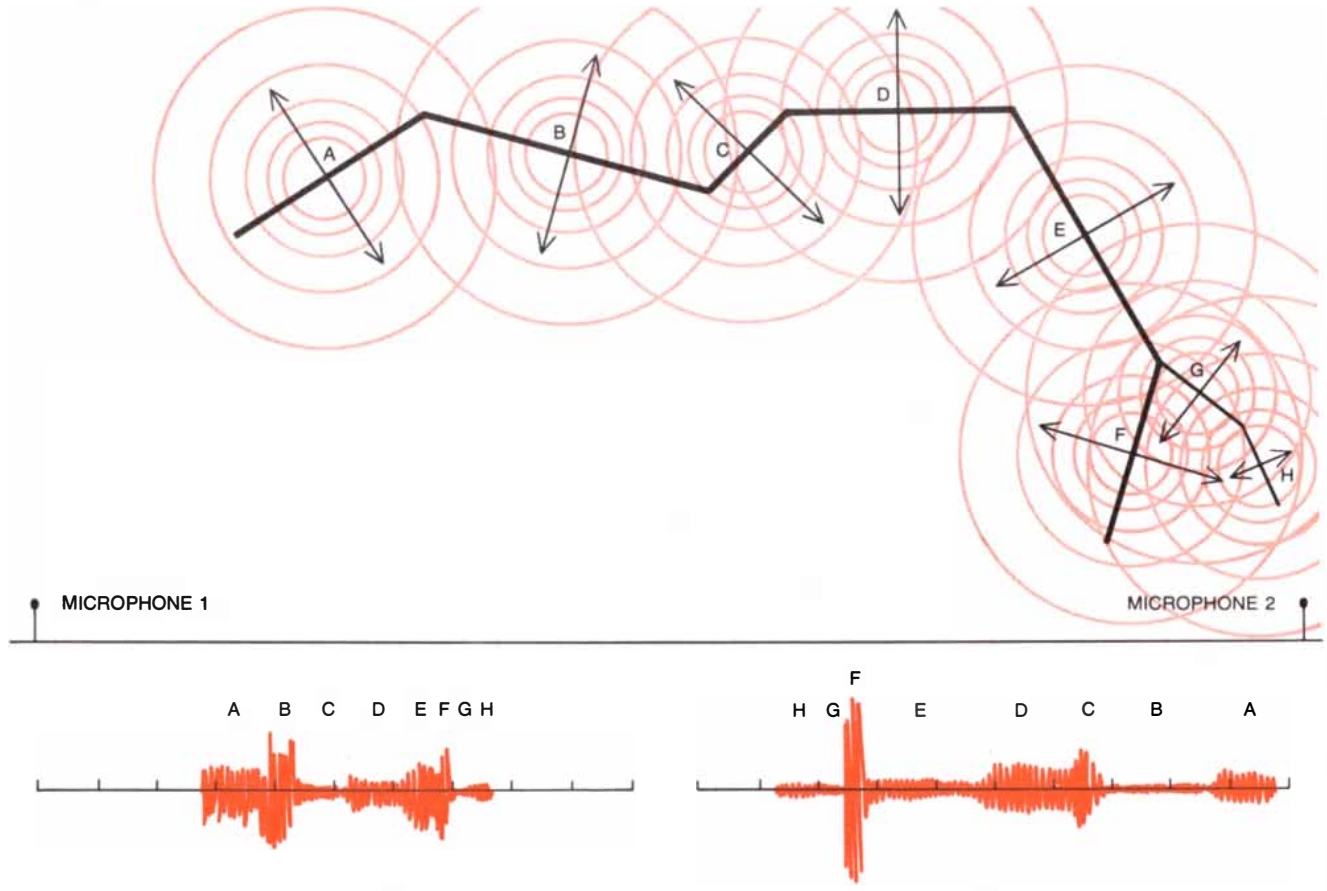


EXPANDING CHANNEL of hot gases produced by a lightning stroke propagates as a shock wave and then as an acoustic pulse of thunder. Because the initial shock wave compresses and heats the air, it quickly dissipates its energy. A few meters from the lightning channel it "relaxes" to yield an acoustic wave of lower amplitude, and the pressure in the region behind the wave is briefly reduced. Only about 1 percent of the energy of the shock wave is transferred to the acoustic wave; the rest is expended in heating the air near the channel.

Between the lightning channel and the listener the thunder signal is altered by the medium in which it travels. The atmosphere attenuates, scatters and refracts the signals; in addition they are subject to nonlinear propagation effects



MESOTORTUOUS SEGMENTS of the lightning channel are the primary radiators of thunder. Each of the four segments shown can be considered an independent point source of sound. Acoustic pulses are emitted from the segments almost simultaneously, but the pattern of sounds perceived on the ground depends on how the channel is oriented with respect to the observer. Sounds emitted perpendicularly to a segment (*color*) are more powerful than sounds emitted parallel or nearly parallel to it (*black*). Since each segment differs in orientation from the adjoining segments by only a small angle, signals from several segments reach an observer whose position is perpendicular to the channel almost simultaneously; the result is a brief but loud clap of thunder. An observer looking up the length of the channel, on the other hand, receives wave fronts at greater intervals and hears a prolonged rumble.



MACROTORTUOUS ELEMENTS of the lightning channel determine the overall pattern of claps and rumbles in the thunder signature. The amplitude and duration of the signal produced by each segment are determined by the orientation of the segment and its distance from the observer. Microphones placed at different positions will record different signatures from the same lightning channel. For this hypothetical lightning flash the segments have been labeled along with the sounds they would produce at two locations.

and, once they reach the ground, to reflection.

"Nonlinear propagation" refers to processes that affect one part of a wave more than others or some frequencies of sound more than others. Nonlinear propagation therefore alters the waveform of the signal (the shape of the individual waves) or its spectrum. One such effect tends to lengthen each pulse as it propagates; sound waves of large amplitude are the most severely affected, so that the process is most significant close to the lightning channel.

The attenuation of thunder in the atmosphere is caused by two independent processes. "Classical" attenuation results from the viscosity of air, that is, from the fact that air is not a perfectly elastic medium. Classical attenuation is well understood, and its magnitude can be predicted. "Molecular" attenuation results from a complex interaction of sound waves with water molecules and oxygen molecules in which acoustic energy excites internal vibrations of the molecules. It can be evaluated only if the temperature and humidity are known for all points along the path traveled by the thunder signal. Molecular attenuation is usually the more important of the two effects. Attenuation increases as the square of the frequency of the signal; thus a 20-hertz wave will travel four times as far as a 40-hertz wave before it is attenuated to the same degree.

The scattering of the thunder signal is even more difficult to predict than its attenuation. The principal agents of scattering are turbulent eddies in the atmosphere, which range in size from microscopic disturbances a few microns in diameter to the thunderstorm cell itself. For the scattering of thunder the most significant eddies are those that are approximately the size of the wavelength of the thunder (less than 50 meters). Again, the losses are severer for higher frequencies.

Because of scattering and attenuation, when a thunder signal travels several kilometers through a turbulent medium, only the lowest-frequency components of the original spectrum survive without major modification. As a result a lightning flash of low energy, which produces little low-frequency sound, will not give rise to audible thunder except at close range.

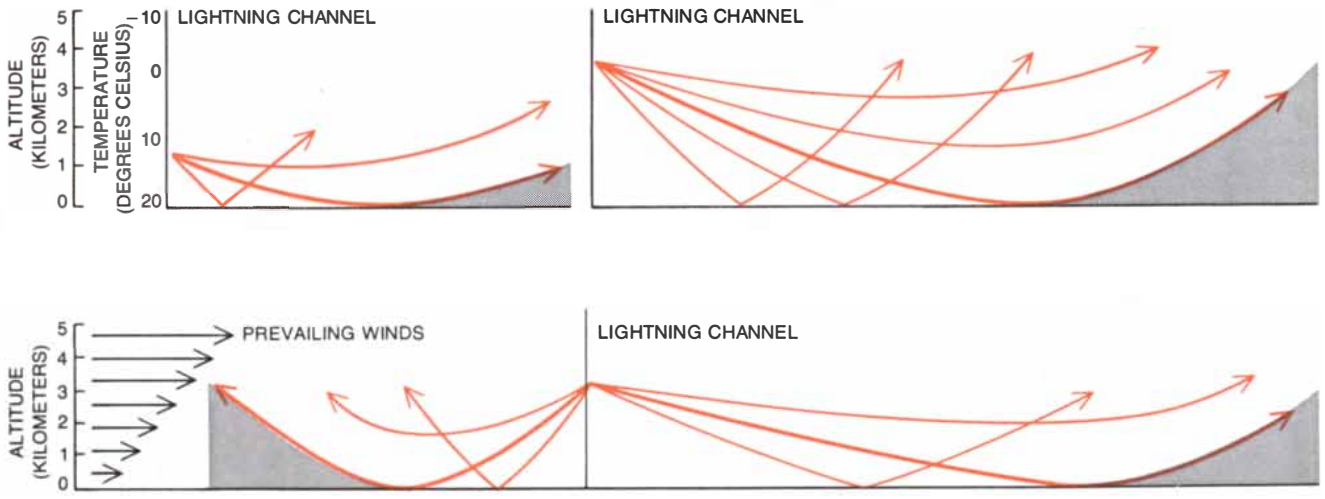
The refraction of thunder is a large-scale phenomenon caused by variations in the speed of sound in the atmosphere; a refracted "ray" of sound is bent away from the straight-line path between the source and the observer. The laws governing refraction are well understood, and given sufficient information on atmospheric conditions, the curved path followed by the acoustic ray can be calculated. The most important variables are temperature and wind.

In the lower atmosphere temperature ordinarily decreases with altitude, typically at a rate of about 6.5 degrees C. per kilometer. Below a thundercloud the temperature gradient is generally steeper, reaching a maximum of 9.8 degrees per kilometer. Because sound travels faster in warm air than it does in cool air, the temperature gradient tends to curve the acoustic rays upward [*see illustration below*]. For this reason thunder generated by the lowest parts of a lightning channel cannot be heard at a distance, and there is a distance beyond which thunder cannot be heard at all, since all the sound passes over the head of the observer.

Wind has two effects on the propagation of sound waves. First, the actual velocity of a wave front is the sum of the speed of sound in air and the wind velocity. Sound therefore propagates downwind faster than it does upwind, and in a precise analysis of thunder signals this difference must be taken into account. Second, wind shear, the variation of wind velocity with altitude, imparts a further bend to the acoustic rays. Wind velocity generally increases with altitude. In the upwind direction the effects of wind shear add to those of temperature refraction; in the downwind direction they subtract from them.

In combination with attenuation and scattering, the temperature gradient and wind shear impose an ultimate limit on the range over which thunder can be heard. The maximum range can be as little as 10 kilometers, but it is sometimes much greater, depending mainly on the altitude of the lightning channel and the wind velocity.

Reflection adds a final modification to the thunder signal before it reaches the listener. For low-frequency sound the amplitude of a reflected wave is roughly proportional to the angular size of the



REFLECTION AND REFRACTION alter the thunder signal and limit its range. The temperature gradient in the atmosphere bends sound waves upward, so that thunder emitted from the lower portions of the lightning channel can be heard only nearby (*top left*); there is a maximum range beyond which thunder cannot be heard (*top right*). The distribution of sound produced by temperature refraction is altered by the wind, which usually has progressively higher velocity at higher altitude. Upwind of the channel the refraction is increased; downwind it is diminished (*bottom*). Low-frequency sound waves striking the ground are efficiently reflected.

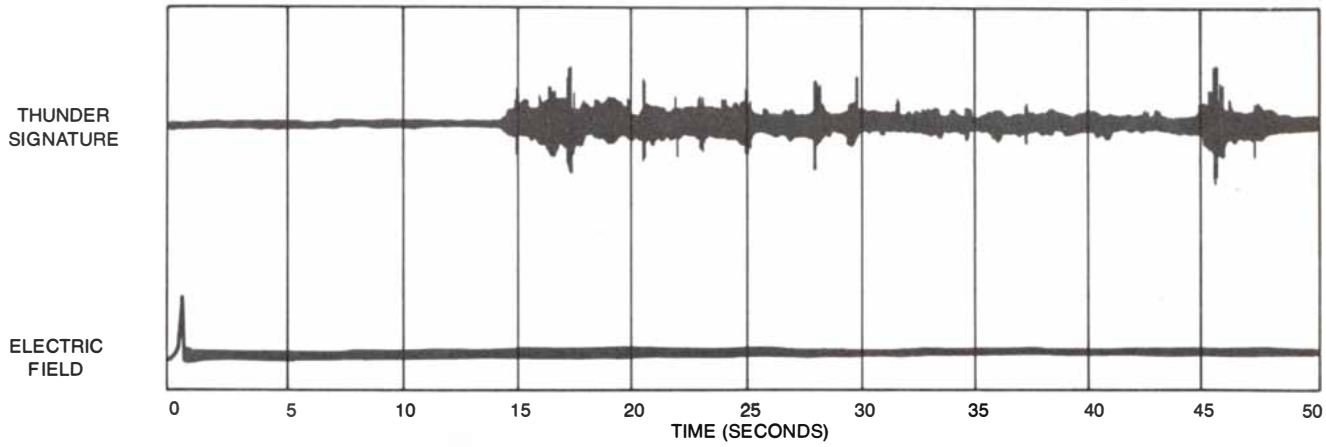
reflecting surface as it is viewed by the observer. In most cases there is only one reflecting surface large enough to produce amplitudes comparable to those of signals received directly: the ground. Since flat ground completely fills the field of view when one is looking downward, the amplitude of the reflected wave will be equal to that of the directly received wave if no sound is absorbed at the surface, which for low frequencies is probably a correct assumption. The perceived sound is thus the sum of the direct and the reflected waves.

Depending on one's height above the ground and the angle of the incident signal, the direct and the reflected waves will add constructively for some wavelengths and destructively for others. In recording thunder signatures we avoid destructive interference by placing our microphones at a height equal to a fraction of the shortest wavelength to be measured.

The thunder perceived at any given position is unique to that point of observation. A microphone placed a few meters above an identical microphone at ground level will detect a slightly different signal. If the two microphones are both near the ground but are separated by a horizontal distance of 20 to 30 meters, the recorded thunder signatures will be similar, but minor differences will be distinguishable because the two positions give a significantly different perspective on the mesotortuous elements of the lightning channel. If the microphones are placed 100 meters apart, few details of the recorded signatures will correspond, but the underlying pattern of claps and rumbles will be preserved; the spacing is now comparable to the size of the macrotortuous elements. Microphones separated by more than a few kilometers will record thunder signatures that may have only one or two features in common.

A thunder signature contains a great deal of information about the lightning channel that produced it and about the atmosphere between the lightning and the observation site. The signature is a complex waveform, but it is possible to analyze it and recover much of the information. When that is done, one can reconstruct the lightning channel, a particularly valuable capability when the channel runs inside a cloud and photography and other optical methods of study are not possible.

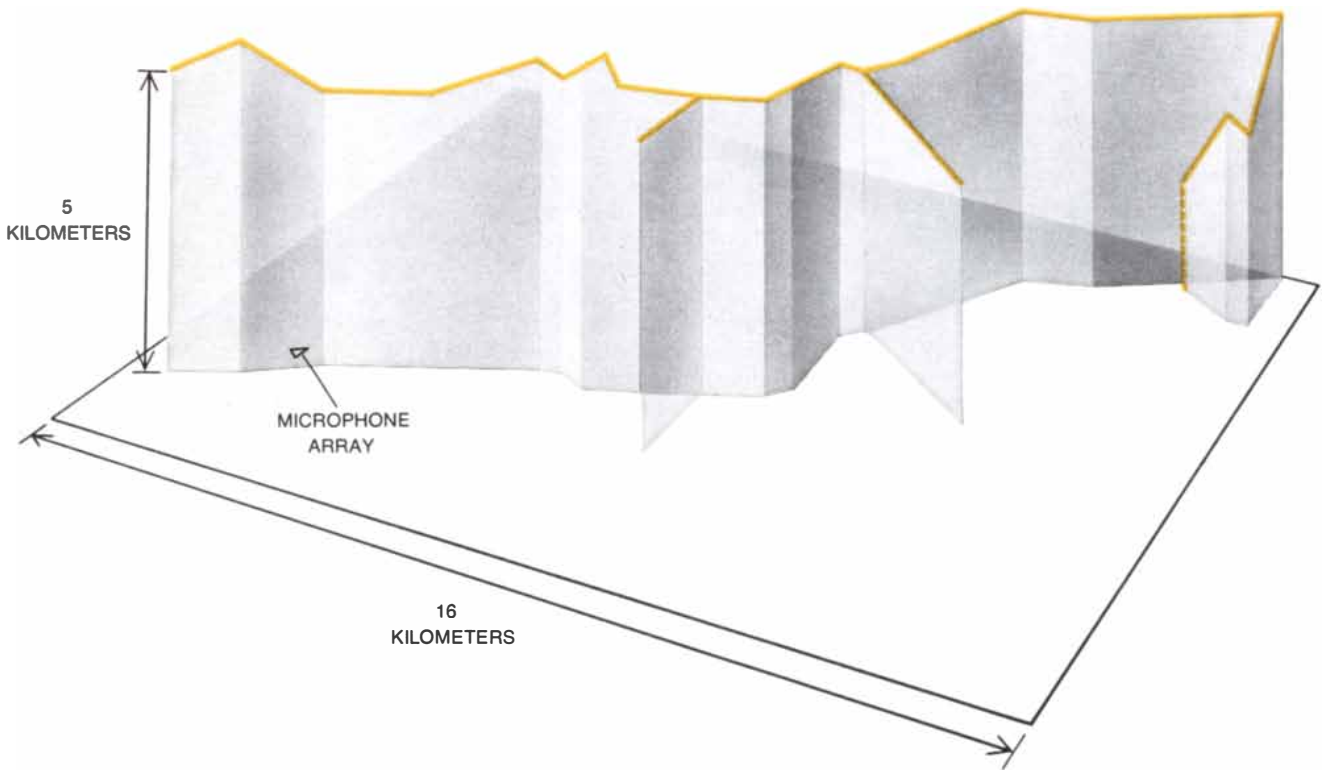
The thunder signature is recorded by an array of microphones on flat ground. For the purposes of analysis the thunder



THUNDER SIGNATURE recorded by one microphone in an array of several microphones contains detailed information about the lightning channel that produced it. The time of the lightning flash is recorded by the momentary change in the electric field at the beginning of the graph. From the time of arrival of each feature of the thunder signature the distance to the corresponding segment of the channel can be calculated; with information from additional microphones the direction can also be determined and hence the location of the channel. Corrections must also be made for effects introduced by the wind and by temperature variations in the atmosphere.

signal is broken up into short sections, each one-fourth of a second to half a second long. The microphones measure the direction of arrival of each section, and in addition the time of the lightning flash and the time of arrival of each section of the signature are known. With the aid of a digital computer and a mathematical model of the atmosphere it is then possible to determine the origin of each section, that is, to deduce its position at the time of the lightning flash. In this way the entire channel can be reconstructed in three-dimensional form. The technique is sensitive enough to locate the main channel and some of the larger branches, but most small branches are unrecoverable because the thunder produced by them has too small an amplitude.

By studying the acoustic record of an entire thunderstorm we have been able to compile a graphic history of the large-scale electrical discharges during the lifetime of a storm. From the same data we have derived information on the process by which the cloud acquires elec-



RECONSTRUCTION of a lightning channel was derived from the thunder signature at the top of the page. Most of the channel is horizontal and is hidden from visual observation by cloud. The macrotortuous segments of the main channel and the larger branches are recorded, but smaller features cannot be resolved and small branches cannot be reconstructed because the thunder they produce is too feeble. Thunder from lowest portion of the channel (*broken line*) could not be detected at the microphone array because of refraction.

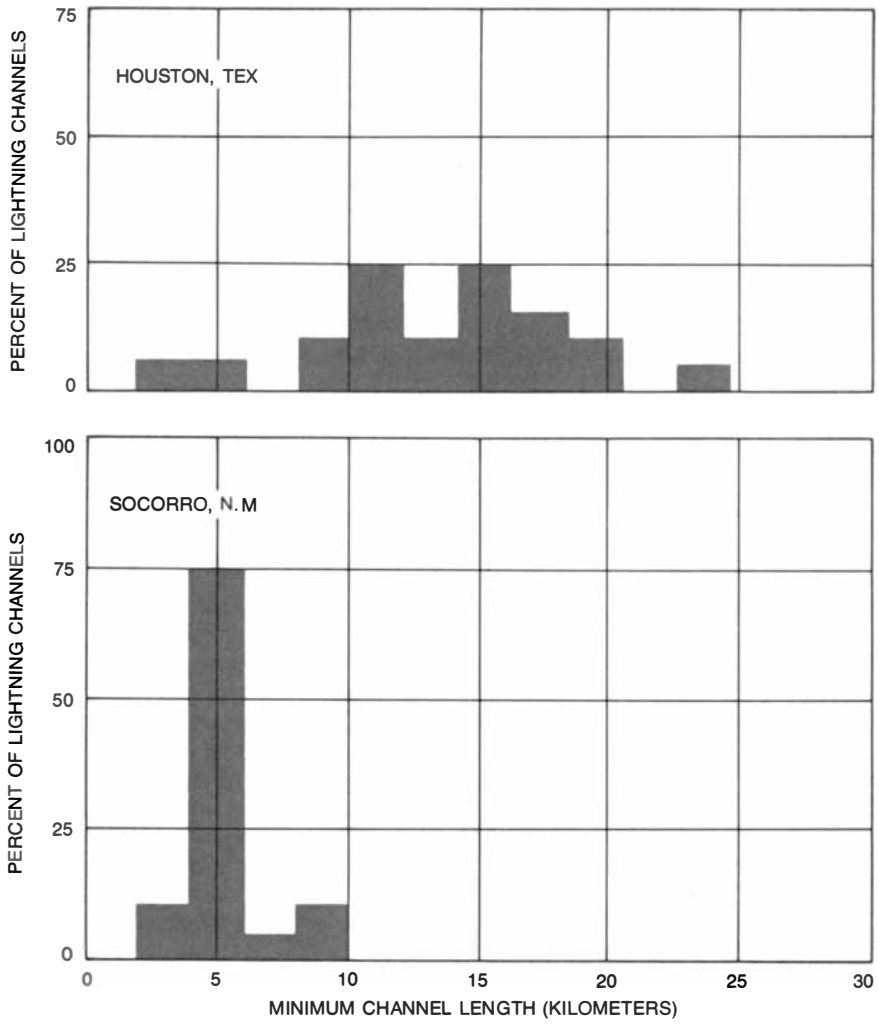
tric charge, the volume in which the charge is stored and the time required to replenish the charge after a lightning stroke. The study of thunder has thus made possible tentative generalizations about electrical activity in the atmosphere and has led to some surprising discoveries about lightning inside clouds.

We have found that intracloud lightning (lightning that does not discharge to the ground) is predominantly horizontal, and so is the intracloud portion of cloud-to-ground lightning. The horizontal lightning channels tend to be aligned so that most of them are roughly parallel.

Our results also show that the negative charge center in the lower part of the cloud is commonly disk-shaped, about two kilometers thick and 10 kilometers in diameter. The positive charge, on the other hand, appears to be dispersed throughout the upper part of the cloud. The negatively charged region is consistently located near the −10 degree isotherm, an indication that the development of the charge is related to the freezing of droplets or to the coexistence of ice and droplets in the same part of the cloud. The development of charge also seems to be correlated with regions of inflow or updraft, which contain small droplets, rather than with areas where raindrops are found.

Lightning early in the life of a storm is confined to the lower, negatively charged region; the upper zone becomes active late in the cycle. Successive lightning flashes draw charge from different volumes of the cloud, but the channels frequently intersect at some point. It is as if one flash took up where the previous one left off. Moreover, lightning in one part of the cloud frequently triggers a discharge in another region. Certain processes important in the physics of clouds, such as the growth of cloud particles, appear to be strongly influenced by the electric field and are correlated with lightning activity.

Finally, the average length of the lightning channel seems to vary with the type of storm. Small, local storms have relatively short lightning discharges (typically about five kilometers long), and all the lightning channels in the storm fall within a narrow range of lengths. In storms associated with large frontal systems, on the other hand, the lightning channels have a broad distribution of lengths and a large mean value (about 15 kilometers). In the larger storms much of the channel length is horizontal.



LENGTH OF LIGHTNING CHANNELS varies from storm to storm. A thunderstorm near Houston, Tex., associated with a frontal system, had a broad distribution of lengths, including many quite long channels. In a local storm at Socorro, N.M., however, there were no very long channels and the great majority were clustered around a single value. The lengths measured were the minimum possible lengths, as determined by duration of the thunder signal.

The complete reconstruction of a lightning channel requires arrays of microphones sensitive to low-frequency sound, equipment capable of recording the sound and a computer. It is possible for the amateur observer, however, to recover some of the information from the thunder signature with apparatus no more elaborate than a wristwatch. You can obtain the approximate distance to the source of a feature in the thunder signature by multiplying the time elapsed between the lightning flash and the arrival of the acoustic signal by the speed of sound. (The approximate distance in kilometers is given by the time in seconds divided by 3, the distance in miles by the time in seconds divided by 5.) Similarly, you can estimate the minimum length of the channel from the total duration of the thunder, again multiplying the elapsed time by the speed of sound [*see illustration on next page*].

With careful observation you may be able to distinguish some of the individual events that make up a cloud-to-ground lightning flash: the creation of the tortuous, branched channel forged by the stepped leader, the brightening of the channel with the first return stroke and a flickering produced by multiple return strokes. Occasionally a lightning flash forks when a dart leader deviates from the path of the original channel.

During a thunderstorm at night you might also want to try photographing lightning. Set up the camera on a tripod and point it toward the most active region of the storm. Close the iris to the smallest aperture possible (that is, set the lens to the largest *f* number), focus at infinity and make time exposures of

from 20 to 30 seconds. Typically one in every three or four frames will contain a lightning photograph, although the number depends on the activity of the storm.

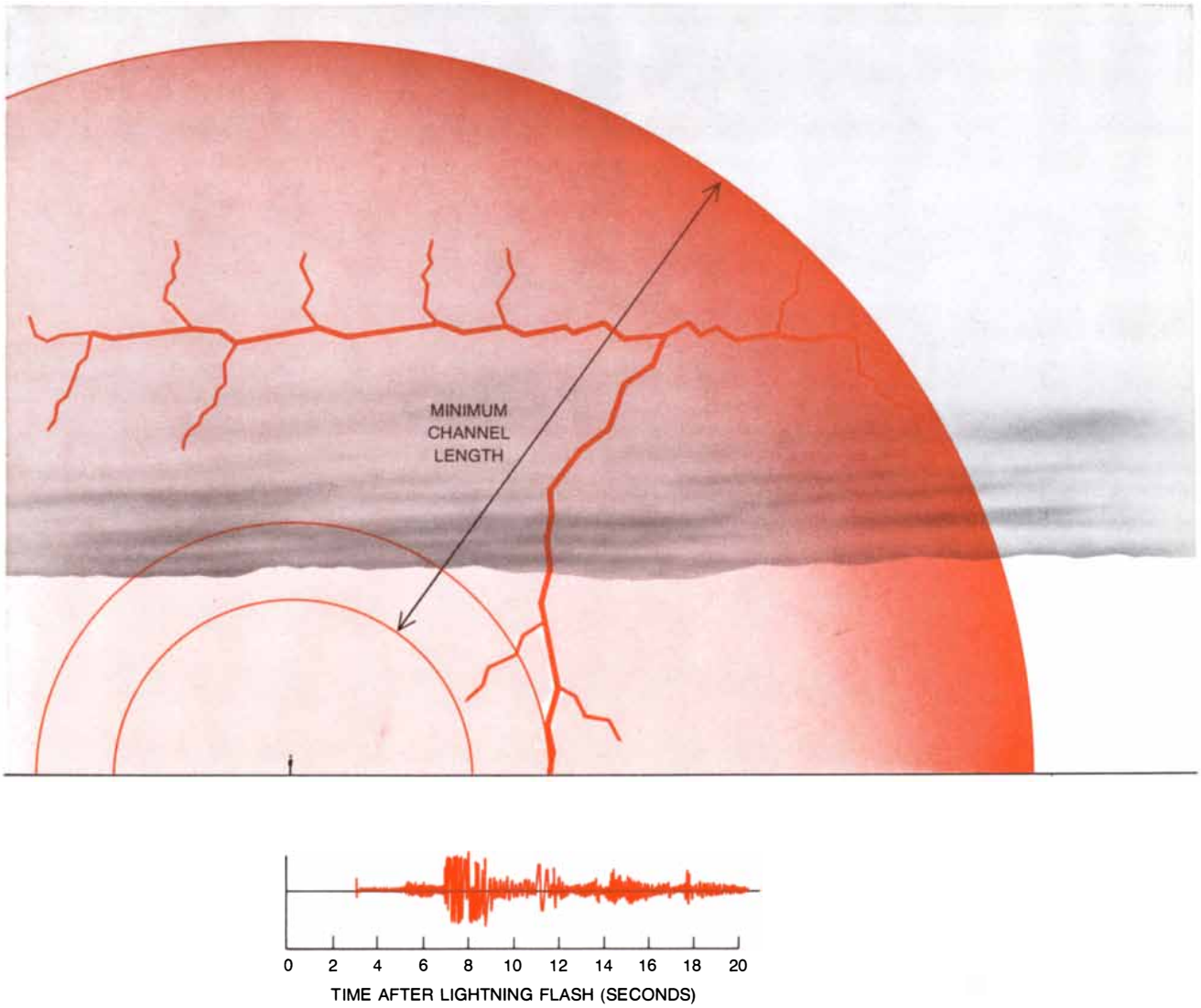
More information about the lightning channel can be obtained through careful attention to the thunder signature. Measure the delay between the flash and the first thunder heard, the loudest clap and the final rumble. From these times you can estimate respectively the distance to the branch nearest you, to the main channel and to the farthest branch. Also note the total duration of the thunder in order to calculate the minimum length of the channel.

A discharge that has struck the ground nearby gives rise to a loud crack, sometimes preceded by a brief rumble or a ripping noise that probably originates in a small branch extending from the main channel toward the observer. When a nearby flash is made up of several strokes, it is sometimes possible to distinguish the acoustic pulses produced by each stroke. The sound is somewhat like a short burst of machine-gun fire.

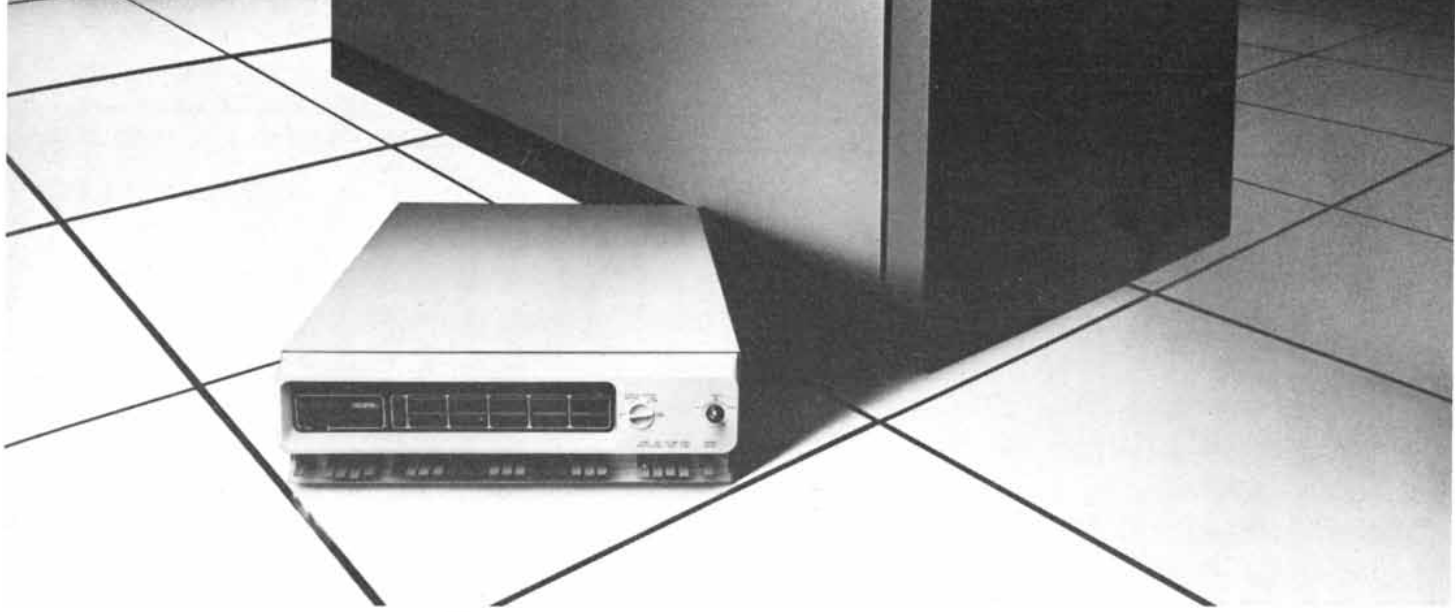
When thunder powerful enough to shake windows is heard as a boom rather than as a clap or a crash, it is usually the product of an energetic flash but a distant or high-altitude one. At a distance of a few kilometers the high frequencies are attenuated with respect to the low frequencies, and the resulting thunder is felt as much as it is heard. In some cases the pitch of the thunder grows progressively lower as sounds arrive from higher or more distant portions of the channel.

Finally, a somewhat rare thunder signal is a ripping noise that can be imagined as the tearing of some cosmic cloth. It is usually attributed to a stepped leader that fails to reach the ground. When the leader does complete a path to the ground, the sound it generates is overwhelmed by that of the subsequent return stroke.

Obviously caution is in order when you are observing a thunderstorm. Do not expose yourself to the risk of personal participation in a lightning discharge by standing in open country or near trees, power lines, fences or other objects likely to be struck. The safest observation post is a closed space, such as a building or an automobile, provided that you avoid touching exterior walls and conducting surfaces.



AMATEUR OBSERVATIONS of thunder require only a wristwatch. By measuring the time elapsed from the lightning flash to the first thunder heard, the loudest clap of thunder and the last rumble, one can determine the distance to the nearest branch of the lightning channel, the main channel and the farthest feature. The approximate distance in kilometers is given by the time in seconds divided by 3. By the same method the minimum length of the channel can be calculated from the total duration of the thunder signal.



You haven't outgrown your big computer. It just needs a little help.

There's no reason for you to go through the trauma of pulling out your big computer and putting in an even bigger one. You can get your big computer a little assistant. A small computer from Data General.

Haggar Slacks brought in a Data General computer to help out their big computer for a lot less than it would have cost to upgrade. Our computer took over a job called "Pattern Layout Optimization." (The computer figures out how to get the most slacks out of every bolt of cloth.)

The Data General computer did the job exactly like the big computer. Just as fast. Just as efficiently. But at a cost so much lower that it paid for itself in less than two years.

And at the same time, our computer took a heavy load off the big computer. So instead of upgrading, Haggar could keep their big computer right where it was.

Lowe's Companies (a group of 130 building materials stores) put our computers in their stores so their salesmen could get instantaneous prices and stock levels. And our computer cost a third less than the IBM 370 it would have taken to do the job.

One of the largest banks in the Northeast put a Data General computer in their trust department to give their portfolio officers access to vital investment information in their large IBM computer. And our computer does the job for much less than the communications costs incurred when the large computer did it alone.

There are hundreds of other ways you can profitably use Data General computers. Sixteen of them are described in our brochure, "The Sensible Way to Use Computers."

Write for a copy.

You may discover that instead of upgrading to a bigger computer, you'll be better off buying a small one.

NAME SCI-775

TITLE

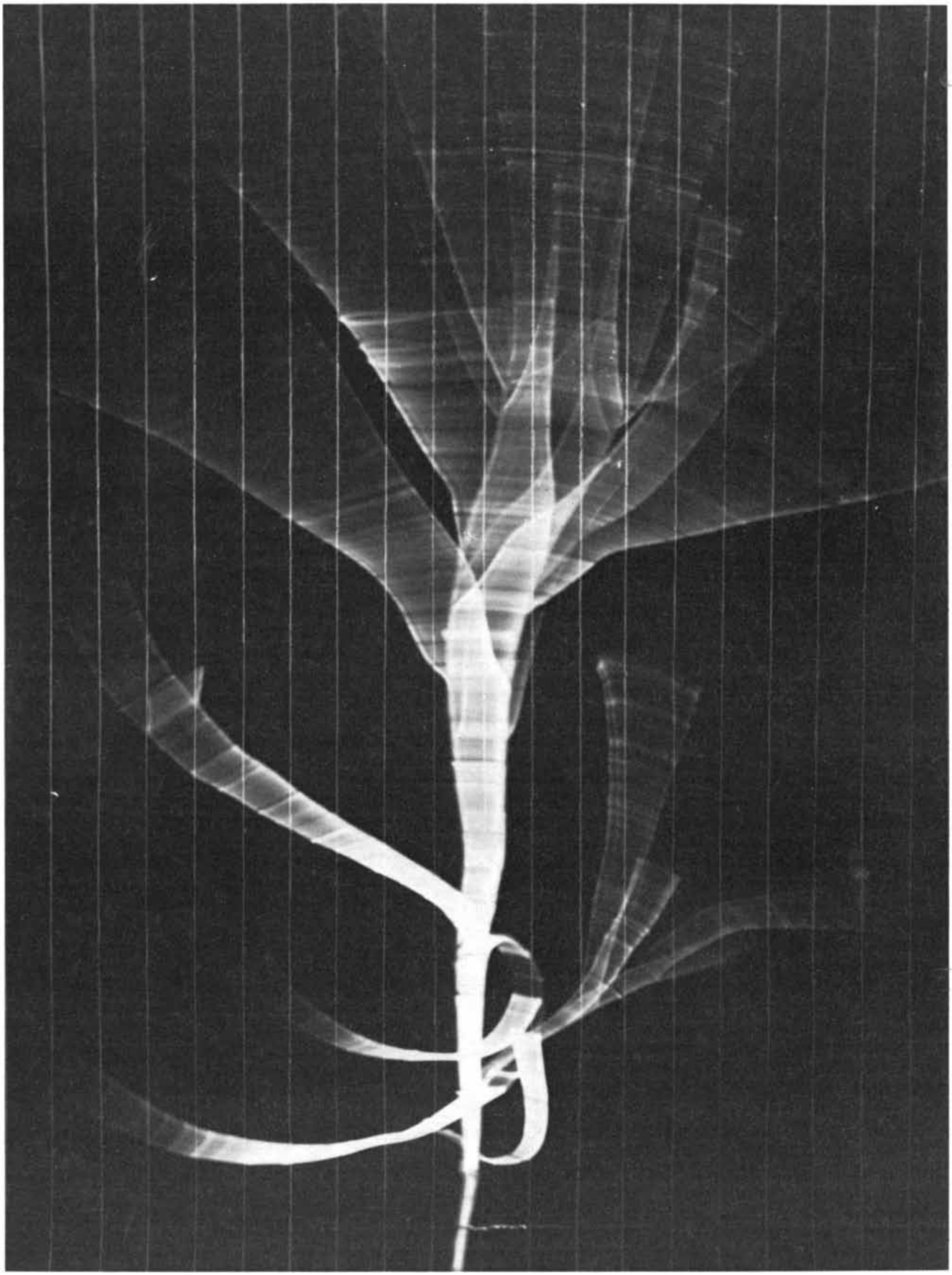
FIRM

ADDRESS

CITY STATE ZIP

DataGeneral

Data General, Dept.H, Route 9, Southboro, Mass. 01772 (617) 485-9100. Data General (Canada) Ltd., Ontario, Data General Europe, 15 Rue Le Sueur, Paris 75116, France. Data General Australia, Melbourne (03) 82-1361/Sydney (02) 908-1366.



YOUNG WILLOW TREE in this time exposure is being shaken at the rate of 90 cycles per minute. The experiment is designed to test a prediction of the author's model related to the principle underlying the mechanical design of trees. The prediction is that the natural frequency of vibration of a branch or stem will be inversely proportional to the square root of the length of the unit tested.

The Mechanical Design of Trees

As the trunk and branches of a tree grow they get not only longer but also thicker. What law governs the relation between the two dimensions? The question is examined by hypothesis and experiment

by Thomas A. McMahon

The branching forms of trees are deeply pleasing to us and would seem to serve the needs of the trees themselves. In their proportions there is a harmony that makes us wonder if we could discover a principle for their mechanical design. Out of the multitude of natural principles that might dictate the diameter of a given limb or stem, could there be a single most important one?

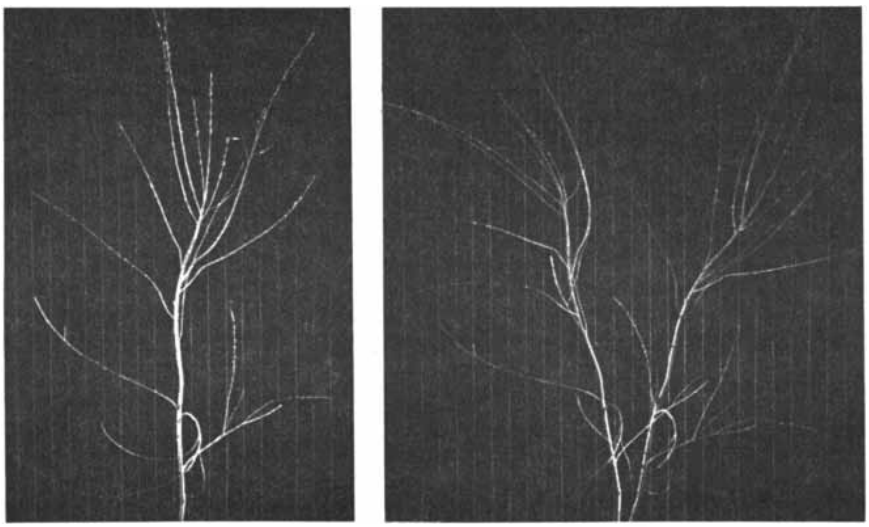
Usually, of course, there is a single most important design principle both in nature and in engineering. In choosing the diameter for the cables of a suspension bridge the engineer must make them large enough to support the load. The struts of the same bridge are more difficult to design, because the mechanism of failure may be either buckling or yielding in compression, but a short effort with pencil and paper will settle which of the two presents the greatest threat, and the bridge can be designed to avoid the most obvious danger.

When the bridge is finished, it may turn out to be considerably more complicated as a system than it is as a mere assemblage of parts. For example, winds may excite unforeseen vibrations; few who have seen the motion pictures of the Tacoma Narrows Bridge breaking up in a high wind will ever forget the spectacle. After such a lesson everyone generally agrees that the only good design is one that considers the mechanical interactions of the entire structure, not just the stresses that might be placed on each part of it. Intuition tells us that the mechanical proportions of trees may be based on just such an integrative principle. In what follows we shall explore this possibility in new theories and experiments, but first let us review some of what is already known about how trees grow.

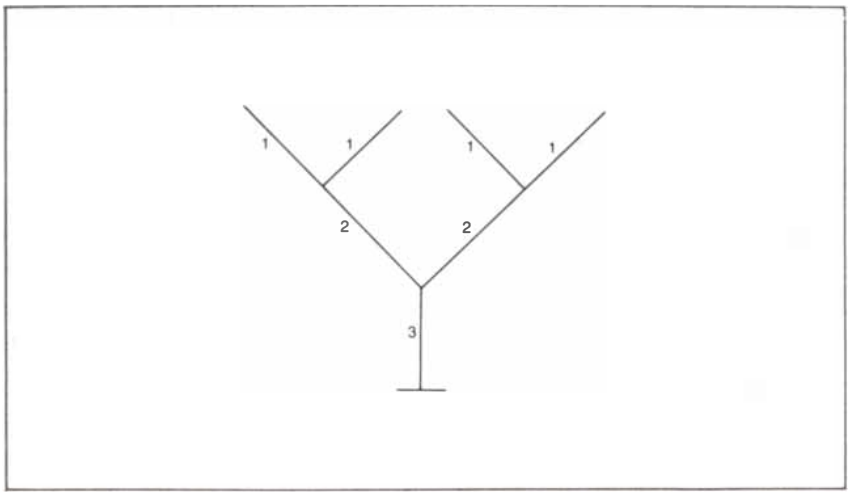
The growth and development of trees, and of plants in general, depends on the effects of various plant hormones and inhibitors. Three major classes of hormones are now recognized: the auxins, the gibberellins and the cytokinins. The hormones that control the predominant factor in tree growth, which is to say the synthesis of new wood, are the auxins. The most important auxin, beta-indolylacetic acid (IAA), is synthesized in apical plant tissues: the tips of shoots.

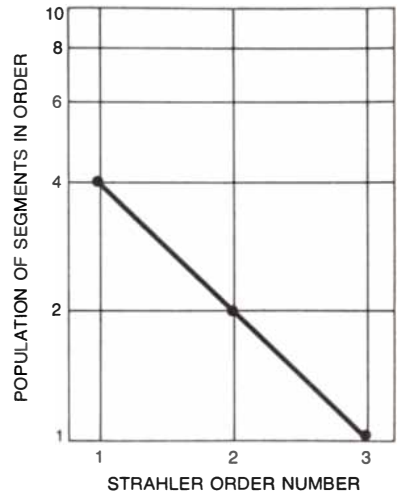
Long before the existence of auxins was recognized their action on growing plants had been examined. Late in the 19th century Charles Darwin and his son observed the preferential bending of grasses toward a light source. In this century the cause was determined to be a greater concentration of auxin in the tissue on the less illuminated side of the plant stem. It was also found possible to reproduce the effect in total darkness. If the tip of a plant stem is removed and auxin is supplied to one side of the stem only, cellular growth on the side of the stem deprived of auxin is slowed. Because the undeprived side continues to grow normally the stem becomes bent.

Like hormones in general, auxin moves from the site where it is synthesized to certain target tissues: the parts of the plant that react to it. In pioneering experiments in the 1920's F. W. Went of the University of Utrecht proved that auxin traveled in a preferred direction and that the direction of travel was independent of both gravity and diffusion. Working with oat seedlings, Went first removed the tip of the coleoptile, or first leaf. Then he removed an additional segment of the seedling stem. When he inverted the stem segment and replaced it and the tip on the stalk, the auxin would not travel through the inverted segment even though gravity favored movement

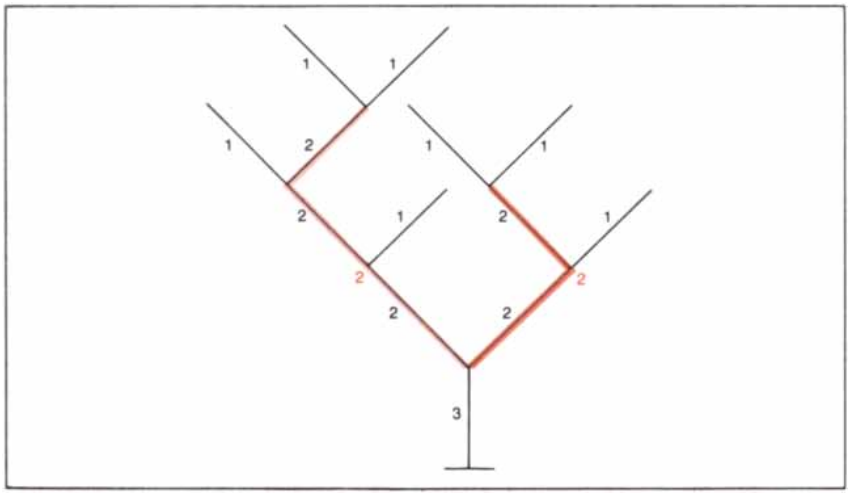


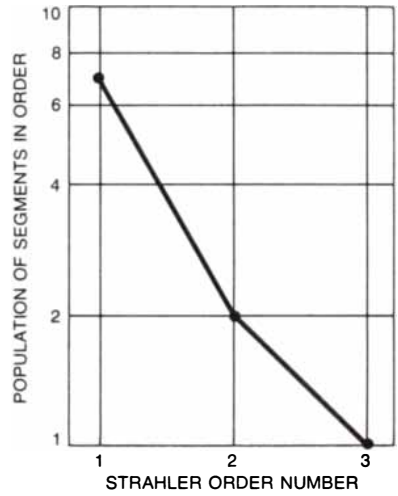
SAME WILLOW is seen at rest (*left*) and at the start of the testing process (*right*). A stroboscopic flash has frozen the image of the tree at both extremes of its motion. Trees, when they are swayed from side to side in this manner, reveal their natural frequency of vibration.





SYMMETRICAL PATTERN of ramification, characterized by the bifurcation of each successive branch (*left*), produces a level descending slope when the population of segments within an order as determined by the Strahler system of ordering is plotted on a semilogarithmic graph (*right*). Starting at the top, the ramifications consist of four first-order segments, two second-order segments and one third-order segment. Strahler's system, devised to describe drainage networks, is applicable to any kind of branching pattern.





ASYMMETRICAL PATTERN of ramification (*left*) also tends to produce a level descending slope when the population of segments within an order is plotted on a semilogarithmic graph (*right*). From the top down the ramifications include seven segments of the first order and one of the third order. The second-order array contains five links, but under the Strahler system of ordering links of the same order, if adjoining, count as one segment only. The second-order array thus contains five links but only two segments (*color*).

in that direction. When he restored the inverted segments to their natural position and again replaced the tips, the transport of auxin was resumed. The rate of transport, about 10 millimeters per hour, was too fast to be the result of diffusion alone.

Where trees are concerned one of the target tissues for auxin is the cambium, the thin layer of living cells that lies just under the bark of tree trunks and limbs. Both the trunks and the limbs of trees increase in girth as a result of cambial activity that promotes the growth of new woody tissue in the underlying xylem layer.

The effect of auxins on cambial activity was also observed long before the discovery of the hormone. The German botanist Ludwig Jost found in numerous experiments during the 1890's that cambial activity was stimulated by growing buds and young leaves. When these parts of a tree were removed, the diameter of the tree did not increase. After the auxins became known the study of their effects continued. For example, observations of many trees show that as spring advances, cambial activity is first evident in the twigs. The activation progresses slowly inward from the twigs along the limbs and trunk for a period of days until finally the tree is increasing in girth over the full length of the trunk.

In experiments where buds or twigs are removed from a tree the cambium remains inactive and no new xylem is formed, with one significant exception. At points on the tree where buds grow directly out of the bark the immediately underlying cambium is found to be active.

In addition to its seasonal role auxin plays a part in the day-to-day life of trees. Consider a forest tree that has been shifted from its normal upright position during a storm. Eventually it will probably grow straight again. A kind of growth known as reaction wood forms along one side or the other of the displaced stem and expands or contracts powerfully enough to restore the tree to an upright position. Experiments indicate the involvement of auxin in the bending process. For example, when synthetic auxin is applied to test specimens of pine, reaction wood grows and

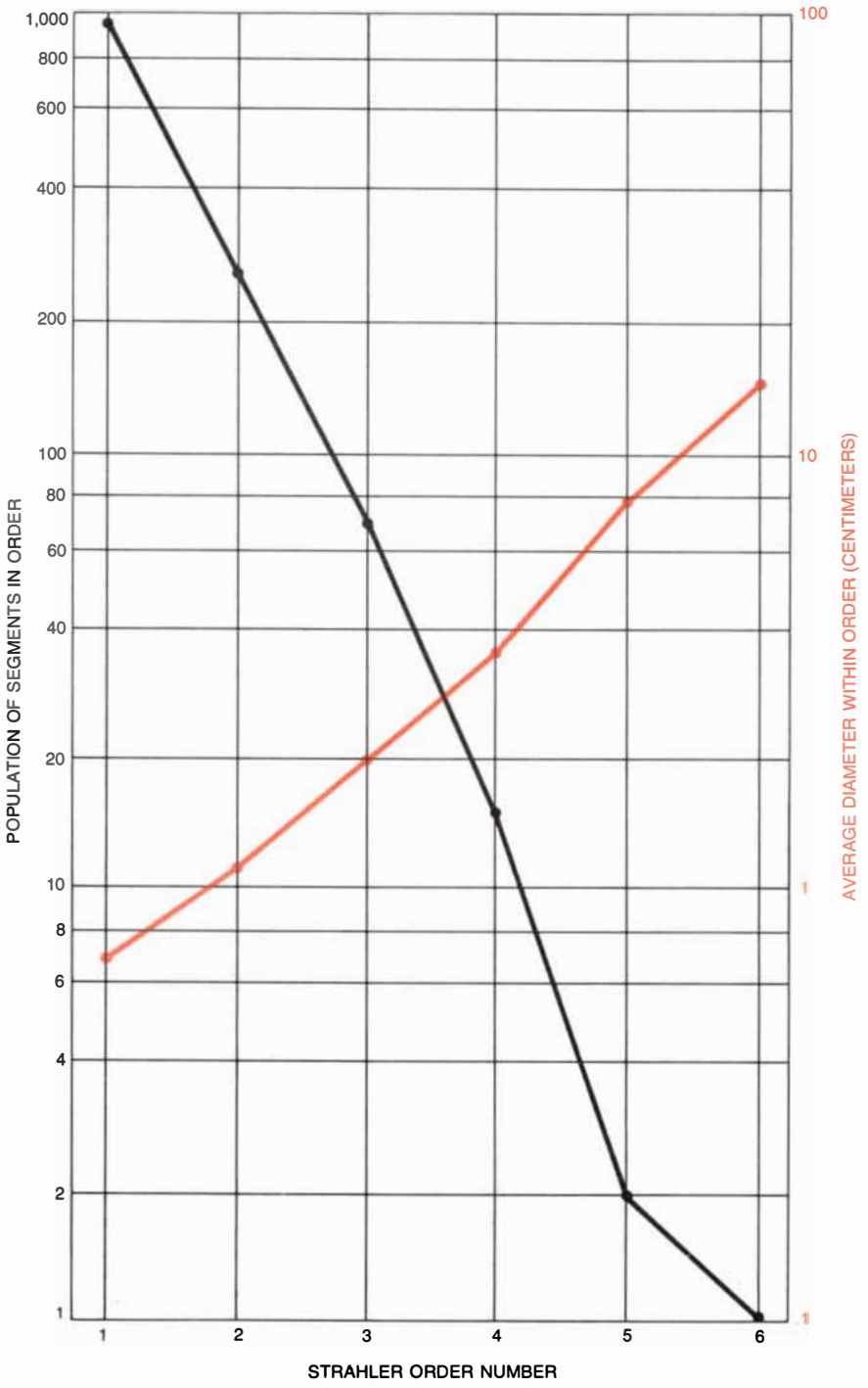
expands at the site of application until the pine stem bends away from the growth site. A similar unequal distribution of natural auxin has been measured in tree branches. The concentration of the hormone is found to be higher in the lower half of the limbs than in the upper half.

All of this suggests that in searching for something that controls the increasing diameter of tree trunks and branches one is well advised to seek a mechanism that is mediated by plant hormones. It is not difficult to understand the function the mechanism fulfills: the goal is the production of a durable tree. But what is it that signals the need for increased girth? And how is an increase recognized as being adequate and the growth signal then attenuated or turned off? It is at this point that analytical models related to the forms and proportions of trees become pertinent. If we are to discuss the strength of a structure and the loads the structure must bear, it is necessary to establish certain facts about the structure itself.

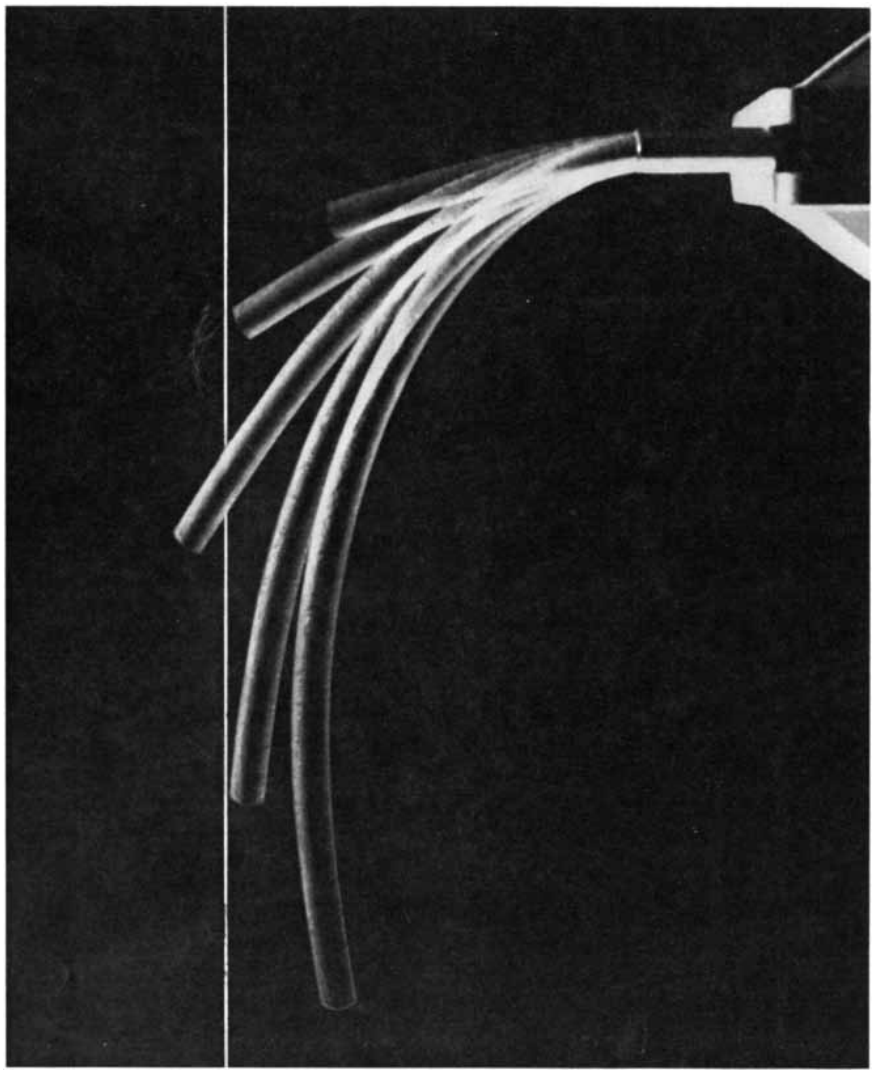
For centuries observers of nature have wondered if common rules could be found that would apply to such fundamentally different kinds of ramification as branching in trees, in watershed drainage systems and in anatomical structures such as blood vessels and bronchial tubes. Over the past few decades the ordering systems that were first developed by geomorphologists to describe drainage networks have been much used in quantitative studies of branching patterns. One of the most widely applied systems was devised by the geomorphologist Arthur N. Strahler.

Starting at the point of ultimate ramification, Strahler assigns to these lowest-order units, which we can call links, the numeral 1. In the description of a drainage pattern these units are the initial feeder streams; in the description of a tree they are the outermost twigs. Whenever two such first-order links meet, they combine to form another link that continues onward. (In a drainage pattern the continuing link is in the direction of the main stream; in a tree it is in the direction of the trunk.) The continuing link is of the next higher order in the system; it is assigned the numeral 2. When, as frequently happens in an asymmetrical pattern of ramification, a first-order link meets a second-order link, the continuation is assigned the numeral of the higher order. Thus the joining of either two first-order links or a first-order link and a second-order one produces a continuation in the form of a second-order link. The joining of two second-order links in turn produces a third-order continuation and so on until the main stream, or the trunk of the tree, is reached [*see lower illustration on opposite page*].

Strahler's system is further simplified, topologically speaking, by the introduction of what can be called the segment concept. For example, in the asymmetrical pattern in the lower illustration on the opposite page, each single first-order link is also a segment. So too are any adjoining links that belong to the same order. Thus seven first-order links are illustrated, and each also ranks as a seg-



SMALL POPLAR TREE, analyzed according to the Strahler system, has 960 first-order segments, 256 second-order segments, 68 third-order segments and 15 fourth-order segments that arise from a main stem bearing only two branches. As with the asymmetrical model shown on the opposite page, when the populations of segments are plotted on a semilogarithmic graph, the descending slope is approximately level. As the size of each population decreases moving from twig to trunk the average diameter of members of each order increases. Ascending slope (*color*) that traces this increase in diameter is also approximately level.



ELASTIC MODELS of tree branches, rubber beams that are the same in diameter but different in length, sag more and more under their own weight as their length is increased. After a critical length (*third from top*) is exceeded the tip of a beam no longer moves farther from the "trunk" (*parallel to line at left*). Instead the sag becomes so great that the beam tip moves closer to the tree trunk (*fourth and fifth from top*). The angle between the horizontal and the chord of a critical-length beam will be the same for beams of any diameter.

ment. The five second-order links, however, comprise only two second-order segments; in one instance two of the links are adjoining and in the other instance three of them are.

The chief utility of this kind of hierarchical ordering system is that it facilitates the recognition of consistent patterns that might otherwise be obscured by the sheer complexity of extended ramifications. One such consistent pattern is the branching ratio, that is, the average ratio between the number of segments that belong to one order and the number of segments in the next higher order. For example, the lowest possible branching ratio in the Strahler system is 2.0 : 1; in order to attain this ratio the ramification must be perfectly symmetrical, so that the Strahler-order numeral changes with each successive bifurcation. Whether or not the branching pattern is symmetrical, however, when one plots the average number of segments in each successive Strahler order as points on a semilogarithmic graph, the points tend to fall on a straight line [*see illustration on preceding page*].

Although the lowest Strahler-system branching ratio is 2.0 : 1, there is no theoretical upper limit to the ratio. The more asymmetrical the ramification, the "deeper" the branching structure and the greater the ratio. Such a structure is said to be deep because the distance between a given segment and the terminal links in the system can vary over a wide range. An everyday example of a deep branching structure is the kind of comb that has a handle. The handle is analogous to the trunk of a tree, the back of the comb to a branch and the comb's teeth to twigs. By adding more and more teeth to the comb one can extend its branching ratio to a theoretically unlimited extent.

Just as the population of segments in an order tends to generate a straight line that rises when it is plotted on a semilogarithmic graph, so do the similarly plotted mean diameters of the segments in successive orders tend to generate a straight descending line. The phenomena that generate both slopes are, to be sure, quite unsurprising. The farther one travels away from the trunk of a tree (or from the main stream when tracing a watershed, or from the heart when tracing a network of blood vessels), the greater the number of ramifications in the system will be. By the same token the dimensions of the segments in each successive order will diminish as the number of segments in the order increases. It nonetheless turns out that the data contain a surprise after all. Work by Michael J. Woldenberg at Harvard University and by Keith Horsfield, Gordon Cumming and their associates at the Midhurst Medical Research Center in England has shown that in several instances the diameter ratio is about half the branching ratio. That is true of pulmonary arteries, of the bronchial network and of several species of trees. Just why there is such a relation between the two ratios remains to be explained.

A branching structure of a kind such that any small portion of the structure is an accurate topological miniature of the whole is called a stationary structure. As far as the study of trees is concerned the principal utility of ordering systems such as Strahler's is that they demonstrate the stationary nature of the branching structure when trees of the same species are compared. In our search for the principle or principles underlying the mechanical design of trees, however, ordering systems offer no help; they are not intended to deal with such quantitative relations as those between the length of a tree limb from the point of its attachment to its tip and the associated progressive decrease in limb diameter. Yet the interaction of these two variables is of great importance mechanically: it determines not only the stiffness and strength of a branch but also its tendency to bend or break.

When one speaks of stiffness, bending and so on, one is actually discussing the mechanical property known as elasticity. This is the property that allows bodies to recover from an imposed distortion in

size or shape or both. All bodies, whether they are solids, liquids or gases, are more or less elastic with respect to size; all solid bodies are more or less elastic with respect to shape.

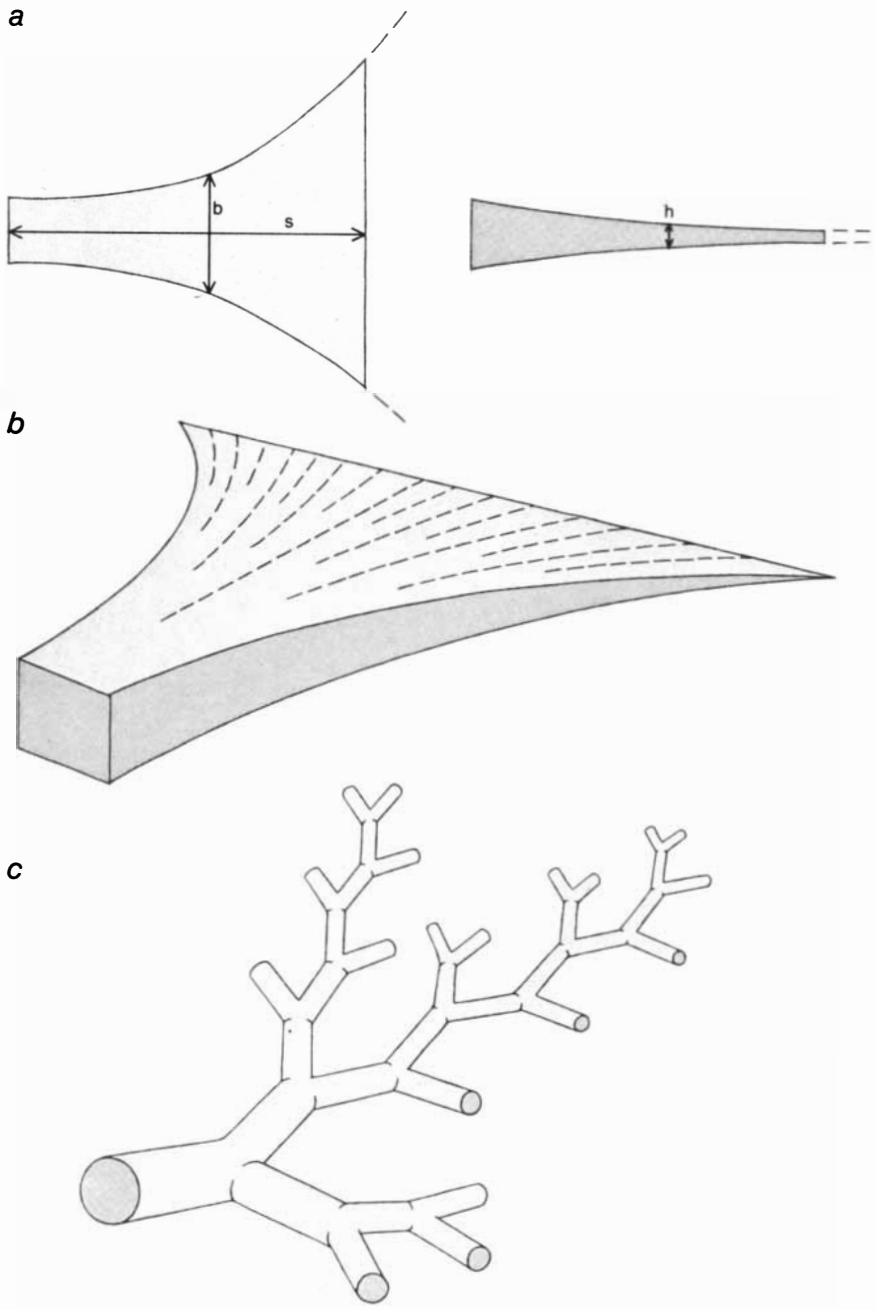
How does elasticity relate to trees? Consider the following experiment. A blackboard occupies one end of a room. A rigid socket is mounted on an adjoining wall so that any cylinder clamped in the socket will extend across the board. We have at our disposal a number of solid rubber cylinders that are identical in all respects but length. Starting with the shortest of these cylindrical beams, we clamp one after another into the socket. We soon discover that the longer the beam is, the more it bends of its own weight. We mark the position of each beam's free end on the blackboard. We are particularly interested in the horizontal displacement of the tip of the beam with respect to the wall where the socket is mounted: it represents the greatest distance from a tree trunk that is available to a leaf. The position is advantageous because leaves in this location escape being in the shade of leaves on higher branches.

As we test beams of ever increasing length we discover that when a beam of a particular length is tested, the horizontal displacement of the tip no longer increases. Beams that are longer than this critical length bend so much that as the beam is lengthened the tip of the beam actually begins to move closer to the wall instead of moving farther away [*see illustration on opposite page*].

If we now use a larger socket and beams that are larger in diameter, keeping the angle of attachment of the socket to the wall the same as before, we achieve the same qualitative results. Quantitatively the critical length is greater; that is because the beams themselves are larger. Nonetheless, the larger beams bend in the same way that the smaller ones do. Consider the droop angle, the angle between a line drawn horizontally to the socket and a line drawn from the socket to the drooping tip of a beam of critical length. Regardless of whether the diameter of a beam is small or large this angle remains exactly the same. It is this very observation, which can be repeated with beams of many different diameters, that gives rise to a generalization with respect to beams that I shall characterize as possessing elastic similarity. Given beams of the same cross-sectional shape, composed of materials of the same modulus of elasticity and the same specific gravity, the generalization states that the diameter of the beams will not be directly proportional to their length but will instead be proportional to the length raised to the 3/2 power, that is, the length times the square root of the length.

Turning from this conclusion that we draw from the branch-beam experiment, let us consider a second experiment involving elastically similar beams. We can call it the trunk-column experiment; the rubber cylinders, identical with those discussed above, are now anchored to the floor of the room, pointing straight up beside the blackboard. Some are short and some are long; some are small in diameter and some are large. If we give the free ends of the columns a sideways push, we shall witness one or the other of two possible outcomes. Either the displaced column will buckle under its own weight and remain bent to one side or it will respond elastically and return to its upright position.

Columns that do not buckle, however, do not all necessarily respond in exactly



"AX HEAD," or tapered-beam, model of tree branches (*a*) relates the decreasing diameter of successive branch ramifications, roughly equivalent to the decreasing thickness of the beam, to the total length of the branch from its tip to the point where the diameter measurement is made. Cuts can be made (*b*) and the model can be unfolded (*c*) so that it more closely resembles the branch of a tree without changing model's static or dynamic features.

THREE LARGE TREES, a giant sequoia (*left*), a Douglas fir (*center*) and a ponderosa pine (*right*), show how the diameter of a tree trunk at its base increases as the height of the tree increases. The mathematical model of elastic similarity requires that the base diameter of a tree increase as the 3/2 power of the tree height; thus taller trees are proportionally thicker than shorter ones. The sequoia has lost its top; its base diameter suggests that undamaged it would have reached a height of 300 feet (*broken line at top*).

the same way. Unless very large damping forces are present, a column in the process of returning to the vertical after displacement will still retain some kinetic energy as it reaches an upright position. This means that it will overshoot the vertical and bend in the opposite direction, then return to overshoot again in the opposite direction, and so on; it will sway back and forth in this way for some time before coming to rest.

We soon note that when we push columns of the same diameter but different lengths, the shorter columns complete a greater number of oscillations in a given time than the longer columns. Now, if a column has such proportions that its tendency to buckle is exactly counterbalanced by its elastic tendency to return to an upright position, it is said to be neutrally stable. As we shall see, the condition of neutral stability in an upright column is analogous to the conditions of critical length in a horizontal beam. This is to say that for any given diameter of a column there is a particular length at which the column is neutrally stable and the number of oscillations following a displacement falls to zero. That critical length differs with different column diameters, but regardless of dimensions a set of neutrally stable columns is elastically similar in the same way that a set of critical-length beams is. Moreover, the diameters of the elastically similar columns prove to be proportional to the 3/2 power of their length. Columns or beams that differ from each other in length by a factor of two will differ in their diameter by two times the square root of 2.

The mechanical fundamentals are now behind us, and we can turn to another kind of tree-structure model. Consider a solid wedge shaped somewhat like the flaring head of a battle ax. Topologically such a wedge is a model of a tree branch that has been folded up much as a folding ruler is. Having established a system of coordinates, we shall use the letter s to designate the distance along a center line running from the wedge's thin edge to a point near its base. Now, like any ax head, the wedge increases in thickness from its thin edge to its base; we shall use the letter h to designate the increasing thickness. Conversely, as the thickness of the ax head increases, its breadth decreases; we shall use the letter b to designate the decreasing breadth.

How are the rates of increase and decrease determined? Given one pair of constants, K_1 and K_2, that vary according to the species of tree involved, and one pair of constant exponents, alpha and beta, with values I shall discuss below, my model proposes that thickness, h, increases proportionally to length, s, raised to the power of beta (a positive value); in this relation K_2 is the constant of proportionality. In turn, breadth, b, decreases proportionally to length, s, raised to the power of alpha (usually a negative value). In this relation K_1 is the constant of proportionality. The tapering edge of the model will be cut off short so that values of length, s, less than some minimum are not admitted. That is because the branching structure of real trees does not continue outward to infinitesimal size. If the wedge were not cut off, it—and the tree branches it represents—would taper down to nothing as s approached zero.

It can be shown that the tapering in breadth is of no consequence in determining the deflection of the tip of a branch under its own weight. The changes in weight that result from breadth tapering are balanced by the changes in stiffness contributed by the same factor. If it is wished to maintain elastic similarity (that is, to keep the deflection of the tip of a branch, divided by its overall length, equal to a constant), that can be accomplished by making beta equal to 3/2, regardless of the values chosen for alpha, K_1 or K_2.

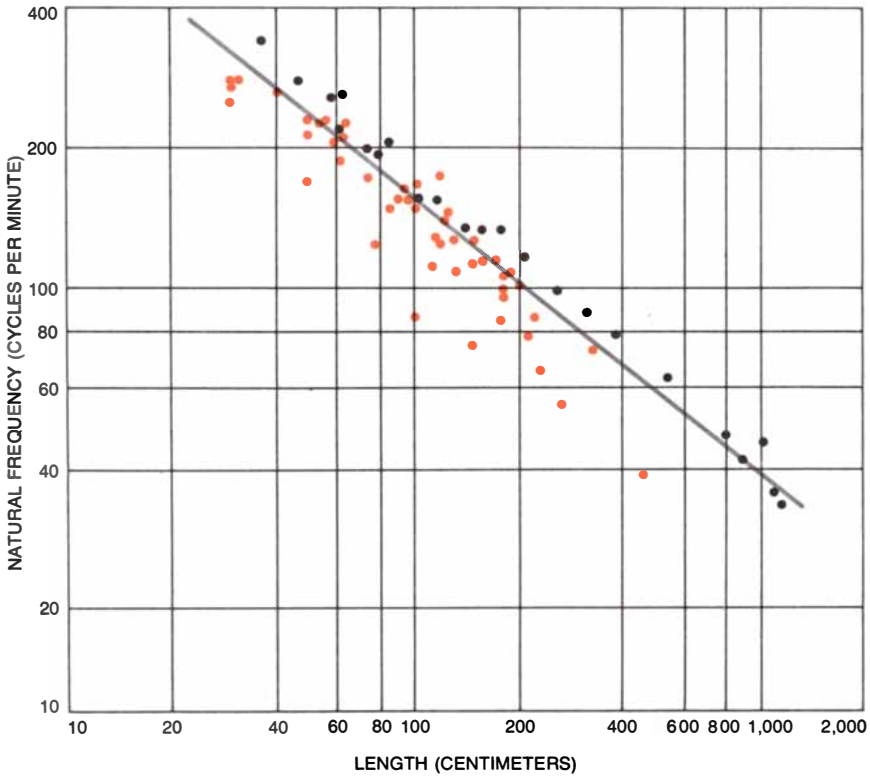
In my elastic-similarity model the measure of thickness, h, corresponds roughly to the increases in diameter of actual tree limbs as the limbs approach the tree trunk. We can now imagine that the base of the ax head has been cemented to the trunk of a tree and that longitudinal cuts, roughly parallel to the direction in which s is measured, have enabled us to unfold the model [*see illustration on page 97*]. What we then see is a tapered branching structure that realistically models an actual tree limb and at the same time retains all the static and dynamic properties of the original model that are of interest to us.

Is the relation between length and diameter proposed by the model actually found in nature? Following the work of the English mathematician A. G. Greenhill, I analyzed a series of height and base-diameter measurements descriptive of specimens of more than 600 species of trees that grow in the U.S.; the data had been assembled by the American Forestry Association. I found that on the average the diameter of a tree at its base was proportional to the 3/2 power of its height. There were, to be sure, great variations in height for any one diameter, but I found that the mean height for each diameter was about 25 percent of the theoretical neutral-stability height for a uniform column of the same diameter. This means that the average tree, being only a fourth as tall as the model column, has a factor of safety with regard to buckling of approximately 4. When an individual tree was particularly tall, of course, the factor was much lower.

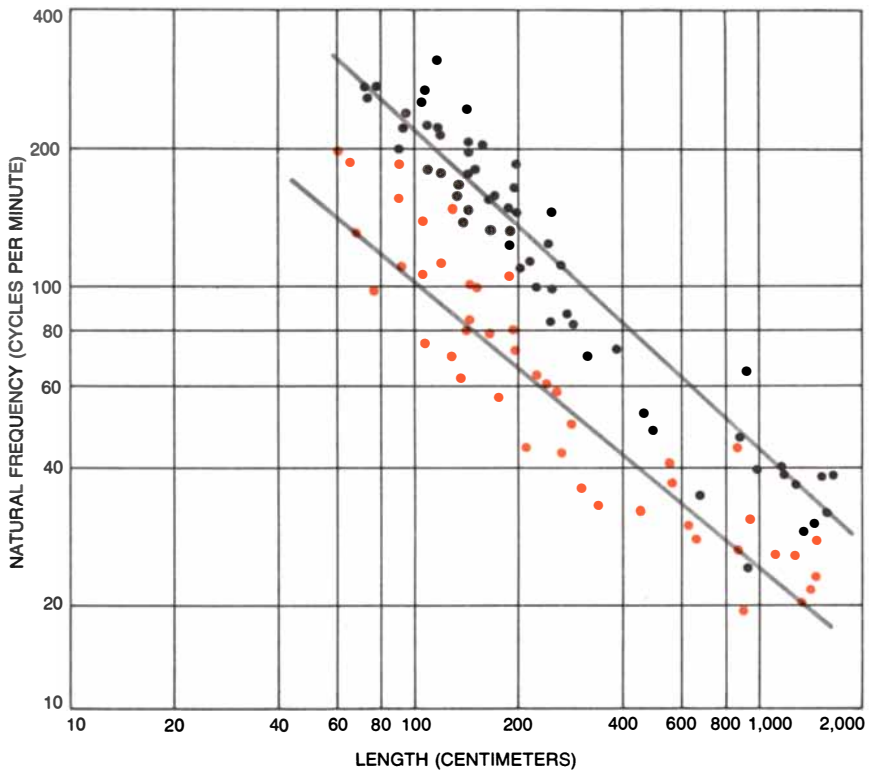
There are at least two alternatives to elastic similarity as a potential principle underlying the mechanical design of trees. One of these is geometric similarity. In such a model the value of beta is set at 1 rather than at 3/2, with the result that two branching structures that differ in length by a given factor will differ in diameter by the same factor. The second model might be called static-stress similarity. Here the value of beta is set at 2. It turns out that regardless of how great the longitudinal extent of my ax-head model is, or at which cross section the stress is sampled, the static tensile stress in the topmost fibers of the beam is a constant. It is interesting to note that the value of beta in the elastic-similarity model, 1.5, lies midway between these alternative values.

Over the past two years, with the help of my students, I have analyzed six trees of modest size that represent four species: white pine, white oak, red oak and cherry. We recorded the length and diameter of every segment in all six trees, using Horsfield's technique of recording so that a computer printout would recreate the topology of each tree. Tracing every path leading from a terminal twig to the main trunk, we tested several alternative descriptions of the relation between local diameters at various points along the path and the distance from each of these points of measurement to the tip of the branch. The best fit to the data was achieved when the value of beta was close to the 3/2 value of the elastic-similarity model. The best-fitting exponents ranged from 1.66 (a small white oak) to 1.37 (a white pine). For one large white oak the exponent was 1.41; for a red oak it was 1.51. For the cherry the fit was exact: 1.50.

A more dynamic method of investigating the mechanical design of trees is the study of their natural frequency of vibration. The oscillations of mechanical systems of springs and masses were analyzed centuries ago by the pioneers of physical science; they soon recognized that the cyclic nature of any such oscillation is a function of the stiffness of



NATURAL FREQUENCY of vibration of a trunk or branch depends on the distribution of stiffness and mass throughout the structure. The graph, indicating the relation between frequency and length for the stems (*black dots*) and branches (*colored dots*) of poplar trees, shows an inverse-square-root relation that is very close to the one predicted by the elastic-similarity model. Tree main stems are mechanically stiffer than branches of the same length.



LEAFLESS TREE, a red maple in this example, turned out to have a natural frequency of vibration two to three times higher (*black dots*) than the same tree did when it was in leaf (*colored dots*). Trees in leaf are heavily damped compared with leafless ones. In either condition, however, the tree evidently still conforms to the proposed model of elastic similarity.

the springs and the magnitude of the masses. The natural frequency of vibration increases as the stiffness of the springs is increased, and it decreases as the magnitude of the masses is increased. It follows that when a tree is shaken in such a way that it sways from side to side (that is, the lowest mode of vibration), the natural frequency of its to-and-fro motion depends on the way both stiffness and mass are distributed throughout the structure of the tree.

The ax-head model of a tree limb predicts that the natural frequency of vibration for a branch bending up and down will be inversely proportional to the square root of the length of the branch (given the value of beta as 3/2). This result is again independent of the choice of value for alpha. The analysis is also valid for small-amplitude shaking of a statically stable tapering column (such as the ax-head model standing upright) and thus applies to whole trees as well as to lateral branches. Small trees should therefore have a higher natural frequency of vibration than large trees of the same species, and individual branches should have a higher frequency than whole trees. The predicted exponent in the frequency-length relation is equal to beta minus 2. By way of comparison, for geometric similarity (beta equals 1) the natural frequency is predicted to be inversely proportional to length and for static-stress similarity (beta equals 2) the natural frequency is predicted to be completely independent of length.

In order to test the prediction of the model, my students and I last year clamped a succession of tree branches in a rigid bench vise, measured the longest path from the point of clamping to the terminal twig and then rhythmically pushed on the branch to make it sway up and down in the lowest mode of vibration. We recorded the number of cycles per unit of time and then clamped the branch at another point and repeated the procedure.

We measured the natural frequency not only of branches but also of the main stems of trees and even of whole trees with and without leaves. The specimens we studied included poplar, larch, red maple, red oak and white oak. We found that the relation between natural frequency and length was close to the prediction of the elastically similar ax-head model: the average power in the frequency-length relation was −.59.

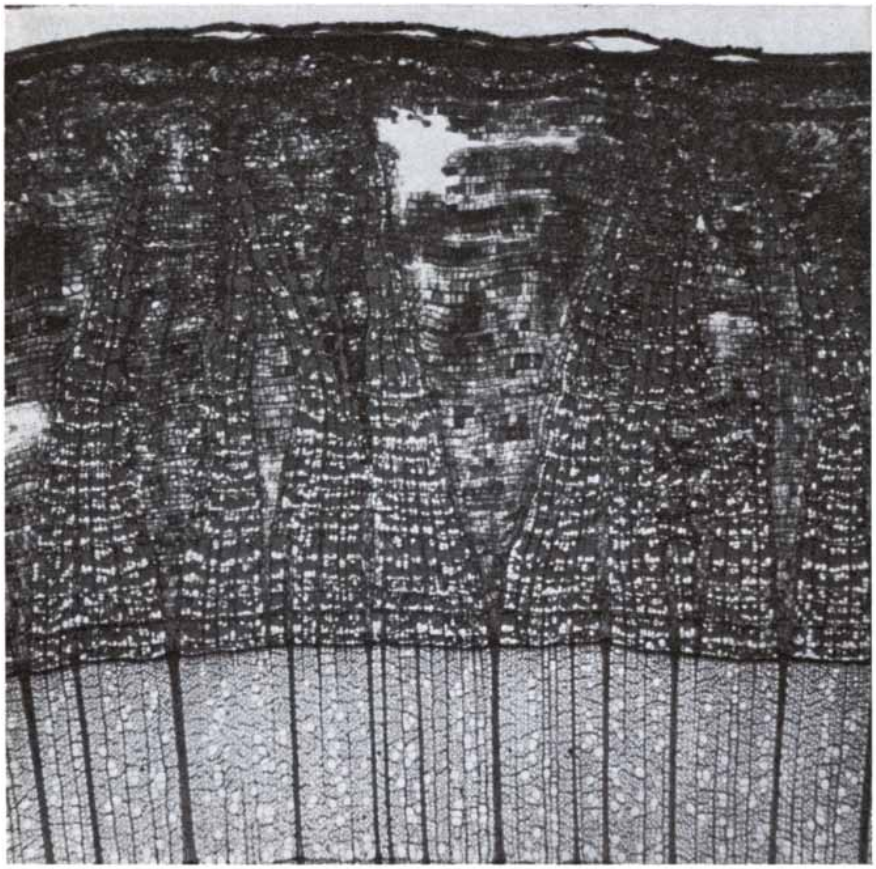
We found that among trees of the same species the natural frequency of the main stem of the tree was slightly higher than the frequency of a branch

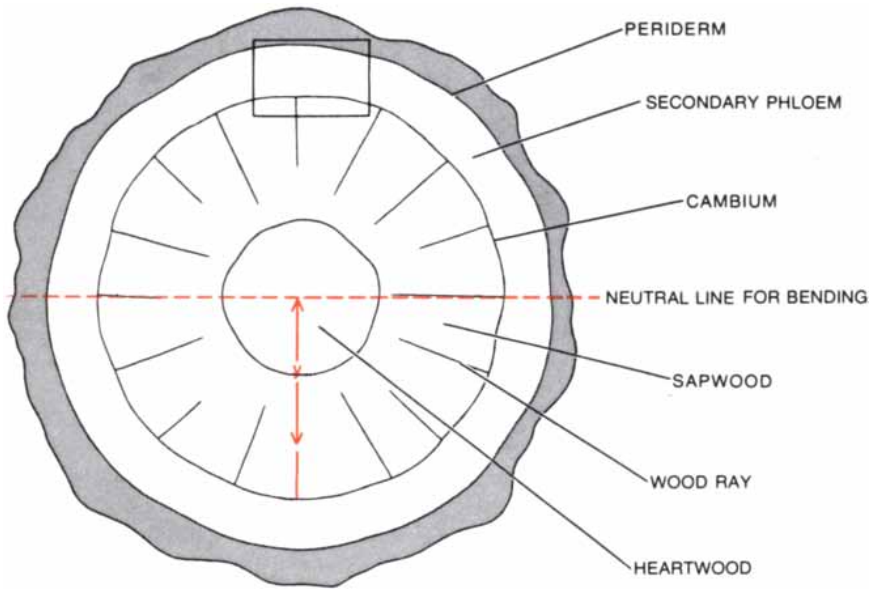
of identical length. This suggests that the design of a main stem follows the same rules as the design of a branch but that the main stem is mechanically stiffer. We also found that the natural frequency of leafless oaks and maples was between two and three times higher than the natural frequency of these species when they are in leaf. Nonetheless, trees in leaf fitted the model as well as leafless ones did [*see bottom illustration on opposite page*].

The vibration experiments proved to be a sensitive test of the value assigned to beta in the ax-head model. As we have seen, when the postulate of elastic similarity was incorporated in that model, it predicted that the frequency of vibration would be inversely proportional to the square root of the branch length. It is not only the close agreement of the experimental data with the model's predictions that allows us to reject the two alternatives to the elastic-similarity model (geometric similarity and static-stress similarity). The first of the alternatives incorrectly predicts that the natural frequency will be inversely proportional to the length rather than to the square root of the length. The second predicts the clearly unrealistic condition that the natural frequency will be completely independent of the length.

The primacy of the principle of elastic similarity is implicit both in the comparison of natural tree morphology with the beam, column and ax-head model studies and in the studies of the relation between branch length, trunk length and the natural frequency of vibration. All our data support the conclusion that the principle underlying the mechanical design of most if not all trees is the maintenance of elastic similarity. This leads to a further conclusion. Something fundamental about the process of growth prevents a tree limb from being deflected by its own weight past some critical point. When the curvature induced by weight exceeds a specific threshold value, it must be true that the limb automatically begins to increase in diameter.

It is conventional wisdom among silviculturists that trees raised in a greenhouse will show greater growth in diameter if given a shaking for a brief interval each day; trees grown under identical conditions but without such stimulation do not grow as thick. By the same token trees that are grown outdoors must not be supported by guy lines for too long; otherwise they will have such tall and slender proportions that they will buckle when the guys are removed. All this information, both quantitative and anecdotal, implies the existence of a mecha-





POSSIBLE SENSING MECHANISM that could inhibit bending of a tree stem or branch in excess of some critical limit is visible in this photomicrograph provided by Martin H. Zimmermann of the Harvard Forest. It shows a transverse section of a linden branch. Cells of the secondary phloem occupy much of the micrograph. The line near the top is the periderm. The lower line is the cambium, an aggregate of cells responsible for the growth of new wood. The cambium includes components known as ray initials; these continue into the sapwood as wood rays (*see diagram*). The wood rays are ideally located to sense bending. Fibers above and below the neutral axis for bending (*color*) are respectively in tension and compression when a branch bends downward. A wood ray perpendicular to the neutral axis could measure the rate of change of these stresses in the direction indicated (*y, color*). If growth of new wood were stimulated when a stress gradient sensed by ray exceeded some set value, radius of curvature in bending would be kept within appropriate limits.

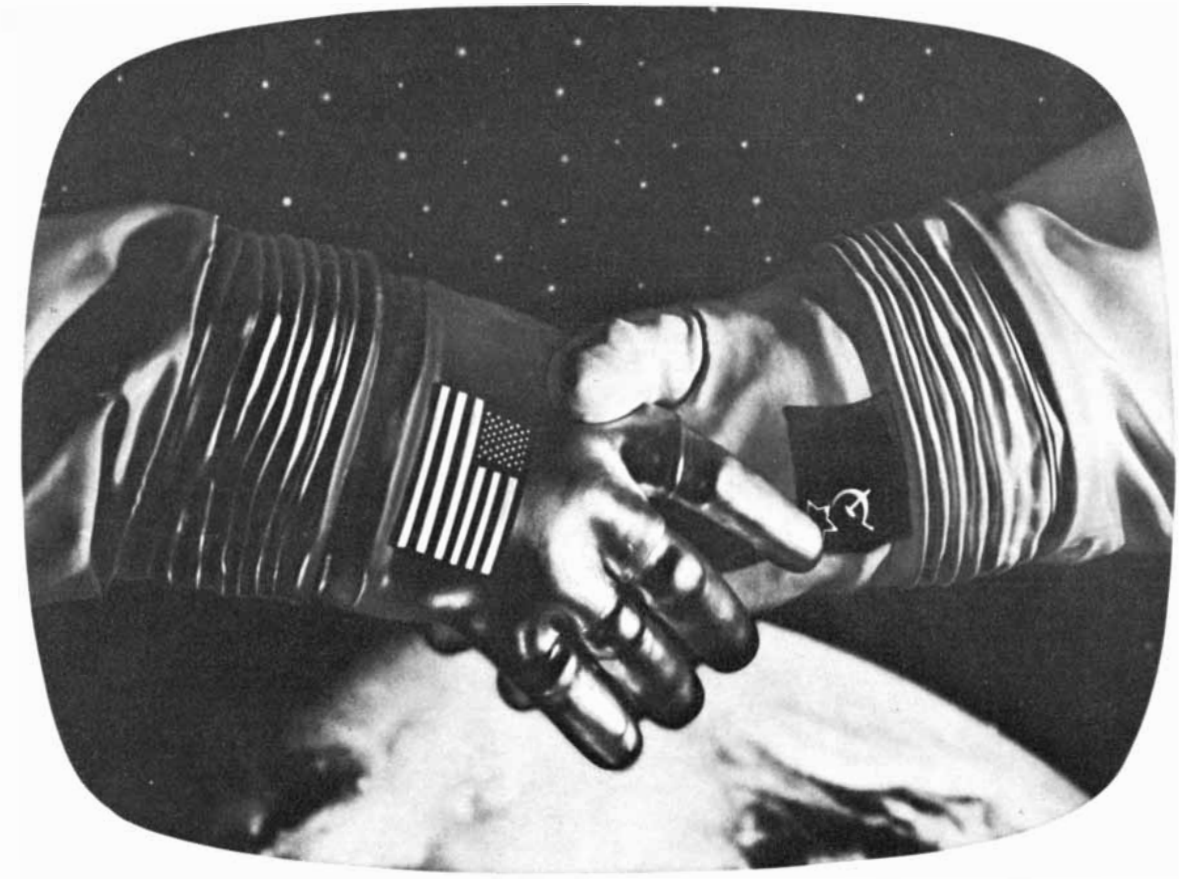
nism that is activated by excessive curvature. How could a property as gross as curvature exercise control over local secondary growth in a tree?

The physical arrangement of the living cells communicating with the cambium layer of woody stems suggests one possible mechanism, analogous to a mechanism proposed by J. D. Currey of the University of York to explain the load-induced remodeling of animal bones. Among the components of the cambium are what are called ray initials; the continuation of a ray initial down into the sapwood of a stem, a branch or a trunk is known as a wood ray [*see illustration at left*]. Now, in a beam that is bending of its own weight the fibers at the top and bottom of the beam are under either tensile or compressive longitudinal stress. The rate of change of stress in a direction at right angles to the plane of the neutral axis (ordinarily the center line) is directly proportional to the local bending curvature of the beam. The wood rays in a tree limb are ideally located for the sensing of such local curvature in a tree. Might there be some mechanism responsive to the gradient in stress along the length of the wood rays? What if, when the gradient in stress exceeded a critical level, extra supplies of auxin were made available to the cambium? Such a response would be a feedback mechanism; by stimulating an increase in limb diameter the mechanism would tend to keep the overall weight-induced curvature of the limb below a fixed threshold value.

This, to be sure, is conjecture. Nevertheless, one important by-product of any mathematical model is the series of new experiments that contemplation of the model suggests. Can it be demonstrated that wood rays are sensitive to a stress gradient? Is auxin released in response to stress-gradient information? Or is it possible that a genetic mechanism, unrelated to the forces of the environment, somehow mediates the maintenance of elastic similarity in trees?

Theories rarely threaten the mystery of natural phenomena. The best theories harmonize with that mystery and make it ever more fascinating. Although it is obvious that trees grow, it is not obvious why they grow the way they do. Part of the answer can now be understood in mechanical terms. Another part must certainly involve the plant hormones that are known to control the secondary growth of the cambium. We shall have to await further developments to know how the parts fit together.

The next giant leap for mankind.



One small step for a man, six years ago, became a giant leap for mankind. Now Soviet cosmonauts and U.S. astronauts prepare for the next giant leap: the union of two spacecraft and two nations 140 miles above Earth.

Sunday July 13, Western Union presents "Union in Space," an exclusive ABC News television documentary on preparations for the Apollo-Soyuz Test Project. Correspondent Jules Bergman traces the progress of U.S. and Soviet space exploration, including an unprecedented visit to Zvezdny Gorodok (Star City), center of the Soviet space program.

Tuesday July 15, "Union in Space" begins 10 days of mission coverage, including live coverage of the Soviet launch, and climaxing when astronauts and cosmonauts rendezvous high above Earth.

The entire ABC News coverage of this historic event is presented by Western Union, pioneer of domestic satellite communications in America.

western union

Western Union presents "Union in Space" on ABC-TV.

Documentary special Sunday July 13, 9:30 p.m. (8:30 p.m. Central). Live coverage July 15-24.

WHY MOSQUITO REPELLENTS REPEL

They are not substances that a mosquito somehow finds distasteful. They jam the mosquito's sensors so that it is not able to follow the warm and moist air currents given off by a warm-blooded animal

by R. H. Wright

In order to understand how mosquito repellents work we must first learn something about how the female mosquito senses and locates the warm-blooded host on which it feeds and then see how the presence of a repellent changes things. At the beginning of my studies I supposed a mosquito "likes" the smell of the host and "dislikes" the smell of a repellent and behaves accordingly. The experiments I shall describe show that this naïve view is completely wrong.

When female mosquitoes in a cage are left undisturbed, they settle on the walls and remain quiescent except for an occasional flight to another part of the cage. A physicist colleague of mine at the British Columbia Research Council, Philip N. Daykin, became curious about these apparently aimless flights. He put about 100 mosquitoes in a cage and waited until they settled down. He then counted the number that had moved after various intervals from an arbitrary time designated zero. He found that as a general rule about half of the mosquitoes in the original population had moved in about an hour.

Daykin's findings reminded him of radioactive decay, which is expressed in terms of the half-life of the radioactive isotope: the time required for half of any given quantity of the isotope to decay. If each mosquito that moves is regarded as being "lost" from the resting population, the rate of loss turns out to be about the same as the rate of decay of a radioactive isotope with a half-life of about an hour.

Since radioactive decay is the result of randomly occurring disintegrations of atomic nuclei, this suggested that the spontaneous flight of undisturbed mosquitoes may also be due to a random process. Nerve cells are known to fire spontaneously. These discharges have a Poisson distribution in time; that is, there are unpredictable short intervals during which there will be an exceptionally large or small number of events. Perhaps a sudden burst of discharges arriving at some kind of command center in the mosquito's brain is responsible for initiating its spontaneous flight.

In another experiment Daykin began with all the mosquitoes at rest on one side of the cage. When about half of those mosquitoes had moved to the other side of the cage, he inserted a partition to separate the apparently active mosquitoes from the apparently sluggish ones. He then recorded the spontaneous flights in each half of the cage and found that the resting population in each group of mosquitoes had approximately the same half-life. This showed that there was really no inherent difference in restlessness between the two groups.

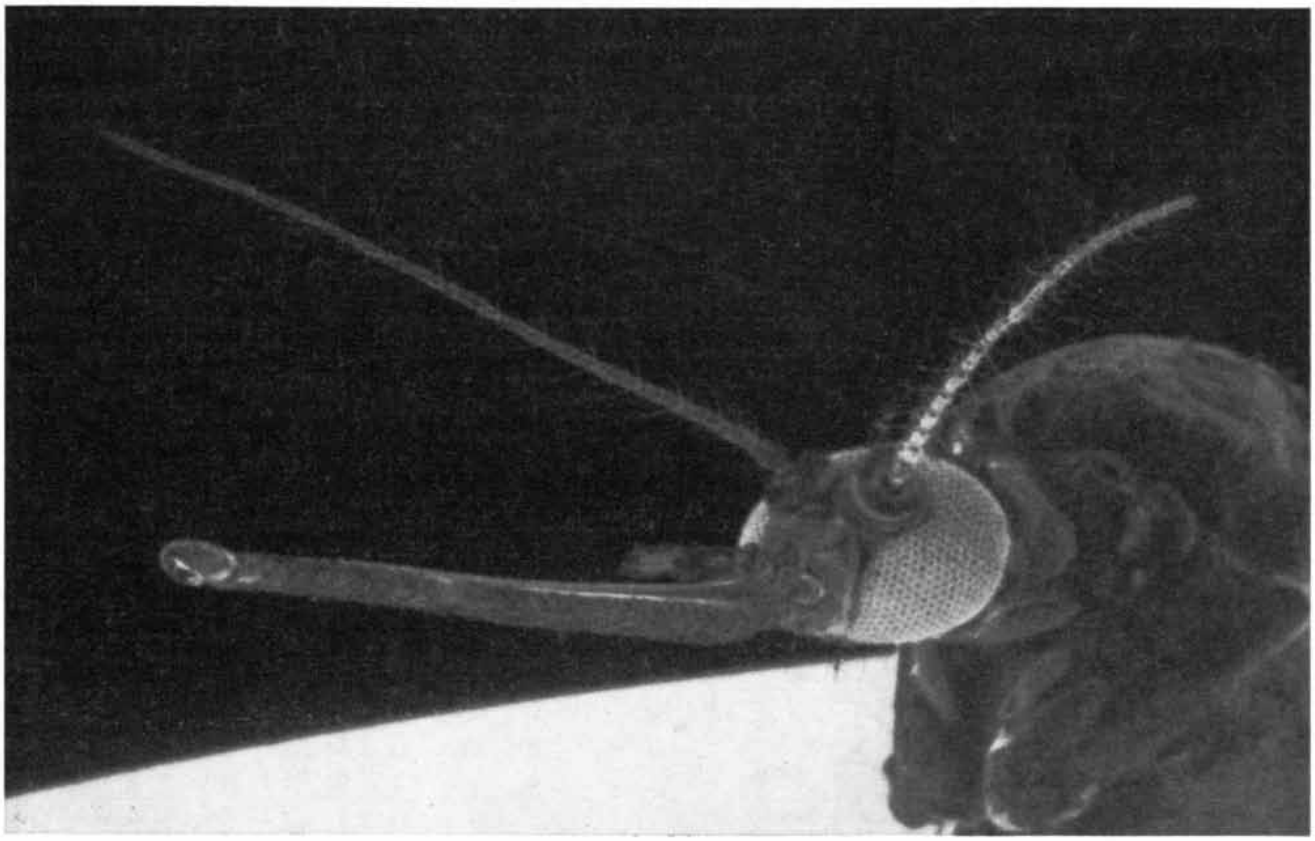
Daykin went on to show that raising the level of carbon dioxide in the mosquitoes' environment shortened the half-life of the resting population to about five minutes but did not change the essential randomness of the flights. This suggests that carbon dioxide increases the number of random discharges reaching the flight command center in the mosquito's brain. In engineering terms these random discharges constitute "noise" in the system, and the effect of the carbon dioxide is to raise the noise level. If we choose to call the process excitation, we must not read too much into the word.

Although raising the concentration of carbon dioxide initially increased the number of spontaneous flights in the resting mosquitoes, when the raised level of carbon dioxide was maintained, the number of spontaneous flights returned to the original frequency. Excitation and flight were restored, however, by a further rise in the concentration of carbon dioxide.

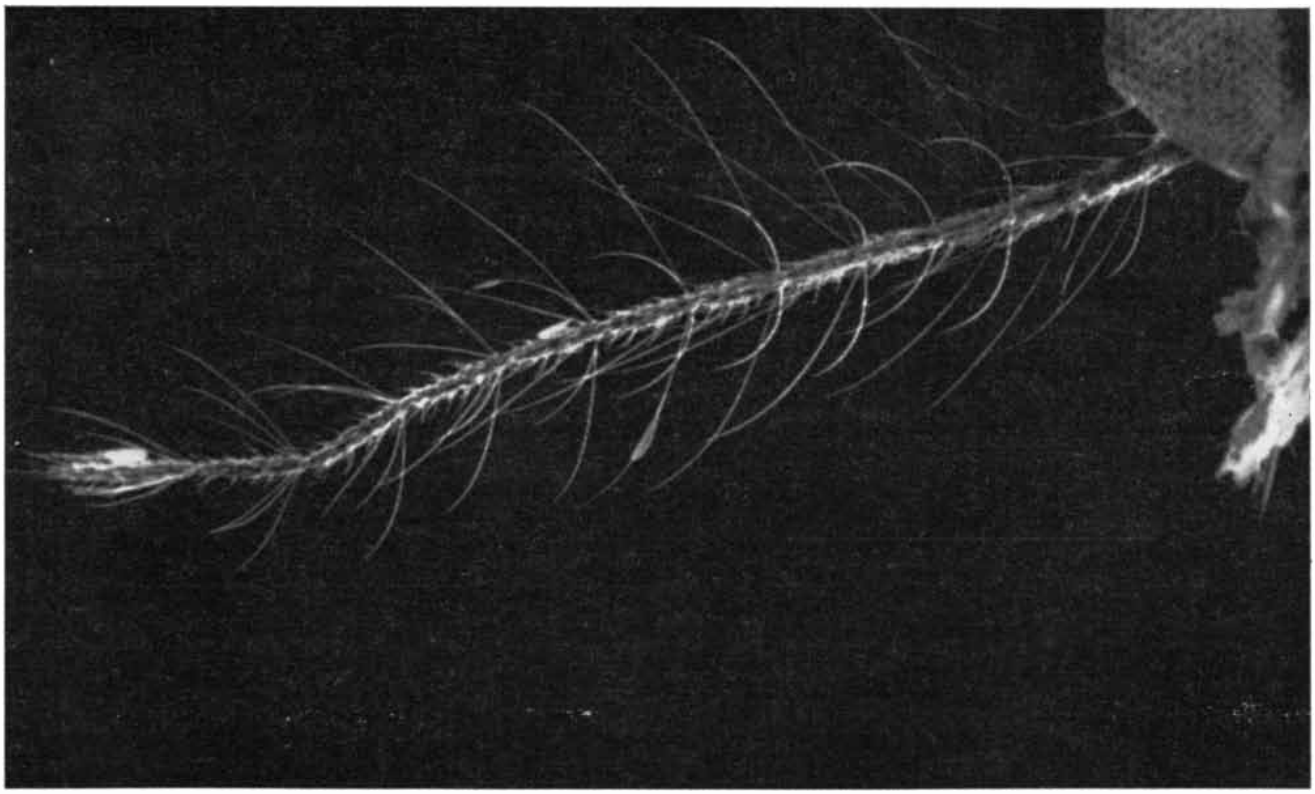
In another series of experiments Daykin and I found that exposure to small concentrations of various common mosquito repellents had a similar effect: excitation followed by adaptation. We then ran tests to find out if adaptation to a repellent would also adapt mosquitoes to carbon dioxide. It turned out that it did.

A practical result followed from these investigations. In an enclosed area mosquitoes can be prevented from responding to the carbon dioxide emitted by warm-blooded animals simply by maintaining a low level of such commonly used and safe repellents as dimethyl phthalate or diethyl toluamide. An area treatment with these chemicals effectively suppresses the activity of mosquitoes. (Mosquitoes seldom begin a search in response to purely visual or mechanical stimuli.)

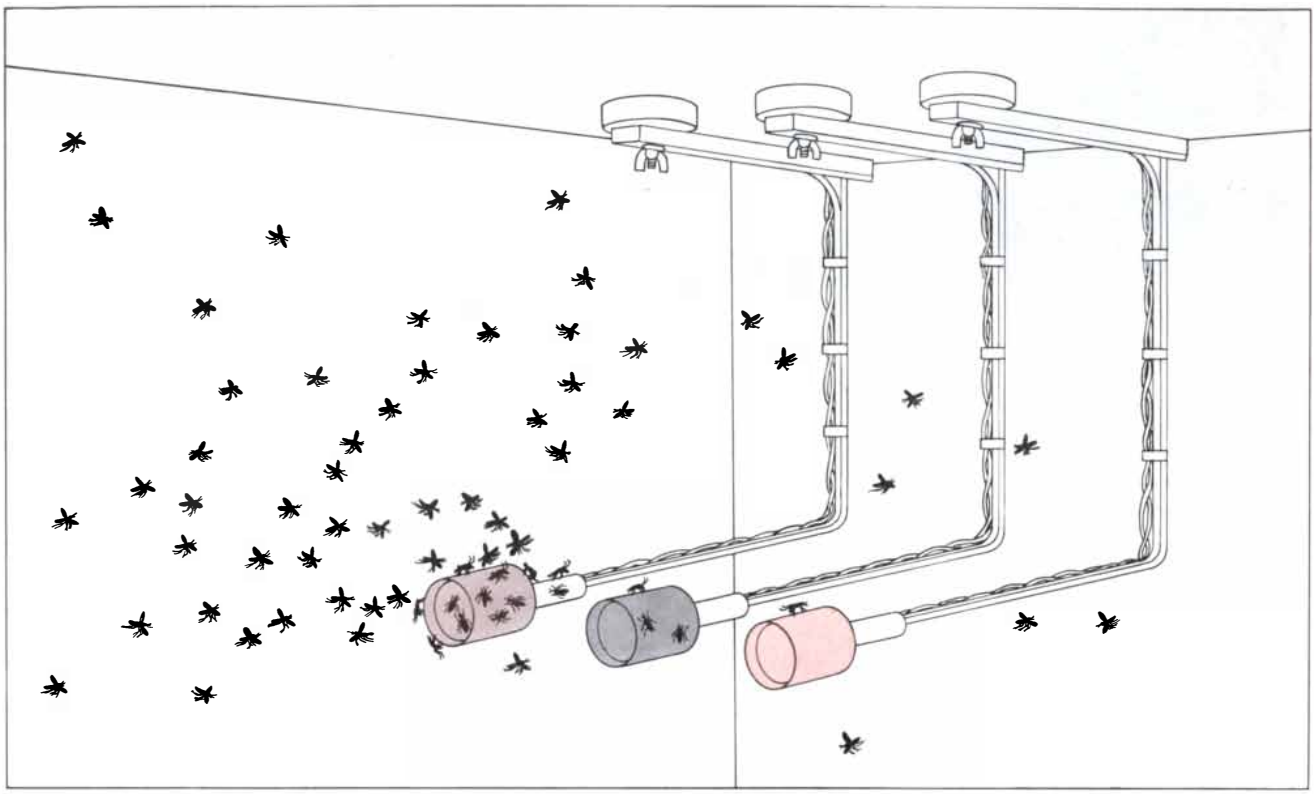
After a mosquito has been stimulated to fly, it must still discriminate between its host and other nearby objects. There have been many suggestions that a special skin odor guides the mosquito, but no such odors have yet been unequivocally identified. Some people are more attractive than others to mosquitoes, but an optimum combination of warmth and humidity is more attractive than the most attractive arm or hand. We were convinced of this by the following experiment. We set up three small cylinders side by side in a wind tunnel. One cylinder was warm, one was wet and one was both warm and wet. After carbon dioxide was added to the air entering the wind tunnel to alert the mosquitoes, we counted the number of mosquitoes that landed on each cylinder. In



PROBOSCIS, HEAD AND ANTENNAE of a female yellow-fever mosquito are enlarged 40 diameters in this scanning electron micrograph. Hairs on the antennae are the sites of sensory receptors that enable mosquitoes to find a warm-blooded host. The body was somewhat deformed in preparing the specimen. The micrograph was made by Edward E. Davis of the Stanford Research Institute.



ANTENNA of a female yellow-fever mosquito is enlarged some 100 diameters in this scanning electron micrograph made by Jack Colvard Jones of the University of Maryland. The longest hairs on the antenna are the *B* receptors, which are believed to detect sound. Various kinds of chemoreceptors are found on the shorter sensory hairs. A portion of the mosquito's compound eye is at top right.

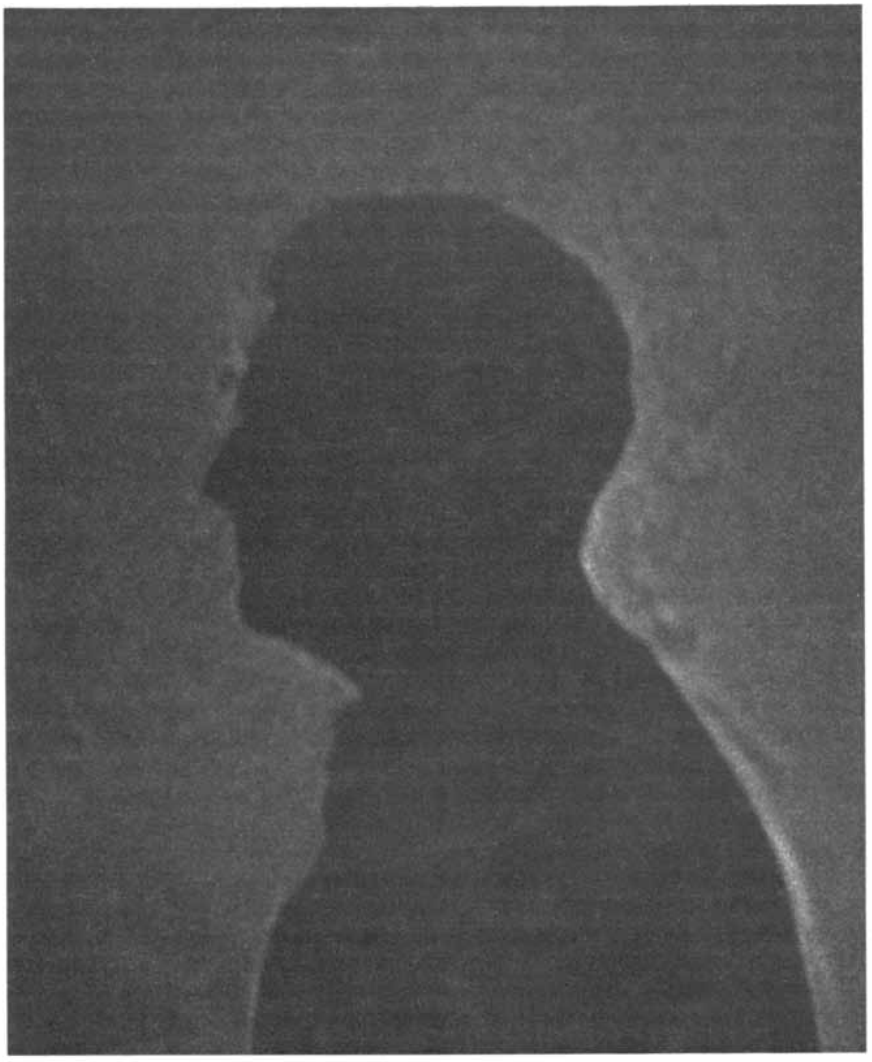


THREE SMALL TARGETS were set in a wind tunnel containing a large number of yellow-fever mosquitoes. One cylinder was warm (*color*), one was wet (*gray*) and one was both warm and wet (*color and gray*). The direction of airflow is from right to left. After the mosquitoes had been alerted by briefly adding carbon dioxide to the airstream, a count was made of the number of mosquitoes that had landed on each cylinder. A large majority of the mosquitoes were attracted to the warm and wet target. This suggests that mosquitoes are able to sense and follow warm and humid convection currents when hunting for warm-blooded hosts on which to feed.

ATTRACTIVENESS TO MOSQUITOES may differ considerably from person to person. Here two subjects both put an arm into the cage at the same time. The arm of one subject attracted nearly 100 mosquitoes; the arm of the other subject attracted fewer than 20 mosquitoes. The arm attracting more mosquitoes had a higher skin temperature and a higher rate of moisture transpiration.

a typical experiment with yellow-fever mosquitoes the target that was warm drew seven mosquitoes, the target that was wet drew 22 and the target that was both warm and wet drew 358. We also demonstrated that the warm and wet target was at least as attractive as a hand, and so there was no need to postulate a special skin odor as an attractant and no need to look for one. There undoubtedly are skin factors that make the mosquito draw blood after it has landed, but our aim was to trace what happens up to the moment of landing and to prevent it.

At this point in our study we met with an instructive example of a misleading experiment. We repeated the three-cylinder experiment with malaria mosquitoes and found that they simply ignored the warm and wet target, but when I put my hand into the tunnel to feel the targets, the hand was instantly attacked. Here was an apparently clear demonstration that malaria mosquitoes responded to some skin factor other than warmth and moisture. Actually it was a matter of target size: my hand was considerably larger than the cylindrical targets. Yellow-fever mosquitoes normally fly in search of food during the day when they can see, whereas malaria mosquitoes fly at night when vision is less useful. The night-flying species therefore need a rather large area of turbulence in the downstream wake in order to work their way to a target. We showed this by setting up a large, streamlined target that was warm and moist. The malaria mosquitoes could not find it until we broke up the streamlines with a baffle. Thus the malaria mosquitoes too are attracted by warmth and moisture.



AURA OF CONVECTION CURRENTS around the human body is made visible by schlieren photography. The currents are created by the warmth of the skin and the transpiration of moisture through it. Such convection currents are readily perceived by mosquitoes.

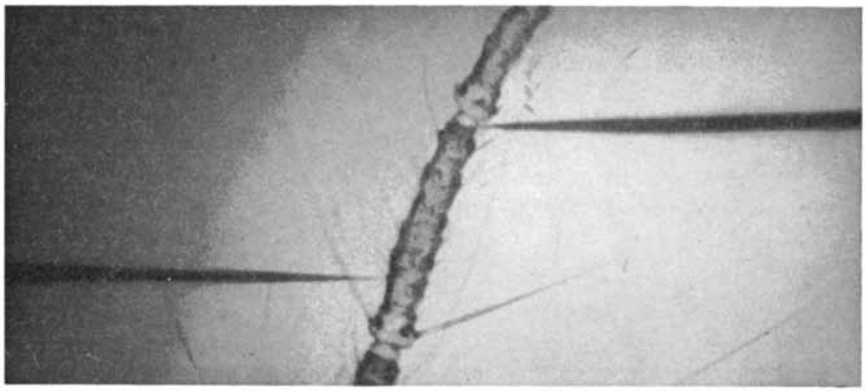
An object that is warmer than its surroundings generates convection currents. Water evaporated from an object also generates convection currents, because moist air is lighter than dry air. An increase in relative humidity from 50 percent to 60 percent has the same effect on the density of air as a rise in temperature of .25 degree Celsius. Although a mosquito may use its eyes to locate an object, vision alone does not tell it whether the object is an inanimate one or a potential host. Convection currents appear to be the clue that enables a mosquito to distinguish between lifeless objects and warm, living animals. The air currents around the two are altogether different.

To find out how a mosquito uses the convection currents we released a stream of warmed and humidified air from a slit into the wind tunnel, together with a trace of nonrepellent smoke to show us where the warm and wet air was going. We then photographed the flight path of a mosquito by a combination of dark-field illumination and time exposures.

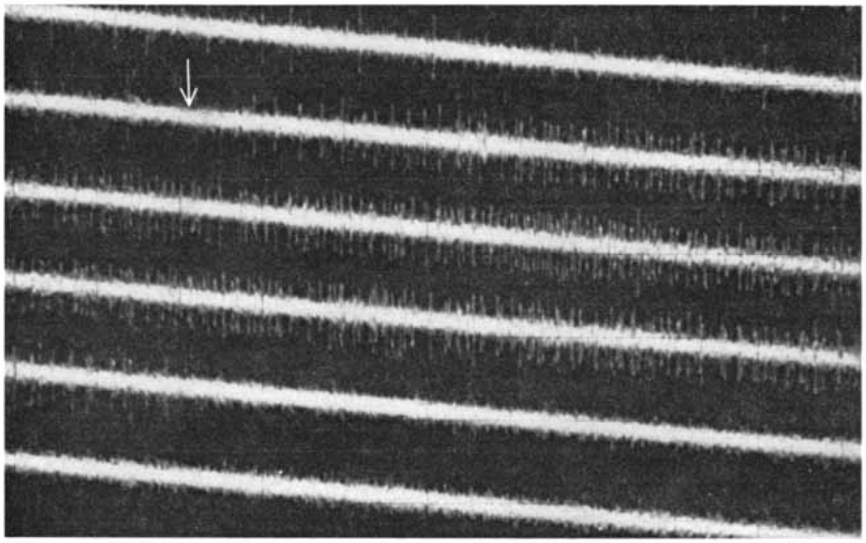
The flight of a mosquito initially is random. When the mosquito encounters a wet and warm convection current, it moves steadily forward. If it passes out of the convection current into cooler and drier air, it turns. The turn is not always in the right direction or sharp enough to bring the mosquito back into the convection current, but it works often enough to enable most mosquitoes to find their prey most of the time.

All of this is a highly condensed account of several years of work. At the end we were able to analyze the "attack program" of the mosquito into a few simple unit responses. A rise in the level of carbon dioxide alerts the mosquito to the approach of a suitable host. The mosquito starts flying, and when it encounters a warm and wet convection current, it keeps on going. When it begins to move out of the convection current, it turns, and the turn usually keeps the mosquito within the current. It follows the current to its source and lands.

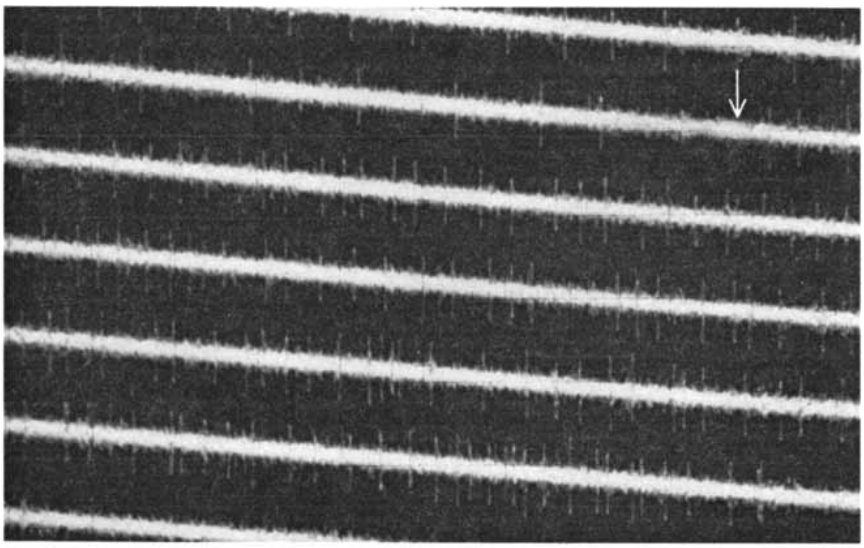
Having clarified how a mosquito behaves when it is searching for a host, we could now look at how a repellent deflects the attack. It became evident to us that a repellent is not an attractant with a minus sign. Attractants can draw insects upwind for a considerable distance, but no known repellent will drive them very far downwind. All a repellent normally does is to turn a mosquito aside just before it lands. Since the mosquito's attack program depends on a certain series of essentially automatic responses occurring in a definite sequence, anything that interrupts that



MICROELECTRODES inserted into the antenna of a female yellow-fever mosquito are shown. The electrode at the left enters the base of a sensillum, or sensory hair, and registers primarily the action potentials that originate in it. The reference electrode is at the right.



DISCHARGE RATE of receptors of a yellow-fever mosquito increases when the relative humidity is increased from 40 percent to 90 percent. Arrow marks point at which humidity was raised. Recording was made by F. E. Kellogg of the British Columbia Research Council.



RESPONSE TO INCREASE IN RELATIVE HUMIDITY from 40 percent to 90 percent is inhibited when the vapor of a repellent, dimethyl phthalate, is simultaneously added. The arrow marks the point at which the humidity was increased and the repellent vapor was added.

sequence has the effect of nullifying the overall success of the program.

Why does the repellent cause a mosquito to turn in its flight? We know that a drop in the humidity and warmth of an airstream will cause the mosquito to turn, and it is reasonable to assume that the impulse to turn comes when there is a sudden drop in the number of signals reaching the mosquito's central nervous system from its sensory receptors for moisture and warmth. To test this assumption my colleague F. E. Kellogg made a series of electrophysiological recordings from individual sensilla, or sensory hairs, on the antenna of the mosquito. He identified the hairs that respond to water vapor, and he showed that there is an increase in their rate of discharge when the relative humidity of the airstream passing over them is raised. When the relative humidity of the air is raised and then suddenly lowered, the discharge rate of the receptors goes up and then drops.

Kellogg tested the effect of raising the relative humidity of the airstream and simultaneously adding the vapor of a repellent. He found that the repellent prevents the moisture sensors from responding normally to the raised humidity. Depending on the relative concentrations of repellent and moisture, the number of nerve impulses does not increase in the usual way in response to a rise in humidity and may even decrease [*see bottom illustration at left*]. A decrease in the discharge rate is what makes the mosquito turn. Normally turning in flight helps the mosquito to find its target, but the turning caused by the repellent sends the mosquito away from the target.

These findings made it clear that a repellent acts in one way on the carbon dioxide sensor and in a different way on the moisture sensor. The carbon dioxide sensor is activated by the repellent, but if exposure to the repellent continues, adaptation occurs and the sensor returns to its normal low rate of discharge. The moisture sensor, on the other hand, seems to be shut off from the outset. That is why mosquitoes have great difficulty in finding a warm and wet object when the entire environment is permeated with a repellent vapor. They fly right through the moist convection currents as though the currents were not there.

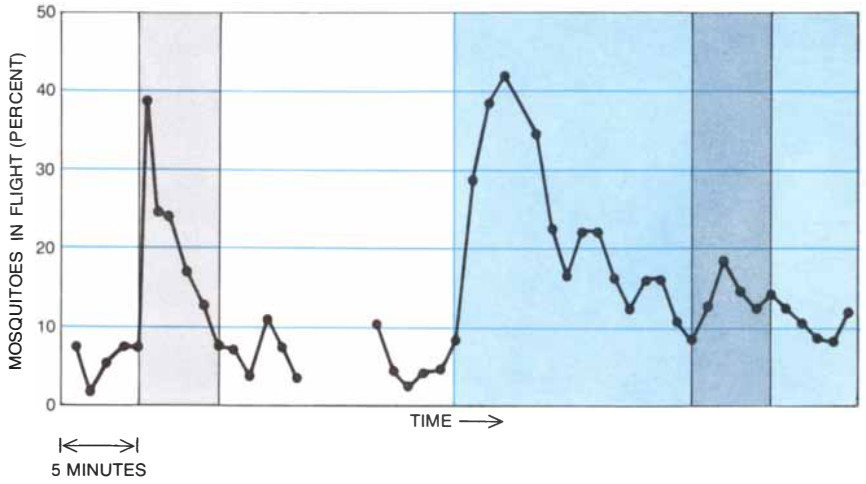
There are other receptors on the antennae of mosquitoes that respond to chemical stimuli, including chemicals that are repellents, such as dimethyl phthalate, and those that are nonrepellent, such as acetone. The fact that a re-

ceptor responds to a certain chemical does not tell us what the biological function of that receptor is. The behavioral response of the mosquito to the chemical must also be studied. We do not know the effect of repellents on the warmth sensors because those sensors have not yet been identified.

More than 25,000 organic compounds have been screened in various Government and military research programs directed toward finding good mosquito repellents. The conventional procedure is to treat the arm of a human volunteer with a standard amount of a chemical and then have him put his arm in a cage containing mosquitoes for three minutes. The exposure is repeated every 30 minutes until the subject is bitten twice within a three-minute interval. About the only good rule that has come out of the arm tests is that an effective repellent for skin application has a molecular weight of between 150 and 250. Apart from that criterion good repellents have been found in almost every class of organic compounds, rendering it difficult to identify the molecular characteristics that make a compound a repellent.

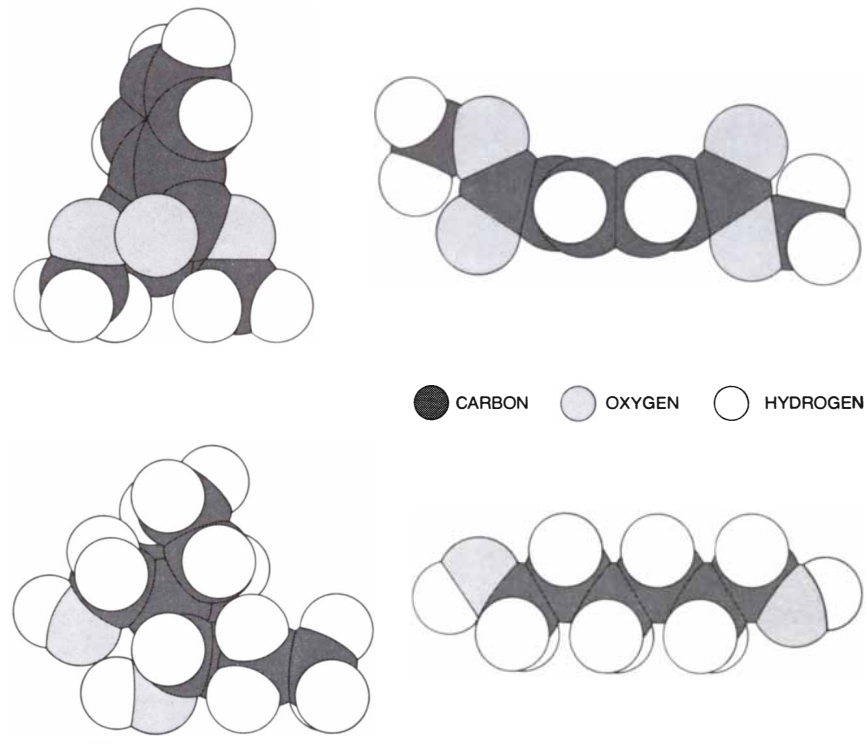
That is because the arm test fails to discriminate between the intrinsic repellency of a compound and its useful service life when it is applied to the skin. In order to measure the intrinsic repellency of various compounds independently of their persistence on the skin, we built a machine that passed two airstreams through a cage of mosquitoes. One stream was warm and humid, and the mosquitoes normally would follow it upwind. They would land on the screen where the warm and wet air was coming in and actively probe it. That part of the screen constituted our artificial target. By means of a machine-driven syringe we could inject a repellent into the airstream at a controlled rate. We found that .01 micromole of diethyl toluamide, a good repellent, per liter of air kept 90 percent of the mosquitoes from landing on the target. To achieve the same effect with citronellol, a less effective repellent, required a concentration about 1,000 times higher [*see upper illustration on next page*].

Micha Bar-Zeev of the Israel Institute for Biological Research compared the protection time of various chemicals with that of diethyl toluamide. The tests were conducted with patches of bare skin on rabbits. In each test one rabbit was treated with diethyl toluamide and a second rabbit with a different compound. The two rabbits were then placed in the same cage and exposed to the same group of mosquitoes. Bar-Zeev demonstrated that several compounds had a substantially longer protection time than diethyl toluamide.



INJECTION OF CARBON DIOXIDE into an airstream passing over a population of yellow-fever mosquitoes induces a sudden rise in the number of mosquitoes in flight. If the increased level of carbon dioxide is maintained for five minutes (*gray area at left*), the number of mosquitoes in flight drops to the original value. Injecting the vapor of a mosquito repellent into the airstream (*colored area*) also causes an initial rise in the number of mosquitoes in flight, followed by a decrease. Continued exposure to the repellent vapor prevents the mosquitoes from responding normally when exposed to carbon dioxide (*gray area at right*).

D. J. Burton and I undertook to compare the intrinsic repellency of some of the compounds tested by Bar-Zeev with that of diethyl toluamide. Our results showed that there is no necessary relation between protection time and intrinsic repellency, although a long-lasting repellent must also have good intrinsic repellency if it is to be effective. The converse does not necessarily hold; for

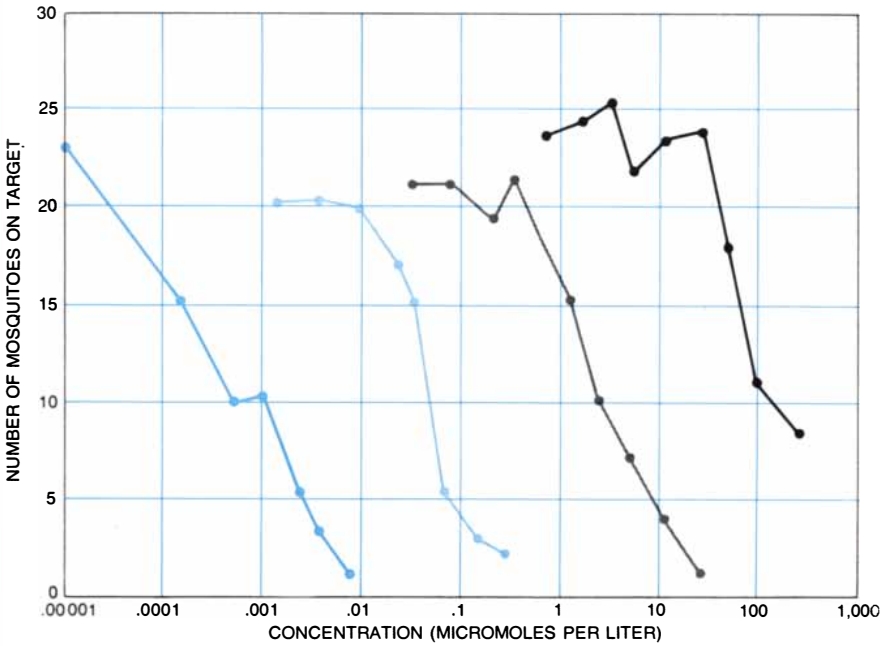


SHAPE OF A MOLECULE of dimethyl phthalate (*upper left*), a good mosquito repellent, is compared with the shape of its isomer, dimethyl terephthalate (*upper right*), a poor repellent. Molecules that are spherical or oval are thought to be more effective blockers of pores in sensory hairs than molecules that are long and flat. Another good repellent, 2-ethyl-1,3-hexane diol (*lower left*), is shown beside its isomer, 1,6-hexane diol, a poor repellent.

example, hexachlorophenol is intrinsically about five times more repellent than diethyl toluamide, but it gives protection for a shorter time [*see lower illustration below*].

From our studies we know that the repellent somehow interferes with the ability of a mosquito to sense water vapor in the air. How does the repellent accomplish this effect? The simplest explanation is that the molecules of the repellent block the pores in the cuticle of the sensory hairs. If that is the case, one would expect the molecules of the repellent to be held in place by the forces of adsorption, and spectroscopic studies have shown that the intermolecular forces giving rise to adsorption tend to be stronger in repellent substances than in nonrepellent ones.

We know the approximate size of the pores and the diameters of the repellent



INTRINSIC REPELLENCY CURVES of four compounds are shown. Diethyl toluamide (*dark colored curve*) requires a concentration of about .01 micromole per liter of air to prevent 90 percent of the mosquitoes in the cage from landing on a standard warm and wet target. The concentration of methyl phenacetate (*light colored curve*) required to achieve the same effect is about 100 times greater. The repellency of citronellol (*gray curve*) is lower still. Heptane (*black curve*) can for all practical purposes be considered a nonrepellent.

COMPOUND	PROTECTION TIME	INTRINSIC REPELLENCY
DIETHYL TOLUAMIDE	1	1
N,N-DIETHYL-2-ETHOXYBENZAMIDE	1.86	.5
N,N-DIPROPYL-2-BENZYLOXYACETAMIDE	1.33	.5
1-BUTYL-4-METHYLCARBOSTYRIL	1.3	2
N,N-DIPROPYL-2-ETHOXYBENZAMIDE	1.26	.3
2-BUTYL-2-ETHYL-1,3-PROPANEDIOL	1.1	1.7
1,3-*BIS*-BUTOXYMETHYL-2-IMIDAZOLIDONE	1.08	.6
N,N-DIETHYL-2-CHLOROBENZAMIDE	.82	1.2
HEXACHLOROPHENOL	.72	.2
1,3-PROPANEDIOL MONOBENZOATE	.53	7.5
DIISOBUTYL MALATE	.39	2.5

PROTECTION TIME AND INTRINSIC REPELLENCY of 10 compounds are compared with those of diethyl toluamide, a commonly used mosquito repellent. Protection time is a measure of the useful service life of a repellent when it is applied to skin. Compounds with a protection time longer than diethyl toluamide's have values greater than 1; compounds with an intrinsic repellency greater than diethyl toluamide's have values of less than 1. In this group of compounds there is no correlation between protection time and intrinsic repellency.

molecules. With these data we can calculate estimates of the blocking and unblocking times. According to these calculations, blocking can occur in a few thousandths of a second even when the concentration of the repellent is only one or two parts per million parts of air. The forces holding the repellent molecule on the pore are relatively weak, and unblocking can take place in less than a second when the mosquito emerges into clean air. These calculated times for blocking and unblocking agree quite well with the observed reaction times of mosquitoes to repellents.

The chemical nature of the cuticle of the sensory hairs is also involved in the adsorption. The cuticle will be somewhat different in different species of mosquitoes, and it is not surprising to find that some chemicals effectively repel one species of mosquito and are ineffective against other species. For example, repellents that incorporate a halogen atom tend to be effective against malaria mosquitoes but not against yellow-fever mosquitoes. Conversely, repellents incorporating an alcohol, an amide or an imide tend to be effective against yellow-fever mosquitoes but not against malaria mosquitoes.

The shape of the molecule of a repellent may also contribute to the molecule's effectiveness in blocking the opening of a pore. Molecules that lie flat on the surface are less likely to be good pore-blockers than molecules that are bulky and stick out from the surface [*see bottom illustration on page 109*].

To sum up, the effect of a repellent on the behavior of mosquitoes depends on whether the repellent is diffused throughout the entire area or is confined to the convection currents arising from a host. Applying a repellent to the skin turns a mosquito away just before it lands. Treating an enclosed area with repellent will suppress the mosquito's arousal response to an increase in carbon dioxide and will also cause the mosquito to fly through a warm and wet convection current without making the usual turning responses. The possibility of preventing mosquito bites by area dispersal of repellent needs more attention. Moreover, if we can gain a better understanding of the physicochemical action of repellents, new and better repellents can be synthesized. Perhaps it will even be possible to develop substances that can be swallowed and subsequently released through the skin, thereby providing long-term protection against the mosquito.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1925: "Professor Shapley, at Harvard, continuing his work on the Magellanic Clouds, has concluded that an irregularly variable star of the ninth magnitude, S Doradus, belongs to the larger cloud and shares its great distance of 110,000 light-years. This makes the star by far the most luminous object yet known. The star is evidently a very hot one, with bright lines in its spectrum, resembling the well-known star P Cygni. Several other stars with spectra of this kind appear in the larger Magellanic Cloud and are exceedingly luminous objects ranging from 15,000 to 60,000 times the sun's brightness. S Doradus excels them all."

"The practical use of the airplane as a means of transportation is making remarkable headway in Europe, whereas in the U.S. but one commercial route has been established. In Europe the airplane is employed mostly for trips to far distant points. The most important line from west to east connects Geneva, Zurich, Munich, Vienna, Bucharest, Constantinople and Ankara. Because of the great distances in Russia the practical use of airplanes has more importance for the Soviet republic than for any other country in Europe. The most important route in Russia goes from Leningrad via Moscow to Baku on the Caspian Sea, with an extended line to Teheran in Persia."

"It was only a little more than 10 years ago that we were given our first information on the success of experiments carried on by Haber under the sponsorship of the Badische Anilin-und-Sodafabrik. It was demonstrated that under certain secret conditions, including the presence of a catalyst, nitrogen and hydrogen could be caused to unite to form ammonia. The original Haber process operated at a pressure of 100 atmospheres and a high temperature; recent work in France indicates that at still greater pressures, up to 1,000 atmospheres, the fixation is more rapid and the yield greater. Although the gain by this method is considerable, it is doubtful whether under present conditions nitrogen from the air can be fixed at a price to compete with Chilean nitrate. We must not overlook, however, the fact that in international relations the food supply is vital. Nitrates are essential to food production."

"A survey of the earth's magnetic field has been completed by the *Carnegie*, a scientific research vessel constructed entirely of wood, brass, copper and other nonmagnetic materials. The timbers in her hull are fastened together with bronze spikes and bolts, the rigging is hemp instead of steel, the cookstoves are built of brass and copper, the anchors are bronze, each weighing 1,900 pounds, and the anchor chains are not chains at all but are manila rope hawsers 11 inches in circumference. All metal fittings, davits and tackle are of bronze or copper. Because of this method of construction no corrections on account of the presence of iron or other magnetic material need be applied to the results obtained with the various magnetic instruments on board. The yacht is charged with the study of the earth's magnetism and with seeking out the cause of variations in the magnetic and electric fields."

SCIENTIFIC AMERICAN

JULY, 1875: "To the followers of Pythagoras the world and its phenomena were all illusion. Centuries later the Egyptian mystic Plotinus taught the same doctrine, that the external world is a mere phantom, and the mystical schools of Christianity took it up in turn. In every age the mystically inclined have delighted in dreaming that everything is a dream, the mere visible reflection of an invisible reality. In truth the delusion lies in the mind of the mystic, not in the things seen. The alleged untrustworthiness of our senses we flatly deny. We frequently misinterpret the messages they bring, it is true, but that is no fault of the senses. The interpretation of sense impressions is something to be learned; we never learn it fully; we are liable to blunder through all our days, but that gives us no right to call our senses liars. It is our judgment, not the sense of sight, that is occasionally deceived. We not only wrong our honest senses but also lose our grip upon this most substantial world when we let mistaken metaphysics persuade us to doubt the testimony they bear."

"The products of the combustion of tobacco are quite numerous and complex. In the smoke the following have been recognized with distinctness: cyanhydric acid; sulphuretted hydrogen; the fatty acids formic, acetic, propionic, butyric and valerianic acid; carbolic acid; creosote; pyridine, picolin, collidin and other similar alkaloids; ammonia; nitrogen; oxygen, and small quantities of marsh gas and carbonic oxide."

"The works of James D. Forbes on glacial phenomena rank among the classics of physical science. He was a precocious Scots youth (whose mother, by the way, came very near to being the wife of Sir Walter Scott). At 17 Forbes was contributor to the *Philosophical Journal;* at 24 he was elected to the chair of natural philosophy at Cambridge and became a member of the Royal societies of Edinburgh and London. This savant, better known as Principal Forbes, made the first accurate determinations concerning the rate of glacial movement. From exhaustive and delicate observations he arrived at the following conclusions: Glacial motion is approximately regular; it is nearly as great during the night as during the day; the center of the glacier moves more rapidly than the sides. In reviewing his results Principal Forbes noted that the movement of a glacier has many resemblances to that of a river. He attributed the motion to the viscous or plastic nature of the ice, in consequence of which it is urged downward by its own weight, like tar or treacle."

"Thirty-nine tons of meat were condemned as unfit for food in New York City during 1874. Under the somewhat lax system of inspection that prevails here, it is likely that this amount was small compared with that which found its way from the hands of the butchers, principally to the poorer classes."

"There is little doubt that the gas escaping constantly from oil wells is of nearly or quite as much value as the oil itself. It is a matter of wonder that means have not long since been adopted to utilize this immense product of the earth. For years the gas has been allowed to pass away into the air uselessly. One well in the Pennsylvania oil region flows with a pressure of 300 pounds to the square inch and is estimated to yield a million cubic feet of gas every 24 hours."

Pioneer's new High Polymer Molecular transducer technology will alter the course of high fidelity.

There's a significant new development in high fidelity that is destined to play a vital role in sound reproduction. It is intimately tied in with the piezoelectric principle.

The piezoelectric effect deals with certain crystal devices that flex when voltages are applied to them. Now, Pioneer has discovered a totally new application of the piezoelectric effect by applying the principle to ultra-thin aluminum coated high polymer film.

By employing this film as a low-mass diaphragm and applying audio signal voltages, the material expands and contracts uniformly generating acoustic energy. For the first time it becomes possible to transform electrical energy to an accurate acoustical equivalent. Such thin-film diaphragms properly mounted are capable of reproducing all music frequencies by means of an incredible "breathing" effect. The ramifications of this unique refinement of the piezoelectric principle are far reaching. Consider such immediate applications as microphones, cartridges, speaker systems and headphones — in fact, any type of electromechanical transducer requiring resonance-free performance.

There have been many attempts to create sound using diaphragm motion. For example, electrostatic speakers and headphones. But in contrast to the electrostatic principle, the new application of the High Polymer Molecular principle as discovered and perfected by Pioneer, requires no dangerous, high polarizing voltages.

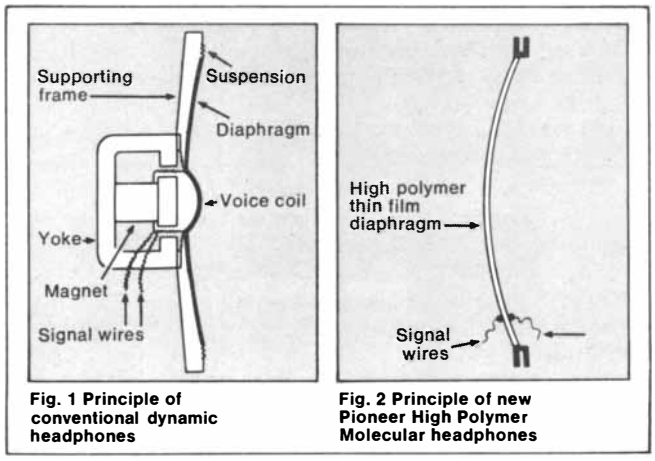


Fig. 1 Principle of conventional dynamic headphones

Fig. 2 Principle of new Pioneer High Polymer Molecular headphones

The first totally new concept in headphones in over a decade.

Pioneer has successfully incorporated the High Polymer Molecular transducer principle in two new headphones that are unlike any others. Conventionally designed headphones use moving coils, miniature loudspeaker elements and other mechanical parts — as shown in Figure 1 — all of which come between you and your music. Pioneer's new SE-700 and SE-500 headphones don't. They employ a single thin-film high polymer piezoelectric diaphragm that reproduces sound directly, as shown in Figure 2. Only the diaphragm moves air — and moves it accurately, in exact conformance with the electrical signal applied directly to it. The accurate, low-distortion signals available from any standard headphone jack on your receiver or amplifier are directly translated to equally precise, low-distortion sound by the action of the high polymer film diaphragm. Nothing, absolutely nothing comes between you and the original sound.

Even though you may now own a pair of headphones, you owe it to yourself to hear these new piezoelectric high polymer transducer headphones. In fact, compare them with other types. You'll find a lower level of distortion-free sound than has ever been achieved — even at unprecedented volume levels. The experience of listening with these new Pioneer headphones is a revelation. In addition, the open-back design, light weight and soft, snug fitting earpieces permit hours of comfortable, private listening. You'll come away from your Pioneer dealer thoroughly convinced that Pioneer has altered the course of high fidelity.

SE-700—$79.95; SE-500—$49.95. Both are supplied with a 9¾ foot connecting cable, standard phone plug and storage case.

U.S. Pioneer Electronics Corp., 75 Oxford Drive, Moonachie, New Jersey 07074. West: 13300 S. Estrella, Los Angeles 90248 / Midwest: 1500 Greenleaf, Elk Grove Village, Ill. 60007 / Canada: S.H. Parker Co.

PIONEER®
when you want something better

MATHEMATICAL GAMES

On tessellating the plane with convex polygon tiles

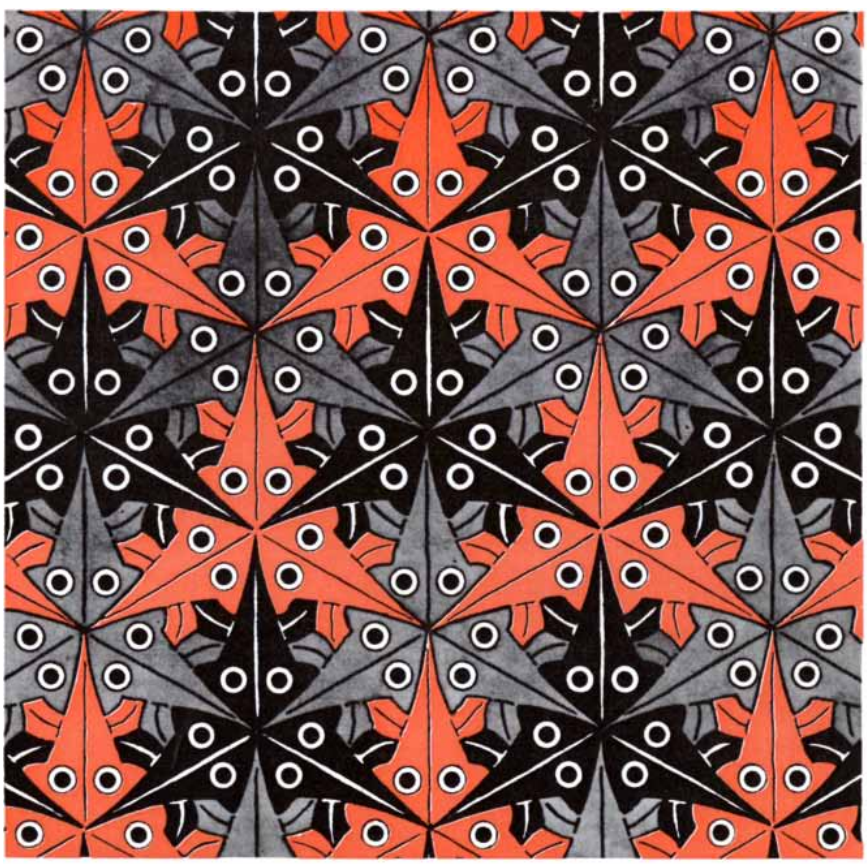
by Martin Gardner

"Many of the brightly coloured, tile-covered walls and floors of the Alhambra in Spain show us that the Moors were masters in the art of filling a plane with similar interlocking figures, bordering each other without gaps. What a pity that their religion forbade them to make images!"

—M. C. Escher

Imagine that you have an infinite supply of jigsaw puzzle pieces, all identical. If it is possible to fit them together without gaps or overlaps to cover the entire plane, the piece is said to tile the plane, and the resulting pattern is called a tessellation. From the most ancient times such tessellations have been used throughout the world for floor and wall coverings and as patterns for furniture, rugs, tapestries, quilts, clothing and other objects. The late Dutch artist M. C. Escher amused himself by tessellating the plane with intricate shapes that resemble birds, fish, animals and other living creatures [*see illustration below*].

A tile that tessellates obviously can have an infinite variety of shapes, but by imposing severe restrictions on the shape the task of classifying and enumerating tessellations is reduced to something manageable. Geometers have been particularly interested in polygonal tiles, of which even the simplest present formidable problems. This month we are concerned only with the task of finding all convex polygons that tile the plane. It is a task that was not completed until 1967, when Richard Brandon Kershner, now assistant director of the Applied Physics Laboratory of Johns Hopkins University, found three pentagonal tilers that had been missed by all predecessors who had worked on the problem.



Tessellation by M. C. Escher

Let us begin by asking how many of the regular polygons tile the plane. As the ancient Greeks knew and proved, there are just three: the equilateral triangle, the square and the regular hexagon. The hexagonal tiling, so familiar to bees and users of bathrooms, is a fixed pattern [*see top illustration on opposite page*]. The patterns formed by equilateral triangles or by squares can be infinitely varied by sliding rows of triangles or squares along lattice lines.

If we remove the restriction that a convex polygon must be regular, the tiling problem grows in interest. It is not hard to show, using Euler's famous formula $v - e + f = 1$ (where letters stand for the vertexes, edges and faces of a polygonal network) and some elementary Diophantine analysis, that no convex polygon of more than six sides can tile the plane. Thus we need to investigate only polygons of three, four, five and six sides.

The triangle is easy. Any triangle tiles the plane. Simply fit two identical triangles together, with the corresponding edges coinciding as shown at the left in the middle illustration on the opposite page, and you create a parallelogram. Replicas of any parallelogram obviously will go side by side to make an endless strip with parallel sides, and the strips in turn go side by side to fill the plane.

The quadrilateral is almost as easy, although much more surprising. Any quadrilateral tiles the plane! As before, take a pair of identical quadrilaterals, one inverted with respect to the other, join the corresponding edges and you create a hexagon [*see bottom illustration on opposite page*]. Each edge of the hexagon is necessarily equal to and parallel to its opposite edge. Such a hexagon, by a simple translation operation (altering its position on the plane without changing its orientation), will form a tiling pattern. The quadrilateral need not be convex. Exactly the same technique creates a tiling pattern for any nonconvex quadrilateral.

The case of the hexagon was settled in 1918 by K. Reinhardt in his doctoral thesis at the University of Frankfurt. He showed that any tessellating convex hexagon belongs to one of three classes.

Kershner, in an article "On Paving the Plane" (*APL Technical Digest*, July, 1969), explains the three types as follows.

Label the sides and angles of a hexagon as shown in the illustration on the next page. A convex hexagon will tile the plane if and only if it belongs to one or more of the following classes:

1. $A + B + C = 360°$, and $a = d$.
2. $A + B + D = 360°$, and $a = d$, $c = e$.
3. $A = C = E = 120°$, and $a = b$, $c = d$, $e = f$.

The illustration gives an example of each type of convex hexagon tiler and a portion of its tiling patterns. The colored lines outline a "fundamental region" that tiles the plane by translation. Note that Type 2 requires reflection if the hexagon is asymmetric.

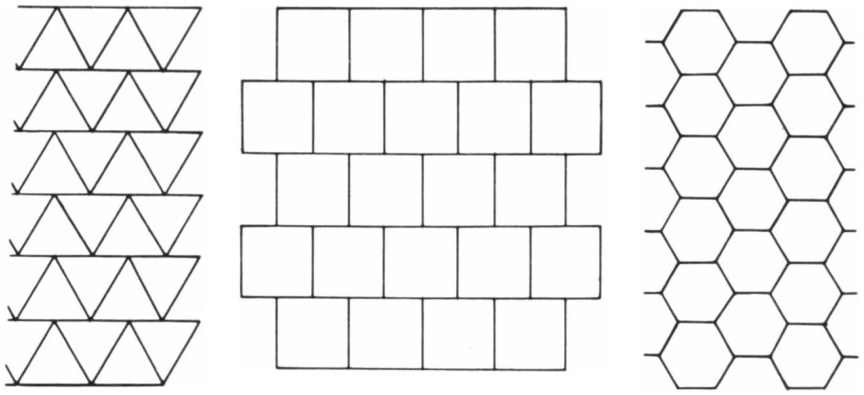
In similar fashion the tessellating convex pentagons can be classified in eight ways. Five were found by Reinhardt. Kershner describes them by labeling the pentagon as shown in the illustration on page 115. A convex pentagon paves the plane if it belongs to one or more of the following classes:

1. $A + B + C = 360°$.
2. $A + B + D = 360°$, and $a = d$.
3. $A = C = D = 120°$, and $a = b$, $d = c + e$.
4. $A = C = 90°$, and $a = b$, $c = d$.
5. $A = 60°$, $C = 120°$, and $a = b$, $c = d$.

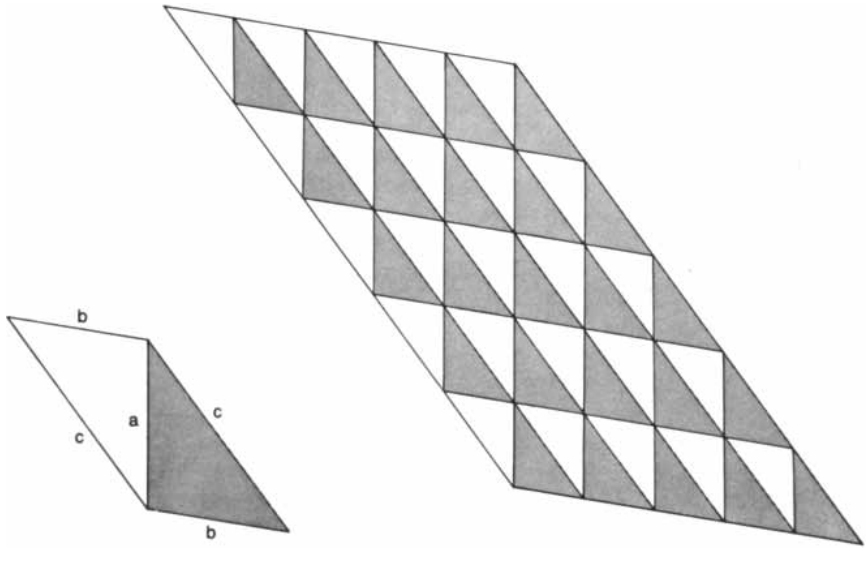
Examples of each type and its tiling pattern are reproduced in the illustration with colored lines outlining fundamental regions. Only Type 2 requires reflection.

"At this point," writes Kershner, "either [Reinhardt's] technique or his fortitude failed him, and he closed the thesis with the statement that in principle it ought to be possible to complete the consideration of pentagons along the lines of his considerations up to that point, but it would be very tedious and there was always the possibility that no further types would emerge. Indeed, it is quite clear that Reinhardt and everyone else in the field thought that the Reinhardt pentagon list was probably complete.

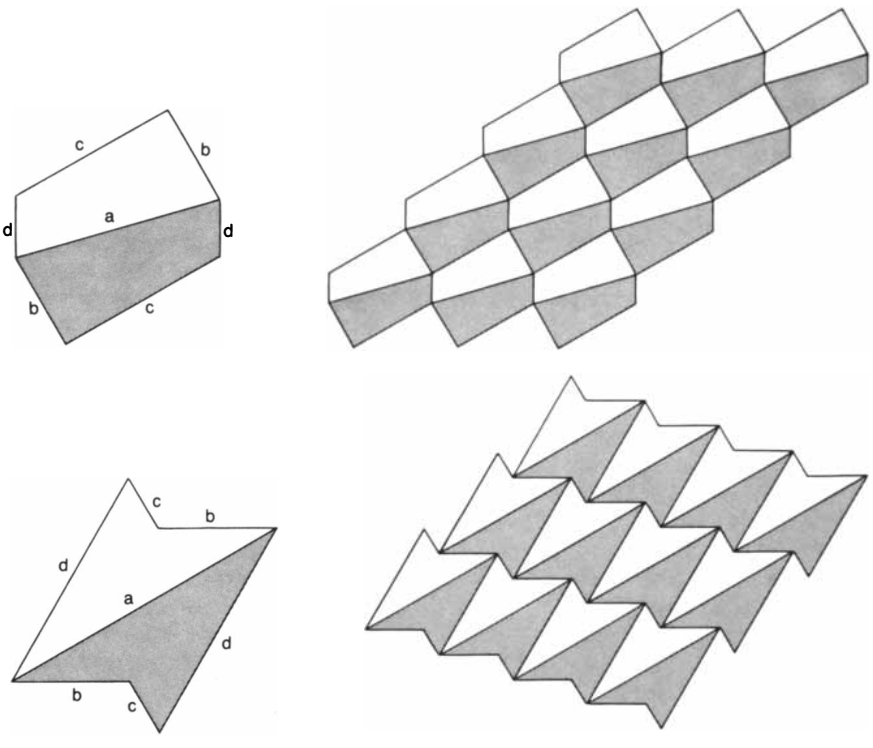
"For reasons that I would have difficulty explaining I have been intrigued by this problem for some 35 years. Every five or 10 years I have made some kind of attempt to solve the problem. Some



The three regular polygons that tile the plane



Any triangle tiles the plane



Any quadrilateral tiles the plane

two years ago I finally discovered a method of classifying the possibilities for pentagons in a more convenient way than Reinhardt's to yield an approach that was humanly possible to carry to completion (though just barely). The result of this investigation was the discovery that there were just three additional types of pentagon... that can pave the plane. These pavings are totally surprising. The discovery of their existence is a source of considerable gratification."

The three additional types [*see top illustration on page 116*] are described by Kershner as follows (Type 7 and Type 8 require reflection):

6. $A + B + D = 360°$, $A = 2C$, and $a = b = e$, $c = d$.
7. $2B + C = 2D + A = 360°$, and $a = b = c = d$.
8. $2A + B = 2D + C = 360°$, and $a = b = c = d$.

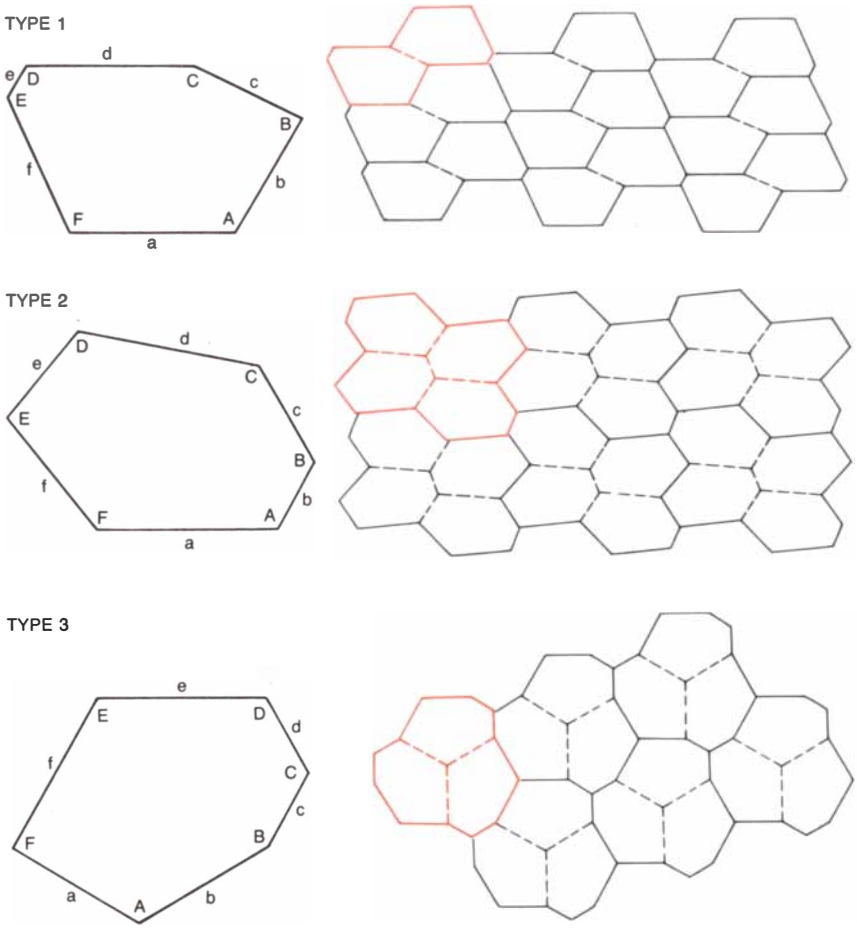
Kershner's paper does not include a proof that there are no other convex pentagons that tile the plane, "for the excellent reason," reads the editor's introductory note, "that a complete proof would require a rather large book." Note that Kershner has deliberately drawn his tessellations with polygons that are as irregular as possible, within the limits of their type, in order to bring out the nature of the tessellation. The most regular hexagonal tessellation is, of course, the familiar beehive mosaic. One can readily see that it belongs to all three hexagon types.

If the beehive hexagons are bisected, the result is a pentagonal pattern that belongs to Type 1 [*see "a" in bottom illustration on page 116*]. The pattern formed by six pentagons in a flowerlike arrangement ["*b*"] belongs to Type 1, Type 5 and Type 6. The most remarkable of all the pentagonal patterns is a tessellation of *equilateral* pentagons ["*c*"]. It belongs only to Type 1. Observe how quadruplets of these pentagons can be grouped into oblong hexagons in two different ways, each set tessellating the plane at right angles to the other. This beautiful tessellation is frequently seen as a street tiling in Cairo, and occasionally in the mosaics of Moorish buildings.

The equilateral pentagon is readily constructed with a compass and a straightedge [*see illustration on page 117*]. First draw a side of the pentagon, *AB*. Construct its perpendicular bisector, *CD*, then draw lines *CE* and *CF* at 45 degrees to *AB*. With center at *A* and radius *AB*, draw a circular arc cutting *CE* at *P*. The same construction is repeated on the other side, with center at *B* and the arc cutting *CF* at *R*. Keeping the compass with radius *AB*, let *R* be the center and strike an arc that cuts the perpendicular bisector, *CD*, at *Q*.

The pentagon's corner angles at *P* and *R* are right angles. The corner at *Q* is a little more than 131 degrees, and corners *A* and *B* are a trifle more than 114 degrees. The length from *Q* to *B* is the product of a side of the pentagon and the square root of 2. The pentagon's area (it is easy to prove) is precisely the square of line segment *CR*.

Among the infinite tessellations of the plane that can be made with congruent nonconvex polygons, combinatorial geometers have given special attention in recent years to tiling with polyominoes and their cousins the polyiamonds and polyhexes. (Polyominoes are formed by joining unit squares, polyiamonds by joining equilateral triangles, and polyhexes by joining regular hexagons.) Many fascinating problems have been formulated, some solved and some not. That will be the topic next month.



The three types of convex hexagon tiler

The answers to last month's problems are that the second player has the win on 1-by-*n* tromino cram, and White wins the 4-by-4 rex (reverse hex) game by taking the cell at the intersection of row *C* and diagonal 1. Although no winning strategy for the first player of order-4 rex is known, a winning second-player pairing strategy for the order-5 game has been found by David L. Silverman. All higher orders remain unsolved.

G. W. Lewthwaite's game generalizes in an obvious way to rectangular boards of any size and shape. If the rectangle has an odd number of cells, the second player wins; if it has an even number of cells, the first player wins. (In the latter case the domino covering strategy, explained last month, includes the initially vacant cell.) Karl Fulves has suggested that instead of visualizing a domino pattern, you play the game with counters secretly marked so that you can place them on the board in orientations that group them in pairs. For example, a small pinhole on the rim of a pawn or a checker would enable you to orient the pieces so that the pinholes of each pair are adjacent. You could then play the domino strategy without having to re-

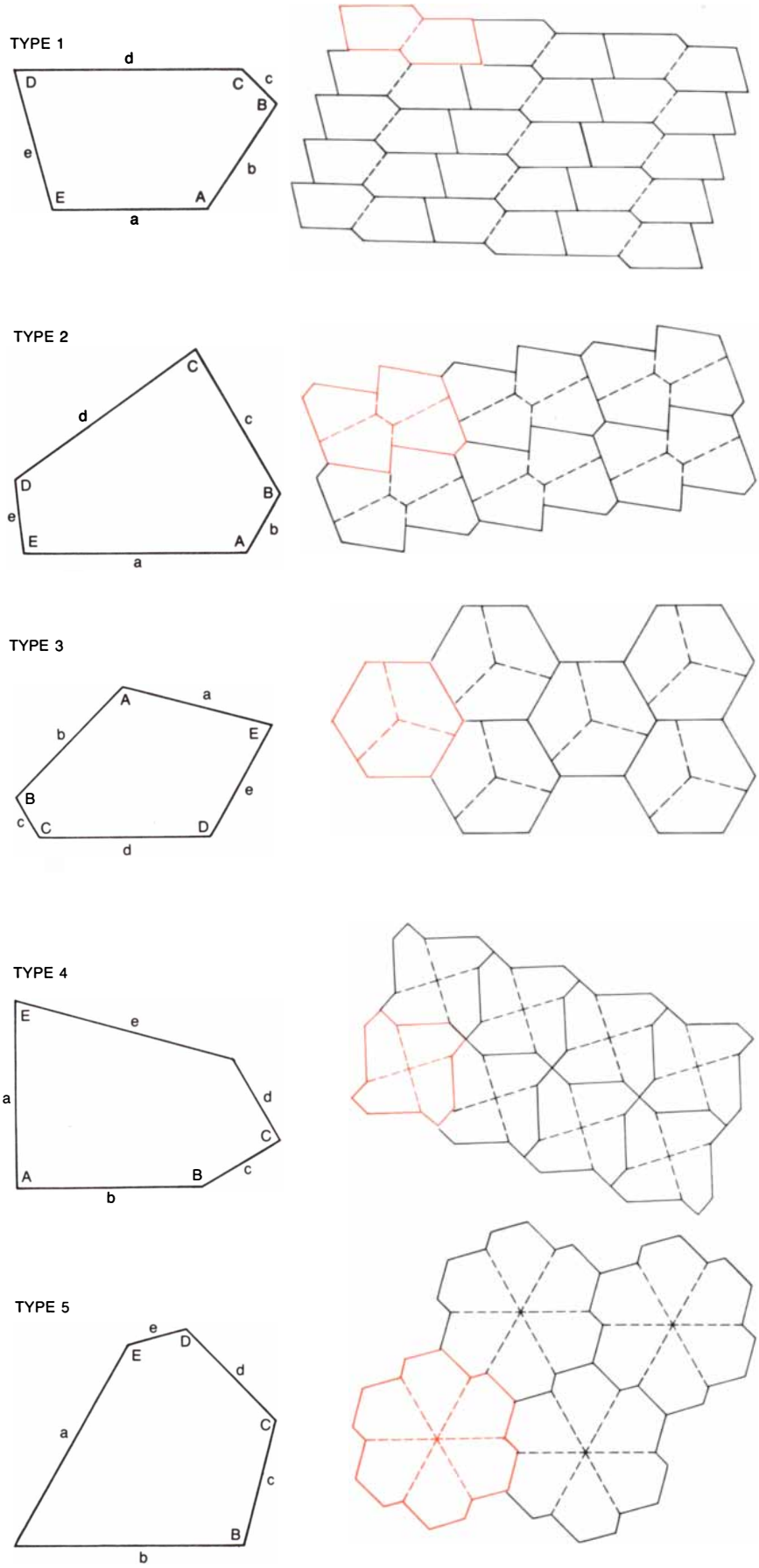
member a domino pattern. If coins are used, designate a spot on the rim of each side of each coin, say the N in the ONE on a penny, as your mark and orient the coins accordingly.

The April Fool hoax column produced more than 1,000 letters from readers who failed to see the joke in spite of the outlandish names and preposterous ideas. Hundreds of readers sent colorings of the four-color-theorem map. William McGregor, who designed the map as a puzzle, has given me permission to use his name.

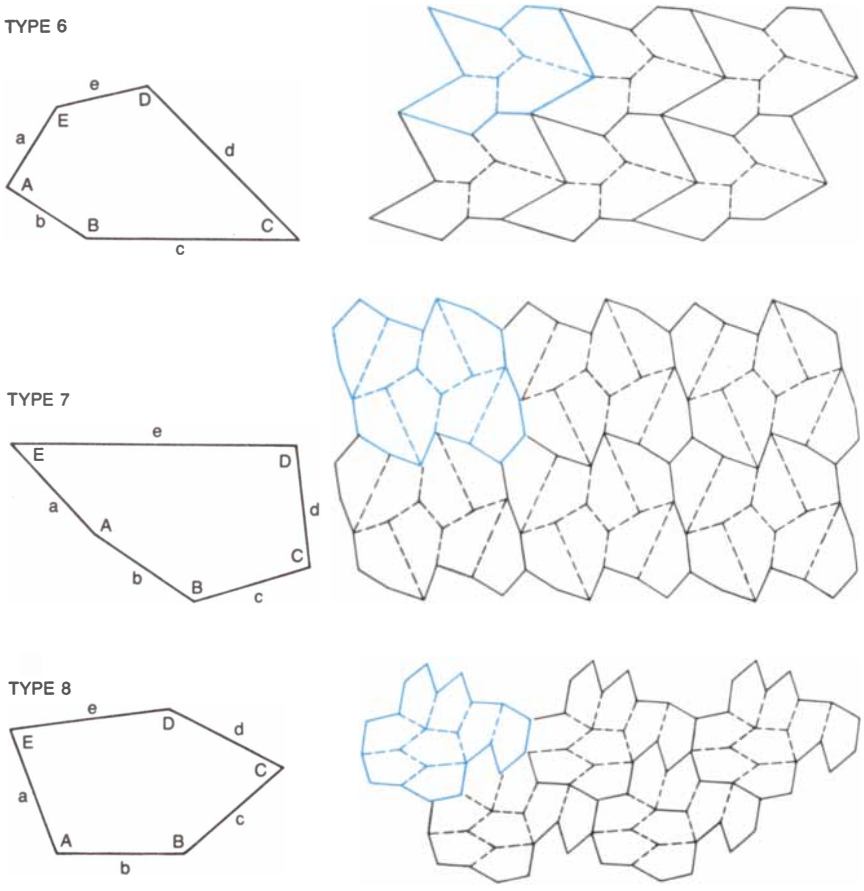
When e is raised to the power of the product of π and the square root of 163, the result is the 18-digit number I gave, minus .000 000 000 000 75.... John Brillo, to whom I attributed this hoax, is a play on the name of the distinguished number theorist John Brillhart. The reference to Ramanujan's paper is legitimate. In it the Indian mathematician discusses a family of remarkable near-integer numbers to which this one belongs, but of course he knew that none were integral. Indeed, as many readers pointed out, it is not hard to prove they are transcendental.

The description of Richard Pinkleaf's chess-playing program, MacHic, is a play on the chess program MacHack, written by Richard Greenblatt of the Massachusetts Institute of Technology. The relativity paradox that I hung on Humbert Pringle (a play on the name of Herbert Dingle, the British physicist who maintains that relativity theory is disproved by the famous twin paradox) is well known. It appears as a problem on page 99 of the paperback edition of *Spacetime Physics*, by Edwin F. Taylor and John A. Wheeler (W. H. Freeman and Company, 1966), and the solution is given on page 25 of the answer section. The paradox is discussed at greater length by George Gamow in *Mr. Tompkins in Wonderland* (Macmillan, 1947), W. Rindler in *American Journal of Physics*, Vol. 29, 1961, pages 365 ff., R. Shaw, *ibid.*, Vol. 30, 1962, page 72 ff., and P. T. Landsberg in *The Mathematical Gazette*, Vol. 47, 1964, page 197 ff.

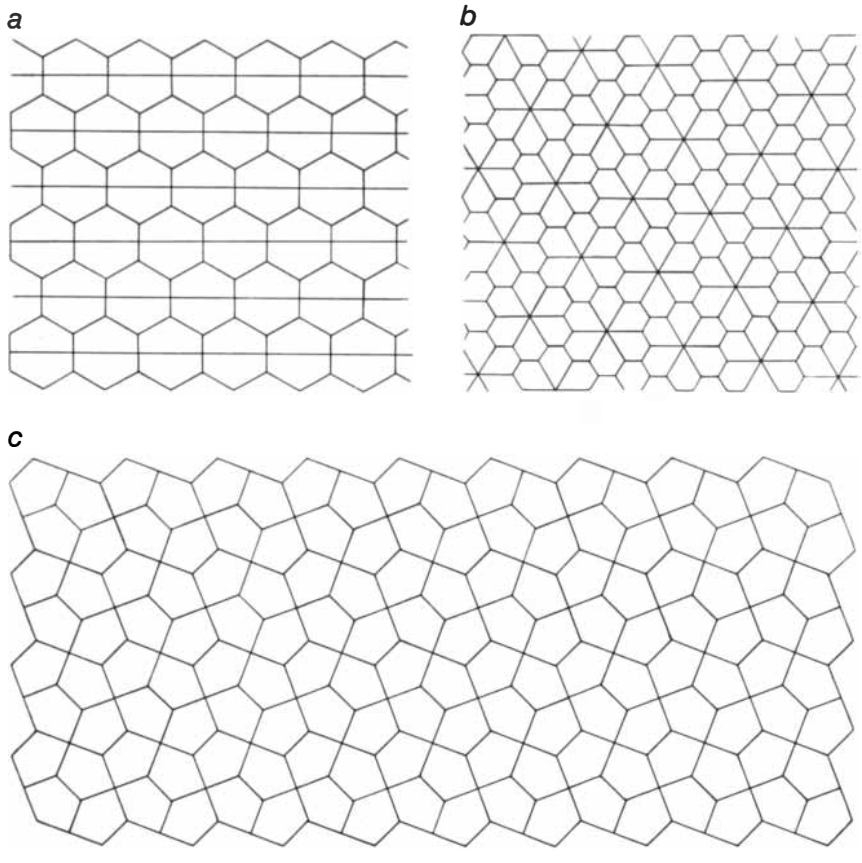
A stationary outside observer will see the meterstick just make it through the hole. If the plate and the stick have thickness, the stick must of course be a trifle shorter than the hole to prevent an end from catching. To an observer on the plate the stick will appear Lorentz-contracted, but it will also appear rotated so that it seems to approach the hole on a slant. The stick's *back* end actually seems to go through the hole before its front



The five types of tiling convex pentagon known in 1918



Three new types of tiling convex pentagon discovered in 1967



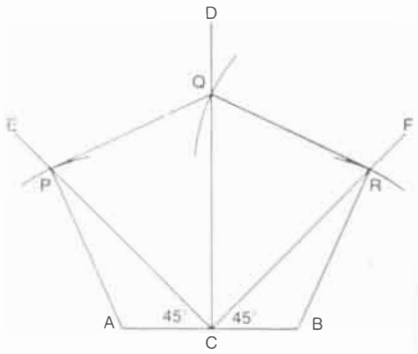
Pentagonal tessellations of unusual symmetry

end, so that it gets through with the same clearance as before. To an observer on the stick the plate will appear Lorentz-contracted, its hole becoming elliptical, but the plate also appears to be rotated. In this case the hole first goes over the *front* end of the slanted stick, again with the same close fit. "Contractions" and "rotations" are ways of speaking in a Euclidean language. In a four-dimensional, non-Euclidean language of space-time the objects retain their shapes and orientations. Having at one time written a book on relativity, I was abashed to receive more than 100 letters from physicists pointing out the stupid blunder I had made.

Those who enjoyed explaining the paradox may wish to consider how to escape from the following variant. Assume that the meterstick is sliding at high speed along the surface of an enormous flat plate of metal toward a hole slightly larger than the stick. We idealize the thought experiment by assuming that there is no friction and that the stick and the plate are extremely thin. When the stick is over the hole, gravity (or some other force) pulls it down and through. For an observer on the stick the sheet slides under it, and the hole is Lorentz-contracted enough to prevent the stick from dropping through. In this case the stick and the plate cannot rotate relative to each other. How does the stick get through? (Please, no letters! I know the answer.)

The Leonardo da Vinci drawing was done by Anthony Ravielli, a graphic artist well known for his superb illustrations in books on sports, science and mathematics. It was an earlier version of the sketch that suggested to me the idea of a hoax column. Many years ago a friend of Ravielli's had jokingly made a bet with a writer that Leonardo had invented the first valve flush toilet. The friend persuaded Ravielli to do a Leonardo drawing in brown ink on faded paper. It was smuggled into the New York Public Library, stamped with a catalogue file number and placed in an official library envelope. Confronted with this evidence, the writer paid off the bet.

Augusto Macaroni is a play on Augusto Marinoni, a da Vinci specialist at the Catholic University of Milan, and Ramón Paz y Bicuspid is a play on Ramón Paz y Remolar, the man who actually found the two missing da Vinci notebooks. My data on the history of the water closet is accurate except for the reference to Thomas Crapper. The book by Wallace Reyburn, *Flushed with Pride: The Story of Thomas Crapper*, does exist, and it is the funniest plumbing hoax since H. L.



How to construct equilateral-pentagon tiler

Mencken wrote his history of the bathtub. Reyburn's book persuaded many readers that the slang words "crap" and "crapper" derive from water closets made in England by one Mr. Crapper, but I shall cite two indications that the book is a joke: (1) the entry on "crap" and "crapping case" in *The Slang Dictionary* (London: Chatto and Windus, revised, 1873) and (2) Reyburn's latest book, *Bust-up: The Uplifting Tale of Otto Titzling and the Development of the Bra* (Prentice-Hall, 1972).

The Ripoff Rotor is a modification of a psychic motor described in Hugo Gernsback's lurid magazine *Science and Invention,* November, 1923, page 651. Prizes were awarded in March, 1924, to readers who gave the best explanations of why the cylinder turned. The motion can be caused by any of three forces: slight air currents in the room, convection currents produced by heat from the hand and currents from breathing. The three forces combine in unpredictable ways, so that if a person who believes he has psychokinetic powers is willing the motor to turn, it may turn in the direction willed, or it may go the other way.

I had no individual in mind when I mentioned Ms. Henrietta Birdbrain, but there *is* an *East West* newspaper in Boston, and if anyone bothers to check the issue cited, he will find a sober report by Stanley Krippner on a psychic motor that was demonstrated to him in Prague by Robert Pavalita. (On Pavalita, see Chapter 28 in *Psychic Discoveries behind the Iron Curtain,* by Sheila Ostrander and Lynn Schroeder.)

Many readers, in the spirit of the hoax, sent hilarious explanations for the Ripoff Rotor. Unfortunately the best are too long to quote. Mark J. Hagmann found that the rate of the relative rotation of the rotor and the room was a function of the contents of the liquor bottle he used to support the needle. The rotation increased as the level of the liquid went down.

Introduction to Contemporary Psychology

Announcing a New Text

by EDMUND J. FANTINO and GEORGE S. REYNOLDS, University of California, San Diego

An instructor's guide and study guide will be published along with the text. A free test item file will be sent to teachers upon class adoption of the text.

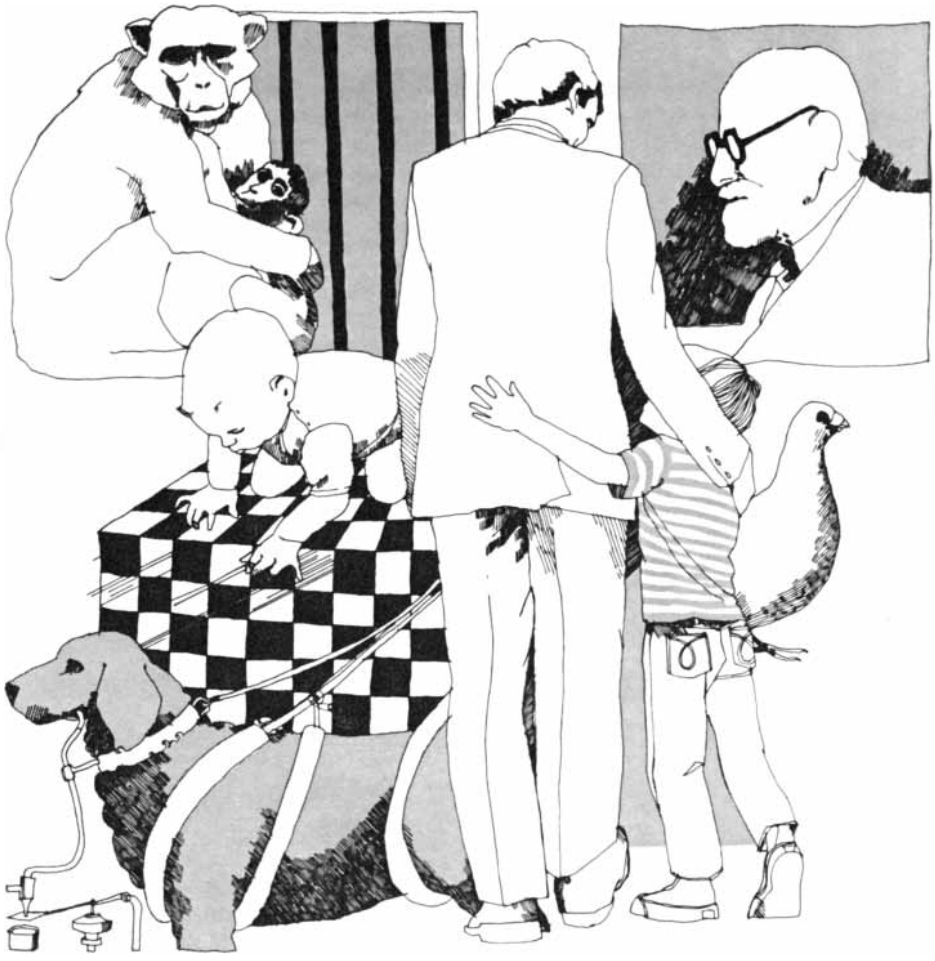
"Although we feel that there is not necessarily any incompatibility among prevailing psychological points of view, we also feel that the introductory student deserves a fair look at, and a sure feel for, all of them."—The Authors

Many students take an introductory psychology course to satisfy an area requirement, others take it as a prerequisite to more advanced courses in psychology, still others take it because they find the subject fascinating. *Introduction to Contemporary Psychology* is a text that will serve all these students well.

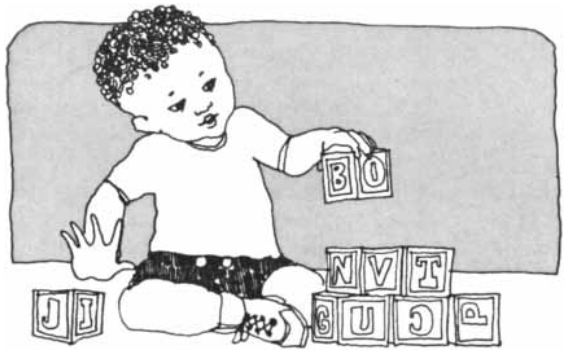
This book takes a new approach to introductory psychology—it is an intentionally eclectic text, treating the subdisciplines and many schools of thought that constitute modern psychology from the point of view of the psychologists actually involved in them. Such an approach gives students an appreciation of the great scope of the science and prepares them for subsequent specialization.

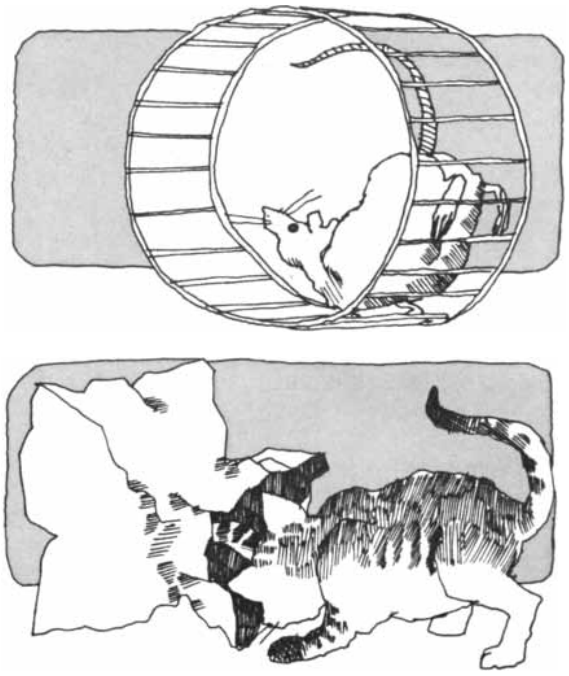
Introduction to Contemporary Psychology is divided into two major parts. The first half of the book deals with such basic psychological processes as learning, remembering, thinking, perception, motivation, instinct, and heredity. Whenever appropriate, the authors have included a discussion of the physiological correlates of behavior. The second half of the book discusses the major subdisciplines of psychology and the important theoretical controversies of the science.

Publication date: February 1975



Good Reading for Anyone Interested in the Many Dimensions of Psychology





Psychology in Progress

Readings from SCIENTIFIC AMERICAN

With Introductions by

RICHARD C. ATKINSON, Stanford University

Psychology in Progress is a collection of 39 *Scientific American* articles, selected to give readers an appreciation of the scope of modern psychology, an acquaintance with the frontiers of psychological research, and clear explanations of basic psychological principles. Some of the topics included are "Pleasure Centers in the Brain," "Teaching Language to an Ape," "Schizophrenia," "The Effects of Observing Violence," and "Intelligence and Race." The book includes a study guide and helpful introductions to the articles, making this valuable supplementary reading for an introductory course.

Publication date: Spring 1975, cloth $12.00, paper $6.25

A Guide to Psychologists and Their Concepts

VERNON J. NORDBY and CALVIN S. HALL

This book belongs in the library of every reader with more than a passing interest in psychology. Here are clear and accurate discussions of the lives and works of 42 men and women who have helped shape the modern behavioral sciences. The book includes a portrait, biography, summary of major ideas, and list of important writings for each psychologist.

1974, 187 pages, 42 illustrations, cloth $8.00, paper $3.50

A Primer of Drug Action

ROBERT M. JULIEN, University of California, Irvine

Here is a sensible new approach to drug education. Organized along the lines of a textbook of pharmacology, this book presents a medically accurate discussion of the major classes of psychoactive drugs—the sedatives, tranquilizers, stimulants, opiates, and psychedelics—at a level easily comprehended by college and advanced high school students. The style throughout is objective, allowing readers to form their own conclusions about the social and moral issues of drug use and abuse.

Publication date: February 1975, cloth $10.00, paper $4.95

A Teacher's Manual is available.

A Primer of Psychobiology

Brain and Behavior

TIMOTHY J. TEYLER, Harvard University

A Primer of Psychobiology is a brief, informative introduction to what is known about the structure and function of the nervous system and how these relate to behavior. Free of jargon and specialized terminology, this book is designed to be readily understood by the introductory level psychology student and the general reader.

Publication date: January 1975, cloth $6.95, paper $3.50

Conceptual Blockbusting

A Guide to Better Ideas

JAMES L. ADAMS, Stanford University

Conceptual blocks are mental walls that get between a problem and its solution. *Conceptual Blockbusting* is a freewheeling but serious guide to identifying and eliminating these blocks. The book is also a good introduction to the classical theoretical writings on creativity—by Freud, Jung, Maslow, and others—and to the modern techniques of applied creativity such as brainstorming and synectics. The book includes exercises and problems designed to improve the reader's "idea-having" abilities.

1974, 137 pages, 42 illustrations, cloth $5.95, paper $3.45

Some Must Watch While Some Must Sleep

WILLIAM C. DEMENT, Stanford University

This brief, fact-filled book is an exciting introduction to the modern scientific study of sleep, dreams, and sleeping disorders. Sleep appears to be a highly dynamic state with a complexity rivaling waking consciousness. Among the topics discussed are: the "jet lag" phenomenon; the amount of time spent dreaming; sleep learning; sleep deprivation; and how sleeping pills make insomnia worse.

1974, 148 pages, 30 illustrations, cloth $5.95, paper $2.95

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

THE AMATEUR SCIENTIST

Among other things, a greenhouse shade that re-creates the lighting of the Tropics

Conducted by C. L. Stong

Most apparatus for automatically regulating the amount of direct sunlight that enters a room involves the operation of shutters with a servomotor controlled by a photocell. One can also obtain glass that darkens in approximate proportion to the intensity of the incident light, but this material is expensive and generally cannot be found in sheets of large size. Neither alternative appealed to Milan D. Fiske (215 Lake Hill Road, Burnt Hills, N.Y. 12027), who set out to control the sunlight in his greenhouse.

Fiske, who is a research physicist, makes a hobby of growing orchids that he collects in the jungles of Ecuador. The exotic specimens require equivalent light to thrive in Burnt Hills. Fiske hit on the scheme of approximating the correct light by letting the seasonal change in the elevation of the sun control the transmission of direct rays through a uniquely proportioned set of opaque slats at the south end of his greenhouse. He determined the geometry of the slats for the installation at Burnt Hills with the aid of a pocket calculator, but the dimensions for other locations can be worked out in a few hours with paper and pencil. The project was reported last year in *Orchidata,* the official publication of the Greater New York Orchid Society. I am grateful to the society for permission to adapt Fiske's description as follows.

"Most greenhouses need shade of one kind or another at some season or other. Out of the great range of such needs the one to be discussed involves only one small segment. My apparatus is a relatively simple device known as the ecliptic shade. Once put in place it can stay in place without attention season after season.

"Ever since I have had a small greenhouse I have been plagued by the shade problem. One is forever seeking more light or less. In general the problem is too much light at midday and too little in winter. What is needed is something to diminish the noonday sun. Elegance demands that it have no moving parts.

"A promising solution of the problem might be found in the moiré principle. The trouble is that moiré shades can be expensive to make because two identical grids must be built and mounted with respect to each other within close tolerances. Until a better solution presents itself I have settled for an ecliptic shade at the south end of a greenhouse that runs north and south.

"The ecliptic shade is a set of rigid slats in parallel planes. The slats are proportioned so that their width, thickness and spacing and the angle the plane in which they lie makes with the ecliptic cause them to cast shade that blocks out high summer sun and admits nearly full sunlight in winter. The area of the device does not enter into the calculation of its other dimensions. My ecliptic shade is roughly six feet square.

"The details of the construction are largely a matter of the builder's ingenuity. I slotted one-inch by three-inch risers of white pine at the desired angle, nailed in the slats and mounted the completed shade with metal fittings that attached the risers to the greenhouse. The thing has been in place three years now and I like it just fine. If the reader has a notion to try one, the formula and the graphs that are included in the accompanying illustrations can guide the choice of dimensions.

"The quantities included in the formula are for the most part given by the lettered drawing. The formula requires simple calculations that can become tedious if they are made with paper and pencil. Moreover, most readers will not have at hand the elevation of the sun as a function of the season and the time of day. This information is included in *The American Ephemeris and Nautical Almanac,* published annually by the U.S. Government Printing Office. (The publication is sold by the Superintendent of Documents, Washington, D.C. 20402. Copies are also available on the shelves of many public libraries.) I am indebted for the information on the sun's elevation to C. L. Hemenway, director of the Dudley Observatory in Albany, N.Y. The illustrations also include [*page 122*] a set of three graphs that can remove the chore of doing the calculations if the reader lives within a few degrees of latitude 40 degrees north (or south).

"The graphs are plotted on the assumption that the slats are set at an angle (A) of 27 degrees with respect to the horizontal, have a thickness (t) of 3/8 inch and a width (W) of 1¼ inches and are spaced (S) at intervals of 2½ inches. The first graph shows plots of the ratio of slat width to slat spacing (W/S) to depict the effect of altering this quantity.

"One must be careful about two points in the calculations. First, because the midwinter sun lies below the plane of the slats in the early morning and the late afternoon (angle E is smaller than angle A), one must be careful to use the appropriate signs in the trigonometric functions for calculating the transmission of direct sunlight (T), otherwise T will rise above unity and present the designer with the theoretical problem of keeping his greenhouse cool! Second, as the sun's angle of elevation increases, the transmission (T) decreases smoothly to zero, at which point it goes mathematically negative as E gets still larger. This result corresponds to the point at which one slat begins to shade another. Ignore negative 'transmissions.'

"A few numerical examples illustrate the procedure. Assume that the elevation of the sun E is 45 degrees and that the other dimensions are as listed above. Then $T = 1 - [\{(1\frac{1}{4} / 2\frac{1}{2}) \times 1.414\} \times \{.325 + (\frac{3}{8} / 1\frac{1}{4}) \times .951\}] = 1 - (.5 \times 1.414 \times .610) = .569$. Similarly, if E is 10 degrees, then $T = 1 - [\{(1\frac{1}{4} / 2\frac{1}{2}) \times 1.015\} \times \{.305 + (\frac{3}{8} / 1\frac{1}{4}) \times .891\}] = 1 - (.5 \times 1.015 \times .956) = .7$. Working with another example, if E is 60 degrees, T is .099.

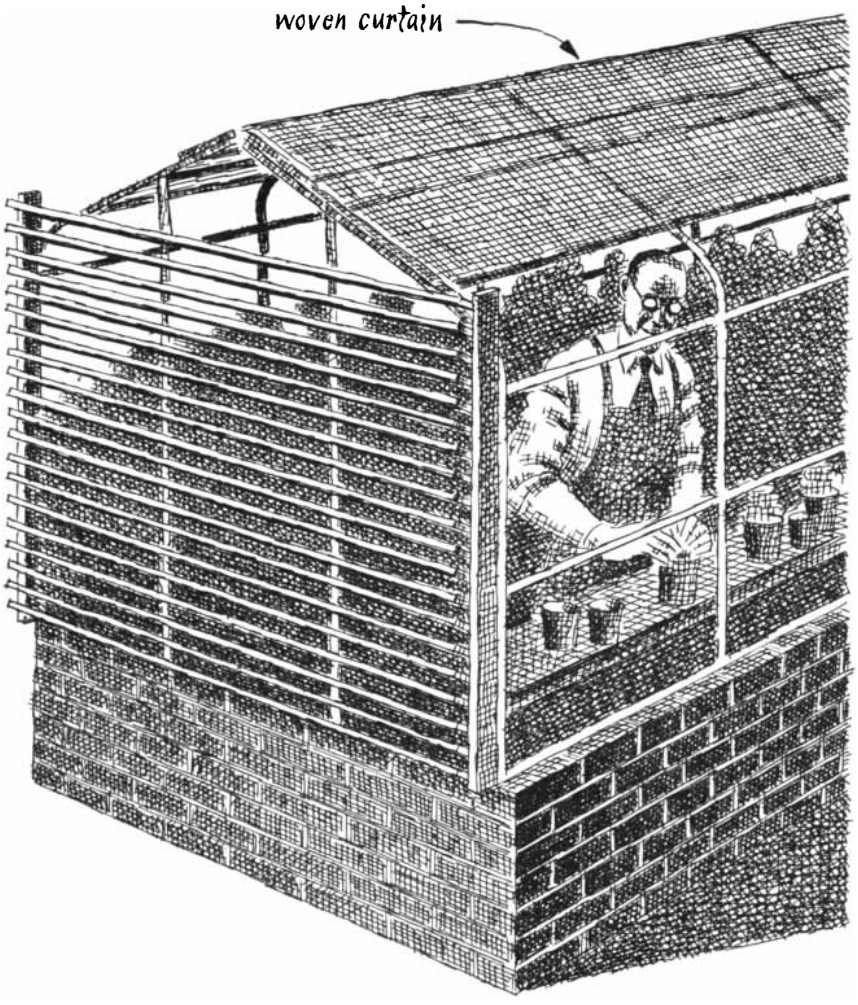
"The second graph plots the transmission (T) against the local time of day for the first day of selected months of the

year. The third graph shows how the transmission varies through the year at various times of day. For example, at 10:00 A.M. in Burnt Hills the transmission of this shade is 79 percent on January 1, 76 percent on March 1, 33 percent on May 1 and only 14 percent on July 1. Observe that the transmission is symmetrical with respect to noon, approximately the same at 11:00 A.M. as at 1:00 P.M., at 10:00 A.M. as at 2:00 P.M. and so on.

"If you build a shade to this prescription, the transmission of direct sunlight will be just what the graphs predict. To the direct sunlight, however, you must add scattered light from the sky that streams through the slats unimpeded, as you can demonstrate with any good venetian blind. In other words, although the sun's direct rays are completely blocked out with a shade of the specified dimensions from 11:00 A.M. to 1:00 P.M. between May 1 and August 4, a lot of light from the sky still gets through. In short, this is a sunshade, not a total-light shade. The resulting illumination is precisely the kind that is needed to prevent leaf burn in orchids without retarding leaf growth.

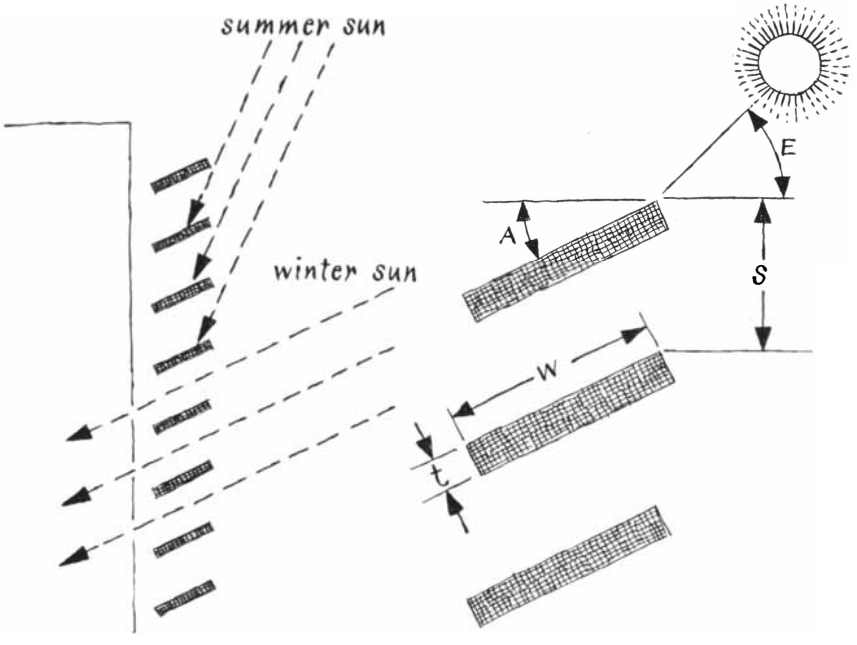
"A shade of the dimensions required for Burnt Hills will work pretty well in most of the continental U.S. On the other hand, readers who have a yen for computing and who own a digital slide rule can with little effort generate graphs for any latitude and combination of shade dimensions. If all this seems like a lot of fuss for an undramatic reward, I can only respond that I think the shade is a dandy and that I fully intend to keep on using mine."

According to theory, big astronomical telescopes show more detail than small ones. The advantage of increased resolution with size is not always realized, however, by the amateur who installs his instrument near communities where the stars must be viewed through polluted air kept aglow by streetlights. Under such conditions a 16-inch reflector shows far less detail than a 3½-inch Questar situated 30 miles from town. For this reason enthusiasts in increasing numbers are substituting modified boat trailers for the imposing backyard observatory that once identified the advanced amateur astronomer. Typical of mobile installations is the 12½-inch $f/8$ Newtonian reflector that Scott Smith, a student at Northwestern University, specifically made two years ago to be transported by a trailer. Smith reports that by keeping an eye on local weather forecasts he can with surprising frequency



Ecliptic shade devised by Milan D. Fiske for the south end of his greenhouse

$$T = 1 - \left[\left\{ (w/s) \sec E \right\} \left\{ \tan (E-A) + (t/w) \cos (E-A) \right\} \right]$$



Formula and principles relating to the slats of the shade

choose a location within reasonable driving range where seeing will be good on almost any date. Smith (11200 60th Avenue SW, Miami, Fla. 33156) describes his rig and its use.

"In general my instrument is conventional. Its equatorial mounting is of the German type in which the upper end of the polar axis carries a bearing that supports the declination axis at a right angle. The shafts for these axes were made of three-inch cold-rolled steel. They turn in heavy bronze bushings.

"The declination bearing was made of a four-inch pipe fitting with a three-inch side inlet. Thrust bearings to support the loads of the polar shaft and the declination shaft were made by sandwiching greased disks of Formica seven inches in diameter between matching pipe flanges. The tube of the telescope is clamped by adjustable bands of galvanized iron to a saddle carried by the declination axis and is counterbalanced by four lead weights of 25 pounds each. The weights were cast in the form of perforated lead disks that slide over the opposite end of the declination shaft, to which they are clamped by pipe strapping.

"At present I am using a commercial clock drive to track the stars. It operates on standard household current. Since I am also interested in observing planets, however, I am working on a new drive. It will include a direct-current motor of the compound type that has a variable but stable tracking response.

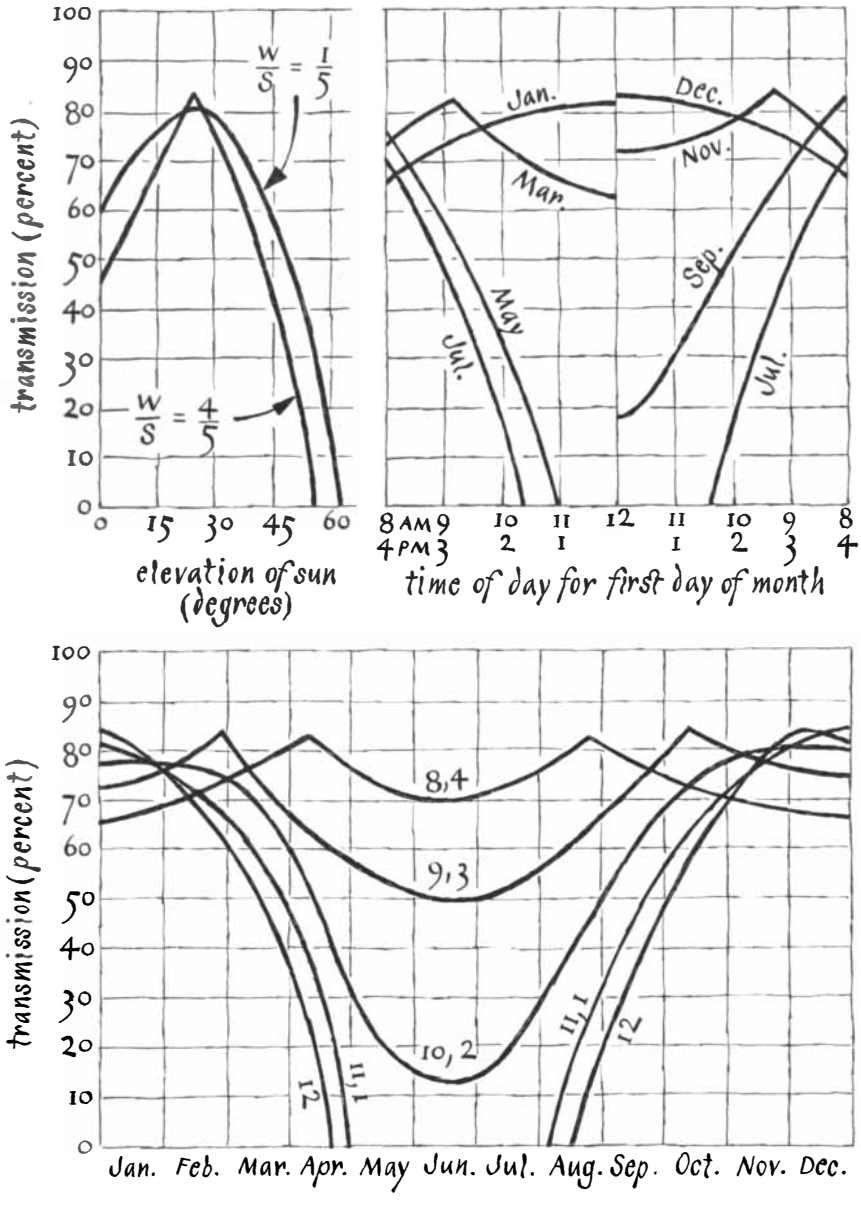
"Notwithstanding heavy construction, it is a sad fact that mountings of the German type are not as free of vibration as one could wish for instruments with apertures larger than eight inches. One can of course use such a large instrument for visual observing, but unless German mountings are protected from the wind they cannot be used for astrophotography that involves long exposures. This is only one of a number of things that must be taken into account when a mobile rig is planned.

"Two major considerations of design for mobile installations are weather and the sustained vibration to which the instrument is subjected when it is moved from site to site. Unlike instruments housed in observatories, a telescope mounted on a trailer is exposed to road dust, weather and the heat of direct sunlight. When the installation is stored in a backyard, it can be shielded from weather with a tarpaulin of canvas or sheet plastic. Remember, however, that coverings of this kind can have a greenhouse effect. Trapped heat can cause parts of the instrument to reach temperatures in excess of 150 degrees Fahrenheit. Therefore one cannot employ in making the telescope many popular plastics and cements that soften at such temperatures. Similarly, one must avoid materials that are subject to damage by prolonged exposure to moisture. Another problem is the possibility that certain fungi will flourish under the tarpaulin.

"With these considerations in mind I made the tube of my instrument from a cardboard form known as a Sonotube. Sonotubes are available from suppliers of heavy construction materials and normally serve as molds for pouring the concrete columns that support highway overpasses. They are manufactured with heavily waxed surfaces.

"I first stripped off the wax so that the cardboard could be soaked with thin epoxy resin of the kind used for impregnating fiberglass. Both the inside and the outside surface were heavily coated. After the resin had set, a layer of fiber and cloth was epoxied to the outside surface. Kits containing fiberglass, resin and full instructions for application are available from dealers in hardware. Tubes thus weatherproofed can be used immediately after the resin sets. One precaution should be observed. Holes drilled through the weatherproofed cardboard during subsequent construction, or any abrasion that would admit moisture to the interior of the cardboard, must be sealed with epoxy.

"All wood parts of the instrument must be similarly weatherproofed. For



Characteristics of the light transmitted by the ecliptic shade

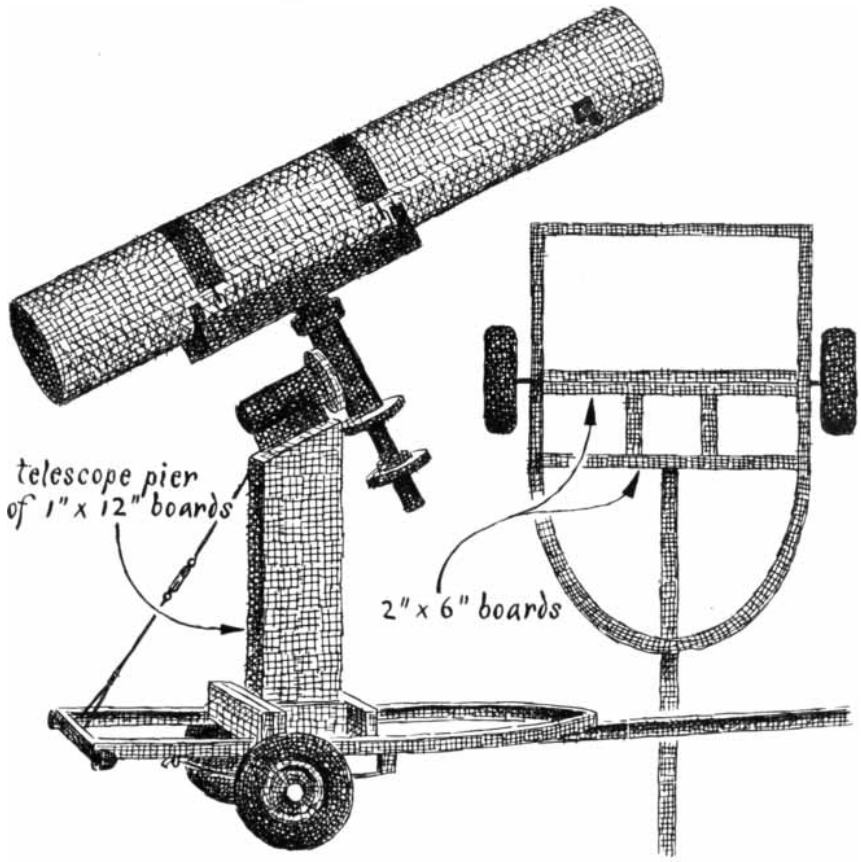
example, the saddle of my telescope, which couples the tube to the declination axis, was made of 3/4-inch plywood cut to the radius of the tube and fastened with screws and waterproof glue to a base of 3/4-inch plywood. The plywood end pieces cradling the tube are braced between two parallel pieces of two-by-four wood. The assembled saddle was heavily coated with fiberglass. All points of contact between the saddle and the tube are covered with pads of indoor-outdoor carpet fastened in place with silicone cement.

"The cell that supports the mirror is also made of moistureproof plywood. It consists of a 3/4-inch disk of plywood 14½ inches in diameter on which the mirror rests. The cylindrical wall of the cell consists of laminated rings of 3/4-inch plywood one inch thick that fit the mirror snugly.

"Four short strips of metal covered with pads of all-weather carpeting are screwed to the upper edge of the cylindrical wall of the cell at equal distances and extend 1/4 inch over the upper edge of the mirror to hold it in place. Three bolts with their heads recessed extend through the base of the cell at separations of 120 degrees. A stiff helical compression spring surrounds each bolt. The threaded ends pass through a 16-inch disk of plywood. Space between this disk and the cell is maintained by the springs.

"The 16-inch disk is rigidly fastened to the end of the telescope tube. A wing nut on each of the bolts can be tightened to put the helical spring under compression and thus adjust the spacing between the cell and the 16-inch disk. The wing nuts therefore adjust the alignment of the optical axis of the mirror with respect to the axis of the tube. All nuts should be of the lock type, including the wing nuts, to minimize changes in adjustment caused by the vibration of road transport.

"Boat trailers are manufactured in a broad range of designs and prices. Doubtless an amateur skilled in the mechanical arts could make one, but he would not save much money. I chose a relatively light vehicle with a wide wheelbase and pliant leaf springs that seemed capable of traveling over many different kinds of terrain. The framework is comparatively open and free of braces that would hamper the movement of an observer. It is a model 1971 Rocket Boat Trailer, one of the most inexpensive rigs that I investigated. I modified it only to the extent of shortening the tongue so that the trailer would be more responsive to small, quick corrections in steering.



Scott Smith's arrangement for making his telescope mobile

"For mounting the pier of the telescope I bolted a platform of two-by-six-inch planks to the bed of the trailer. Two cross members were fastened with *U* clamps above the axle and a single cross member was clamped one foot in front. Fore-and-aft spacers of the same material were then bolted to the cross members to form at the center of the bed a well one foot square and six inches deep. The well supports the pier on which the polar axis of the instrument is mounted.

"The pier is a square tube made of one-inch pine that is closed at the top by a plug four inches thick. The plug consists of laminated pieces of two-by-four wood. When the trailer is level, the upper surface of the plug makes an angle of approximately 26 degrees with respect to the horizontal, which is roughly correct for the latitude (26 degrees north) of my home in Florida, for which the instrument was designed. The telescope and mounting as bolted to this plug weigh about 400 pounds.

"Lag bolts screwed into the plug and through the two-by-six boards of the well anchor the pier to the bed of the trailer. Inertial force set up by the 400-pound instrument during deceleration is counteracted in part by a 3/8-inch cable attached between the top of the pier and the back of the trailer bed. The cable can be adjusted in tension and is easily unhooked and removed for making observations, since at such times it is an obstruction.

"A minimum of two accessories must be included with a mobile rig of this kind. An eight-foot stepladder is essential for reaching the eyepiece when viewing stars near the zenith. I transport the ladder by putting it sideways on the bed of the trailer and lashing it to the pier.

"A minimum of two automobile jacks must also be carried to support the bed rigidly at the front when making observations. Otherwise the flexible leaf springs allow the instrument to jiggle. Other accessories include plastic garbage bags and elastic cords that serve as dustproof and waterproof coverings for both ends of the telescope tube and smaller plastic bags that protect the clock drive and the eyepiece.

"Still further conveniences suggest themselves. For example, the labor of setting up the telescope at frequently visited locations can be minimized by installing a couple of stakes that are accurately aligned on true north. Some

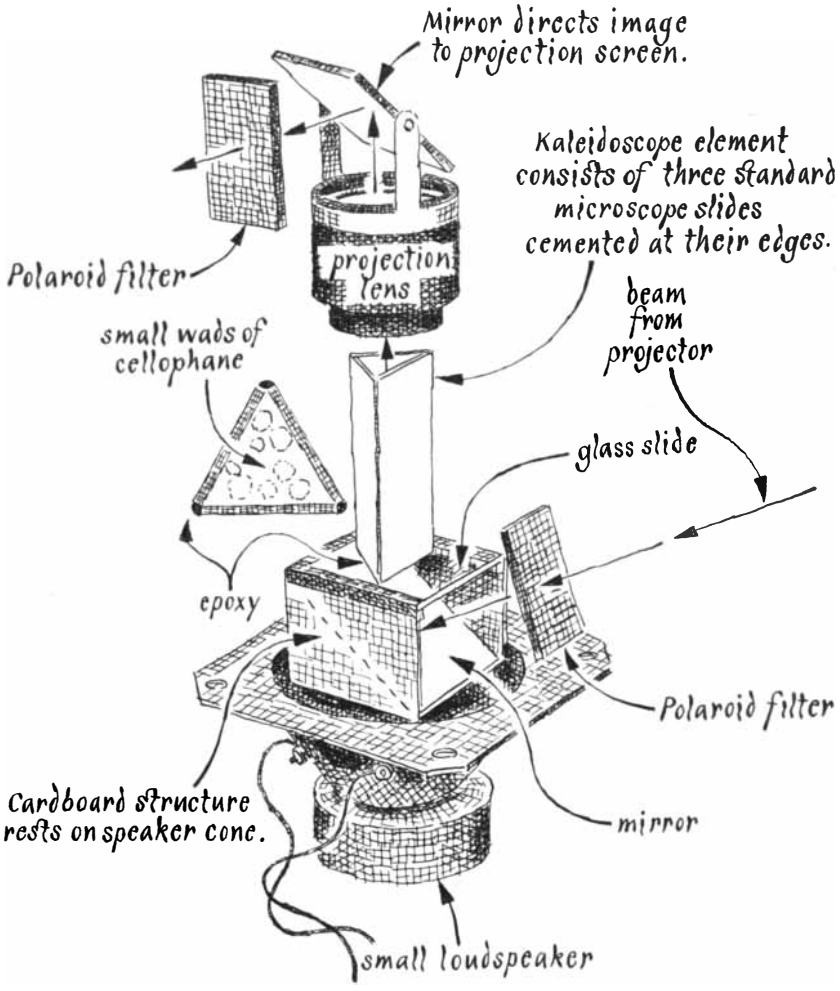
tricks for quickly aligning the polar axis will be found in *Amateur Telescope Making, Book Two,* edited by Albert G. Ingalls."

Don H. Anderson (538 Van Voorhis Avenue, Rochester, N.Y. 14617) has developed a kaleidoscope for projecting symmetrical patterns in color that change form in synchrony with the tempo of music. Anderson, who is director of the industrial laboratory of the Eastman Kodak Company, writes that he made the gadget for the fun of it. He describes it as follows.

"The device involves several variations of the toy that was patented in 1817 by the Scottish physicist Sir David Brewster. Brewster mounted a pair of rectangular plane mirrors facing each other so as to make an angle similar to a partly opened book. He supported the mirrors inside an opaque tube closed at one end with a thin transparent cell that contained loosely trapped fragments of colored glass. A peephole in an opaque disk that closed the opposite end of the tube enabled the observer to view the fragments as images that were multiply reflected by the mirrors. When the mirrors were canted with respect to each other at an even submultiple of 360 degrees, the images appeared as abstract patterns of perfect symmetry. By rotating the tube the observer could cause the fragments to tumble and so could form patterns of endlessly varied detail.

"I have modified Brewster's device by adding a few components that were not available in his time. My basic kaleidoscope consists of three microscope slides, two of which were made into front-surface mirrors with aluminized coatings. They are assembled, along with the uncoated slide, to form a triangular tube of glass with the reflecting surfaces facing inward. The edges are cemented with epoxy.

"In one version of the device one end of the mirror tube is cemented into a closely fitting bottle cap of black plastic. The assembly resembles a triangular box.



Don H. Anderson's projection kaleidoscope

When small objects are dropped into the box and lighted from the side through the transparent slide, they appear as a pattern of threefold symmetry as the observer looks down into the box. Jiggling the box shifts the particles and alters the details of the pattern.

"By strongly lighting the particles, as with a 35-millimeter slide projector, and substituting for the eye an appropriately placed projection lens above the box, an enlarged image of the pattern can be focused on a screen. In other words, the device functions in part as an opaque projector. As a convenience I support a front-surface mirror at an angle of 45 degrees above the device. The mirror deflects the rays horizontally to a wall screen for comfortable viewing.

"My second modification consists of cementing the kaleidoscope through a light framework of cardboard to the paper cone of an inexpensive loudspeaker. Power from a radio or a phonograph drives the loudspeaker and jiggles colored fragments in the plastic cap of the kaleidoscope. The small colored spheres known to confectioners as French's nonpareils, together with short lengths of colored plastic insulation stripped from copper wire of No. 28 gauge, make effective fragments for opaque projection.

"Alternatively the kaleidoscope mirrors can be mounted on a transparent base, such as the two-inch cover glass of a projection slide. This subassembly can in turn be mounted to the cone of the loudspeaker with a cardboard framework. Silicone cement adheres well to glass. A front-surface mirror can be mounted at an angle approaching 45 degrees between the cone of the speaker and the glass slide. A source of light that is incident on the mirror will be reflected upward through the kaleidoscope assembly. If the kaleidoscope is charged with colored bits of transparent plastic, it will function as a conventional projector.

"Another fascinating variation in design consists in polarizing the incident light with a sheet of Polaroid and charging the kaleidoscope with small crumpled wads of cellophane or another highly birefringent material. Rays from the projection lens are similarly transmitted to the screen by a second sheet of Polaroid. When the loudspeaker is energized and the polarizing sheets are approximately adjusted, the details of the projected patterns change continuously in both form and color. I find it very difficult to worry about the problems of the world as I listen to music in the evening and watch the synchronously dancing patterns on my screen."

BOOKS

Is there any real evidence that I.Q. test scores are heritable?

by David Layzer

THE SCIENCE AND POLITICS OF I.Q., by Leon J. Kamin. Lawrence Erlbaum Associates, Publishers, distributed by Halsted Press Division of John Wiley & Sons ($10.95). Conventional wisdom has it that each of us comes into the world equipped with a set of more or less highly differentiated capacities for mental and physical development. These capacities vary greatly from one individual to another—but so do the opportunities our society affords for their development. Inequalities in occupational status and economic rewards therefore reflect environmental as well as genetic differences among individuals. Which of the two sources of inequality is more important? The question obviously has profound social and political implications. If genetic differences are the chief source of social inequalities, one might argue (as many have argued) that efforts to reduce them are misguided. If, on the other hand, environmental factors are to blame, democratic principles would seem to demand far-reaching (and expensive) social reforms. Until recently there was no objective basis for deciding between these two explanations of social inequality, but now (I am still quoting conventional wisdom), thanks to the efforts of psychometricians and quantitative geneticists, the issue can at last be resolved in a scientific way.

Two technical developments are said to be responsible for this breakthrough. One is the I.Q. test, which an American psychologist, Richard J. Herrnstein, calls "psychology's most telling accomplishment." The other is heritability analysis, a statistical technique for analyzing biometric and psychometric data. The I.Q. test (it is maintained) assesses the innate capacities that underlie the acquisition of "higher" cognitive skills: abstract reasoning and problem solving. Heritability analysis is an analytic machine that accepts as inputs statistical correlations between the I.Q. scores of genetically related individuals and yields as output the "heritability of I.Q.," a numerical measure of the relative importance of genetic and environmental sources of the variation of I.Q. within a given population.

Several independent studies of I.Q. heritability have been carried out in recent years, and most of them have concluded that its value is very high: about .8. This implies that no more than 20 percent of the observed variation in I.Q. can be attributed to nongenetic factors, including some, such as intrauterine experience, that are insensitive to education and social forms of intervention. Early in 1969 Arthur R. Jensen, a professor of educational psychology at the University of California at Berkeley, published a summary of the data supporting this conclusion and a discussion of their educational and social implications. He concluded that children who score low on I.Q. tests lack the capacity for abstract reasoning and that efforts to teach them problem-solving skills are not only futile but also unkind; such children, he suggested, should be taught mainly by rote. Jensen also concluded that reported differences between the average I.Q.'s of black children and white children probably reflect genetic differences between the two groups. The article aroused widespread interest, among public officials charged with formulating and implementing educational and social policies as well as among social scientists and educators. Two years later Herrnstein published a popular account of Jensen's discussion of I.Q. heritability, along with his own assessment of its implications (*The Atlantic Monthly,* September, 1971). Herrnstein argued that the elimination of artificial barriers to social and economic mobility would lead to the emergence of hereditary socioeconomic stratification based on I.Q.

Leon Kamin's book is directed against these arguments and the conventional wisdom in which they are rooted. It reaches two major conclusions. "The first stems from a detailed examination of the empirical evidence which has been adduced in support of the idea of heritability, and it can be stated simply. There exist no data which should lead a prudent man to accept the hypothesis that I.Q. test scores are in any degree heritable."

Kamin's second major conclusion represents his answer to the obvious question raised by the first: If the belief in I.Q. heritability has no empirical support, how has it come to be so widely accepted by the experts? Here is Kamin's answer: "The I.Q. test in America, and the way in which we think about it, has been fostered by men committed to a particular social view. That view includes the belief that those on the bottom are genetically inferior victims of their own immutable defects. The consequence has been that the I.Q. test has served as an instrument of oppression against the poor—dressed in the trappings of science, rather than politics. The message of science is heard respectfully, particularly when the tidings it carries are soothing to the public conscience."

The data that have been used to support conventional estimates of I.Q. heritability are of two main kinds: I.Q. correlations between genetically related individuals (particularly identical twins) reared separately and I.Q. correlations between genetically unrelated children reared together. Under certain idealized conditions correlations of the first kind would be produced entirely by genetic similarities and those of the second kind would be produced entirely by environmental similarities. If genetic factors play a major role in the genesis of I.Q. differences, as Jensen asserts and Kamin denies, then the I.Q. correlation between separated identical twins should be large, that between separated fraternal twins somewhat smaller and that between separated first cousins smaller still. And the correlation between the I.Q.'s of unrelated children reared together should provide a direct measure of the environmental contribution to the variation of I.Q.

The problem is more complex than these naïve considerations suggest. One cannot safely assume that genetically related children who have been reared

separately have had uncorrelated environments, since such children are often reared by close relatives living in similar neighborhoods. Nor can one assume that the environments of unrelated children reared together are identical, or even that their genotypes are uncorrelated. These are merely the most obvious of the difficulties one encounters when one comes to grips with the actual data.

Both Jensen and Herrnstein relied heavily on the work of the late Sir Cyril Burt, which, as Kamin remarks, "has had a major impact on all facets of the study of I.Q. heritability. There are, for example, various categories of kinship for which the only existing I.Q. correlations have been provided in Burt's publications. Those publications, and those of his colleagues and students, are almost limitless in number. They furnish us with a veritable treasure of I.Q. data." Kamin has subjected this "treasure" to careful critical scrutiny—and found it to be worthless. Consider, for instance, the I.Q. correlation for separated identical twins, a statistic commonly assumed to furnish the single most direct and reliable estimate of I.Q. heritability. The value reported by Burt in his last paper on the subject, published in 1966, is .771. Now, Burt's twin sample grew over the years as new subjects were added to it. In 1955 it contained 21 pairs, in 1958 "over 30" pairs and in 1966 53 pairs. Yet in all three years the reported I.Q. correlation had the same value, .771! Kamin points out a number of similar coincidences in Burt's reported data. These intimations of subjectivity, disconcerting as they are, are less so than inconsistencies in the procedures by which Burt and his colleagues assessed I.Q. and various environmental factors. Jensen has reported that Burt's original data files have been destroyed, so that it is not now possible to ascertain the sources of the inconsistencies, contradictions and numerical coincidences that abound in the work published since 1955, but Kamin's conclusion seems inescapable: "The numbers left behind by Professor Burt are simply not worthy of our current scientific attention."

The three remaining studies of separated identical twins, all of which deal with adult subjects, are not particularly informative, although the two larger studies (one carried out in the U.S., the other in Britain) do provide clear evidence that the degree of resemblance between the I.Q.'s of separated twins is profoundly influenced by the degree of resemblance between their environments.

Having demolished the central edifice of hereditarian doctrine, Kamin next trains his critical artillery on some of the outbuildings: studies of I.Q. correlations between genetically related people who are not identical twins and studies of adopted children. The difference between the I.Q. correlation for identical twins reared together and that for same-sex fraternal twins reared together enables one to estimate I.Q. heritability under certain simplifying (and, as Kamin rightly points out, unjustified) assumptions. Jensen's review of what he described as "all the major twin studies using intelligence tests" yielded heritability estimates ranging from 47 to 91 percent. Kamin notes, however, that Jensen omitted a number of studies, with "sample sizes comparable to the included studies," that seem equally sound methodologically. When some of these are included, the estimates of I.Q. heritability range from an impossible −7 percent to an impossible +153 percent. As for the studies of adopted children, the observation that "adopted children's I.Q.'s are very much higher than those of their biological parents forces hereditarians to concede some role to environment. There do not appear to be any equivalent data that compel a similar concession by environmentalists. The hypothesis of zero heritability stands unscathed."

Kamin has made a strong negative case. He has gone back to the primary sources and demonstrated with a wealth of circumstantial detail that the data they contain cannot support the interpretation that Burt, Jensen and other hereditarians have placed on them. He has, I think, been less successful in explaining the widespread acceptance of hereditarian arguments by the scientific community. It is undeniable (and understandable) that the most indefatigable students of the genetics of human intelligence, from Sir Francis Galton to Sir Cyril Burt and from Lewis M. Terman to Arthur R. Jensen, have been strongly motivated by a belief in biological determinism and by a Platonic view of what constitutes a just society. It is also true, as Kamin forcefully reminds us in a brilliant chapter titled "Psychology and the Immigrant," that many American psychologists and geneticists have supported and are continuing to support questionable social legislation with even more questionable scientific judgments. Yet Kamin himself states that I.Q. heritability poses "a straightforward scientific question, one which can be answered by a logical analysis of the data." If the scientific issues are as straightforward as all that, why do they seem so ambiguous and value-laden to so many psychologists and geneticists who have no strong political commitments?

I suggest that the perceived ambiguity does not result from (although it may be nourished by) a conflict between class loyalties and scientific objectivity but rather is implicit in the framework within which the issues have usually been discussed—by Kamin as well as by those whose work he criticizes. Both hereditarians and environmentalists accept, implicitly or explicitly, a view expounded by B. F. Skinner in *Science and Human Behavior:* that the scientific study of behavior has no need of a theoretical framework, that it should make do with the weakest possible hypotheses. Kamin considers the weakest possible hypothesis that could account for observed differences in I.Q. to be the "null hypothesis," which attributes all such differences to nongenetic causes. He then argues, in my opinion convincingly, that published experimental data do not compel one to abandon this hypothesis; that is, the data do not establish a causal link between genetic differences and observed I.Q. differences. So far, so good. Let us, however, look at the other side of the coin. Suppose we take as our starting point the hypothesis that 80 percent of the variation in I.Q. is genetic in origin. Are there any data that would compel us to abandon that hypothesis—data establishing a causal link between specific environmental differences and differences in I.Q. scores? As far as I know there are none. (This may help to explain why the views of hereditarians on I.Q. heritability have not been visibly shaken by the collapse of the principal component of their ostensible data base, the work of Burt.) Hereditarians and environmentalists thus hold positions that are equally impregnable.

From another point of view the two positions are equally untenable. I refer to the point of view shared by the two giants of 20th-century psychology, Freud and Piaget: that a science of human behavior must be firmly rooted in biology. It seems to me that both the environmentalist and the hereditarian approach to the problem of human intelligence and its variation are profoundly unbiological. A biological approach would at the very outset recognize these three aspects of the problem:

1. Intelligence is a biological adaptation whose most distinctive characteristic is plasticity. Intelligence manifests itself in the ability to devise effective responses to new and unforeseen environmental challenges and to make creative use of relevant past experience.

Criteria for assessing and comparing intelligent behavior must therefore necessarily vary from one culture or subculture to another. In seeking to devise "culture-free" tests of intelligence psychometricians are pursuing a chimera.

2. Intelligence, like every phenotypically plastic biological character, is not the realization of a genetic blueprint but the outcome of an exceedingly complex interaction of a genetically encoded developmental strategy and a unique environmental history. The outcome of this interaction will of course be affected by variations in both the genotype and its milieu, but there is no reason to suppose a given genotypic variation would produce the same variation in outcome under all environmental conditions, or that a given environmental variation would produce the same variation in outcome for all genotypes. Just such similarity of outcome, however, is the biological meaning of the principal assumption underlying conventional heritability analysis.

3. Heritability is a concept that cannot meaningfully be applied to just any set of numbers someone decides to call measurements. It applies to a certain class of phenotypic characters in populations that satisfy certain rather restrictive conditions. I.Q. scores are not measurements of a character belonging to this class. Although they are valid psychological "measures," they fail to meet the requirements of a biological measurement. And it has recently been shown that, for technical mathematical and biological reasons, the concept of heritability probably does not apply to any phenotypically plastic trait in natural human populations.

In the light of these considerations the debate between environmentalists and hereditarians seems a bit unreal, centering as it does on analyses of inadmissible data by means of an inapplicable theory. Yet the social and educational implications of current hereditarian doctrines are real enough, and Kamin has performed an important service by exposing their scientific inadequacy.

Shorter Reviews

by Philip Morrison

MAMMALS OF THE WORLD, by Ernest P. Walker and associates. Third edition revised by John L. Paradiso. The Johns Hopkins University Press ($37.50). EAST AFRICAN MAMMALS: AN ATLAS OF EVOLUTION IN AFRICA, by Jonathan Kingdon. Volume II, Part A (Insectivores and Bats), Volume II, Part B (Hares and Rodents). Academic Press ($46.75 each). Twelve or fifteen thousand species of our close kin, warm-blooded, hairy and alert, comprise the class Mammalia, found worldwide on the land, in the sea and in the air. Their diversity no less than their relatedness and their size lends them special interest. A bare list of species would be of little use save to specialists, and a formidable work giving even a page to each species would occupy an entire shelf. The late Ernest P. Walker, an enthusiastic mammalogist associated for a lifetime with the National Zoo and the National Museum, devoted some 30 years to the task of putting together a satisfying two-volume compromise. It appeared first in 1964 (along with a bibliographical volume that has not been revised) and is brought up to date for the second time by a colleague in the U.S. Fish and Wildlife Service.

The species of mammals can be grouped into sets of animals judged to be closely related, most of the groups gaining general consensus: the genera of the mammals. Some 900 genera are described here, arranged into orders and families; each genus (with a handful of exceptions) receives a page or more of general description summarizing the appearance, range, behavior and relation to man of typical species, often of several species. An effort has been made to provide a photograph from life to represent every genus; the result has been to display a remarkable number, although some genera must still be represented by a museum skin or an artist's rendering in line or wash. Only recent genera are treated, those with a record of contact during life with our own kind. (The extinct ground sloths just made it!)

They parade past, full of surprises and pleasures, a paper menagerie. Take the genus *Tylonycteris,* the bamboo bats. One species is shown full length and in two close-ups. It is less than three inches long from tip to tail, adapted to enter through narrow cracks into big hollow stems of bamboo, where "as many as 13 have been found roosting by day" in a single joint. Or take the tree sloth. Let no one believe that the sloth, admittedly no traveler, spends its entire life in one tree. Not so, says Charles Handley, sloth-watcher. He has seen them move freely from tree to tree over long limbs and vines and even drag themselves by their arms across open ground for 10 meters or more.

Or consider the three-banded armadillo of the Amazon. Among the armored family only this genus (three species) has the ability to roll into a perfect sphere, completely enclosed, with its head, legs and tail fitted neatly inside; the shelly mammal-melon is shown here very clearly. Consider the courageous, strong, vile-smelling and fearless honey badger. This beast (genus *Mellivora* of the weasel family) has a mutual-assistance pact with the honey guide, *Indicator indicator.* The bird leads the beast to a bees' nest, which the badger shatters so that the allies can share the honey. The honey badger's hide is so loose and thick that it can twist about within its skin enough to "bite an adversary that has seized [it] by the back of the neck." Fangs, quills and stings cannot often penetrate that hide. Caught young, the honey badger can become a fine pet with "incredible strength and energy."

Most of these genera were formally described long ago. It is rare for a new one to be found. The beachcombing zoologist still has a chance, however: the genus *Indopacetus,* a new kind of beaked whale, was created in 1968 on the basis of two skulls thrown up on opposite shores of the Indian Ocean. Somewhat more sadly, the pigmy rock mouse of Ceylon (*Gätamiya*) is known from only one specimen, a little male run over by a car on the road in 1965. Of course the zoologists who first publish proper Latinate descriptions are not the only claimants to sure knowledge of mammals. Local people are generally well acquainted with the mammals of their own forest or savanna. The new mongoose of the Liberian forest, *Liberiictis,* is known in only one species, first described from eight skulls in 1958. The Kran people there call the well-known related cusimanse *quiqui-senna.* ("Quiqui is the way it talks.") Dr. Kuhn of Heidelberg, working from the skulls alone, cannot know any external characteristics of his "new" genus. The Kran, however, have long recognized a second distinct kind of *senna* the size of a water mongoose and completely black, living in groups of from three to five in tree holes of the high forest. "I am sure it will show up one day," writes Kuhn, as *Liberiictis.*

The two-volume mammalian zoo is a reference work of authority and breadth, as plain-looking as an illustrated dictionary. The two volumes of Jonathan Kingdon are quite different, looking rather like a sumptuous new edition of the sketchbooks of some old Florentine master. They bring the reader more of what we had from this unique artist-scientist four years back. The volumes continue to be "singularly handsome... visually without a peer among many years of publishers' output in science." Their content for the eye and the mind is nei-

ther superficial nor library-bound. We read of the direct observations of the author, of the comment of other observers in the field, of the zoogeographic results of present-day work and of the evolutionary lessons to be drawn. Every species mentioned is treated at some length, and we are given wash and line drawings, even watercolors, of the animals in active life, of musculature and skeleton, the wide margins and entire pages crawling and squealing with dormice and slit-faced bats in delicious profusion and diverse antics. Here are the color phases of an elephant shrew mapped onto the soil color by regional types. The entire work is the essence of an informed and reasoned, often quantitative, natural history—or what is fashionably called ecology today.

The details are rich. Here is the larynx of the male hammer bat, which fills most of the animal's chest, so that it has become a "flying loudspeaker" that makes a remarkable clanging sound, at close quarters "like the burst from an electric alarm," far away "a ringing chink." The rats, our commensals, are well treated here. The indigenous East African host of plague (as of the newfound Lassa virus) is the little multimammate rat, "the dominant domestic rat in the more remote villages." It is slowly giving ground to the black rat, the very species that carried plague to medieval Europe. The multimammate rat burrows in floors and walls; the black rat loves the thatch. The black rat began to occupy Uganda in the first decades of the century, spreading from the seaports. Plague of unprecedented severity hit Uganda between 1917 and 1942. The tougher brown Norway rat has now reached the ports and there has displaced the black rat, just as it did in Europe. The brown rat does not carry plague because it is too wary of human contact. Thus does the linkage of rat, flea and man repeat in the Rift Valley the history of 1,000 years of Europe.

The work is much broader than the title indicates, not only in its interests but also in its geographic relevance, since East Africa has both forest and savanna and overlaps with most of sub-Saharan Africa, particularly for the small forms that are covered. Fortunate is the reader or the library that has an interest in drawings, in Africa or in mammalogy and that can manage to buy these expensive but remarkable volumes, a luxurious wonder.

BICYCLING SCIENCE: ERGONOMICS AND MECHANICS, by Frank Rowland Whitt and David Gordon Wilson. The

MIT Press ($12.95). The expensive anchored bicycles on which the businessman seeks exercise and physiologists measure performance are quite deficient. The irony of a fruitless trip in place is not the problem, although it is real enough. The point is that their stationary human engine is not well cooled. A racing cyclist can develop up to 250 watts over a 24-hour stretch, but on the floor-bound ergometer he gives up after half an hour at 150 watts. The windage of open road provides powerful evaporative cooling: no sweat! The other side of that same coin is that the racer's power mostly goes—at speeds higher than about 10 miles per hour—into the cooling wind's air drag; the bearing and linkage friction and the rolling-tire losses are negligible at racing speeds, although they match the air loss at speeds familiar to those who take "the gentle way of the bicycle." Streamlined cycles therefore add some six miles per hour to the usual top speed of 30 m.p.h. for particular races. Here you can see the transparent canopy—very well ventilated below—of a brave cyclist of the Manchester Wheelers moving at a dazzling speed. A cyclist can go more than 100 m.p.h. once his air path is broken by a train or a car; a runner, at much lower speed, would gain only marginally, since his losses have nothing much to do with air friction. (Some actual help may be given the rider by the current behind a moving shield; lateral forces of only a few pounds from overtaking cars can affect cyclists at 70 m.p.h. Stay off the freeway!)

The pedal system is about the best way to take power from the human engine; maybe enabling the hardworking prime mover to use his arms on cranks along with his legs could improve performance a little, but that remains an unsolved acrobatic and technical feat. The diamond frame made of brazed or welded steel tubing evolved nearly 100 years ago and remains virtually unchallenged. Plastics and fiber composites for various reasons seem less suitable, unless ease of production makes them cheaper than steel. Only titanium frames—half the weight of steel and many times the cost—may now outperform steel. The bamboo cycle shown here is no answer. The frame grew in the cycle industry itself, like the wire-suspension wheel, along with much progress in ball bearings. It is no accident that the Wright Brothers were cycle-makers; bicycles helped significantly to "launch the aviation age."

Four chapters on the human engine, then chapter-length treatments of air losses, rolling losses, mechanical friction, braking and the complex mechanics of bicycle steering and balancing, still little understood, bring this brief, handsome and unusual book to its close. It contains a considerable and fascinating set of references to a diverse sporting and engineering literature. The senior author is a British cyclist and designer of long experience, the second author a Massachusetts Institute of Technology engineering professor interested in the human engine in transport. They look a little into the future; it may be that this "wonderful vehicle" will in one way or another contribute more fully to humane and healthful cities. That is on the side of mass utility. For the dreamers there is the hope of the flying bicycle, much sought and not yet realized, the response to the "irresistible challenge" of man-powered flight. The challenge is now fueled in spirit by the $120,000 Kremer Prize, which will go to the first person to complete a figure eight around a one-mile course. It looks as though two-seater vehicles, fragile craft whose wingspan is as great as that of a medium-sized airliner but that weigh less than a man, will yet succeed. The sport is not cheap; so enthusiastic and sophisticated are some teams of amateurs that they work at a development cost of some $250 per pound of machine, more than the cost of standard jets. The modern craft little resemble the intuitive designs that date from the dawn of flight, but their performance still falls short of the old goal. The best flights have covered about half a mile, without the turns. Man will yet fly.

NBS Frequency and Time Broadcast Services. Peter P. Viezbicke, editor. National Bureau of Standards Special Publication 236, 1974 edition. Catalogue No. C13.11:236. U.S. Government Printing Office (60 cents). Nice to the last detail, the broadcast signals by which we Americans distribute our most elegant physical measurement, time, tick off their seconds by no casual tapping. The second pulses are precise trains of just five cycles of a 1,000-cycle wave. (These mark the senior station, WWV in Colorado. Its Hawaiian counterpart uses six cycles of a 1,200-cycle-per-second wave.) On short radio waves and long, at many frequencies, the precision signals travel constantly, including standard audio pitches. Tune your violin to an accuracy of one part in 10^{11}.) Hourly you can also hear up-to-date voice predictions of radio-propagation

conditions, news of major storm locations in the Pacific and the Atlantic, a summary of solar and geophysical events of the day and from time to time other announcements by Federal agencies—of, for example, widely visible high-altitude rocket flights.

This 20-page brochure tells the story, with detailed codes and schedules, a block diagram of how it all works and photographs of the oracle stations themselves. The final resort is the primary frequency standard of the Bureau at Boulder, known as NBS-5, which is a special cesium atomic-beam clock with a current uncertainty "of about 2 parts in 10^{13}." That is about a one-second error in a period as long as the time since our clever forebears chipped stone in Olduvai! The broadcasts themselves are not quite so accurate, controlled only to one part in 10^{11}. Of course the short-wave propagation medium, the inconstant ionosphere, cannot be relied on; such accuracy is perishable in transit. A long-wave coded service at 60 kilocycles supplies the time publicly, with relatively minor propagation errors, to a sponsoring worldwide net of seismic monitors (whose services we doubtless owe to the partial test-ban treaty). It is amusing to see that a Denver television station helps to control the Fort Collins transmissions from the Boulder primary standard: the station's normal horizontal synchronizing pulses provide a common comparison signal. Perhaps this is the most unquestioned program on U.S. commercial television.

The earth's rotation is no longer our time standard. Atomic time (TAI in the official French acronym of the Bureau de l'Heure) is now definitive. These stations announce what is called Coordinated Universal Time (UTC), which more or less steadily gains on the slowing earth, about a second a year. Since astronomers and navigators still watch the stars, UTC is adjusted to fit the earth's rotation time over the long run by using step adjustments of exactly one second—leap seconds—added as directed from Paris, preferably at the last second of the year (or of the fiscal year).

This booklet is indispensable to serious users of WWV and its friends and is of general interest to all concerned with measure. If you have no radio, telephone (303) 499–7111 to hear up to three minutes of the live broadcast. (The telephoned accuracy falls to about 30 milliseconds, since even the gross geography of your call varies from time to time.) This is *not* a toll-free service, but it is cheap enough after 11:00 P.M.

BIBLIOGRAPHY

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

THE ACCURACY OF STRATEGIC MISSILES

REVIEW OF NUCLEAR WEAPONS EFFECTS. Harold L. Brode in *Annual Review of Nuclear Science,* Vol. 18, pages 153–202; 1968.

BALLISTIC MISSILE GUIDANCE. D. G. Hoag. Charles Stark Draper Laboratories, Massachusetts Institute of Technology, June, 1970.

ARMS CONTROL: READINGS FROM SCIENTIFIC AMERICAN. Introductions by Herbert F. York. W. H. Freeman and Company, 1973.

THE ORIGIN OF MIRV—SIPRI RESEARCH REPORT No. 9. Herbert F. York. Stockholm International Peace Research Institute, 1973.

MAKING THE MIRV. Ted Greenwood. Ballinger Press, 1975.

PHYSICS AND CALCULUS OF COUNTERCITY AND COUNTERFORCE NUCLEAR ATTACKS. Kosta Tsipis in *Science,* Vol. 187, No. 4175, pages 393–397; February 7, 1975.

THE MANIPULATION OF GENES

CONSTRUCTION OF BIOLOGICALLY FUNCTIONAL BACTERIAL PLASMIDS *IN VITRO.* Stanley N. Cohen, Annie C. Y. Chang, Herbert W. Boyer and Robert B. Helling in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 70, No. 11, pages 3240–3244; November, 1973.

REPLICATION AND TRANSCRIPTION OF EUKARYOTIC DNA IN *ESCHERICHIA COLI.* John F. Morrow, Stanley N. Cohen, Annie C. Y. Chang, Herbert W. Boyer, Howard M. Goodman and Robert B. Helling in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 71, No. 5, pages 1743–1747; May, 1974.

POTENTIAL BIOHAZARDS OF RECOMBINANT DNA MOLECULES. P. Berg et al. in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 71, No. 7, pages 2595–2599; July, 1974.

FIRST ASM CONFERENCE ON EXTRACHROMOSOMAL ELEMENTS IN BACTERIA in *Microbiology—1974.* American Society for Microbiology, 1975.

REPORT OF THE WORKING PARTY ON THE EXPERIMENTAL MANIPULATION OF MICROORGANISMS. Her Majesty's Stationery Office, London, 1975.

POSITRONS AS A PROBE OF THE SOLID STATE

POSITRON STUDIES OF CONDENSED MATTER. R. N. West. Barnes & Noble, Inc., 1974.

RECENT DEVELOPMENTS IN POSITRON SCINTIGRAPHY. G. L. Brownell and C. A. Burnham in *Instrumentation in Nuclear Medicine: Vol. 2,* edited by G. J. Hine and U. A. Sorenson. Academic Press, 1974.

THE USE OF POSITRONS FOR THE STUDY OF SOLIDS. I. Ya. Dekhtyar in *Physics Reports,* Vol. 9, pages 243–353; 1974.

POSITRON DYNAMICS IN SOLIDS. W. Brandt in *Proceedings of the 3rd International Conference on Positron Annihilation—1973.* Springer-Verlag, Heidelberg, 1975.

THE JOURNAL BEARING

ANALYSIS AND LUBRICATION OF BEARINGS. Milton C. Shaw and Fred Macks. McGraw-Hill Book Company, 1949.

BASIC LUBRICATION THEORY. A. Cameron. Longman, Inc., 1971.

TRIBOLOGY HANDBOOK. Edited by M. J. Neale. Halsted Press, 1973.

THE EFFECTS OF LIGHT ON THE HUMAN BODY

PREVENTION OF HYPERBILIRUBINEMIA OF PREMATURITY BY PHOTOTHERAPY. J. F. Lucey, M. Ferreiro and J. Hewitt in *Pediatrics,* Vol. 41, pages 1047–1056; 1968.

NATURAL AND SYNTHETIC SOURCES OF CIRCULATING 25-HYDROXYVITAMIN D IN MAN. John G. Haddad and Theodore J. Hahn in *Nature,* Vol. 244, No. 5417, pages 515–516; August 24, 1973.

THE EFFECTS OF LIGHT ON MAN AND OTHER MAMMALS. Richard J. Wurtman in *Annual Review of Physiology,* Vol. 37, pages 467–483; 1975.

DAILY RHYTHM IN HUMAN URINARY MELATONIN. H. J. LYNCH, R. J. Wurtman, M. A. Moskowitz, M. C. Archer and M. H. Ho in *Science,* Vol. 187, No. 4172, pages 169–171; January 17, 1975.

THUNDER

LIGHTNING CHANNEL RECONSTRUCTION FROM THUNDER MEASUREMENTS. A. A. Few in *Journal of Geophysical Research,* Vol. 75, No. 36, pages 7517–7523; December 20, 1970.

THUNDER SIGNATURES. A. A. Few in *E ⊕ S Transactions of the American Geophysical Union,* Vol. 55, No. 5, pages 508–514; May, 1974.

HORIZONTAL LIGHTNING. Thomas L. Teer and A. A. Few in *Journal of Geophysical Research,* Vol. 79, No. 24, pages 3436–3441; August 20, 1974.

THE MECHANICAL DESIGN OF TREES

ON GROWTH AND FORM. D'Arcy W. Thompson. Cambridge University Press, 1917.

SIZE AND SHAPE IN BIOLOGY. Thomas McMahon in *Science,* Vol. 179, No. 4079, pages 1201–1204; March 23, 1973.

PATTERNS IN NATURE. Peter S. Stevens. Atlantic Monthly Press/Little, Brown and Company, 1974.

WHY MOSQUITO REPELLENTS REPEL

HOST FINDING AND REPULSION OF *AEDES AEGYPTI.* P. N. Daykin, F. E. Kellogg and R. H. Wright in *Canadian Entomologist,* Vol. 97, pages 239–263; 1965.

ORIENTATION OF *AEDES AEGYPTI* IN VERTICAL AIR CURRENTS. P. N. Daykin in *Canadian Entomologist,* Vol. 99, pages 303–308; 1967.

THE EFFECT OF HUMIDITY ON THE REPELLENCY OF ETHYLHEXANEDIOL, ("6–12") TO *AEDES AEGYPTI.* P. W. Wood in *Canadian Entomologist,* Vol. 100, pages 1331–1334; 1968.

WATER VAPOR AND CARBON DIOXIDE RECEPTORS IN *AEDES AEGYPTI* L. F. E. Kellogg in *Journal of Insect Physiology,* Vol. 16, pages 99–108; 1970.

MATHEMATICAL GAMES

ON PAVING THE PLANE. R. B. Kershner in *The American Mathematical Monthly,* Vol. 75, No. 8, pages 839–844; October, 1968.

TESSELLATIONS WITH PENTAGONS. J. A. Dunn in *The Mathematical Gazette,* Vol. 55, No. 394, pages 366–369; December, 1971. (Also see Vol. 56, No. 398, pages 332–335; December, 1972.)

THE AMATEUR SCIENTIST

AMATEUR TELESCOPE MAKING: BOOKS ONE, TWO AND THREE. Edited by Albert G. Ingalls. Scientific American, Inc., 1950, 1949 and 1953.

The uncommon luxury car.

In a world that frequently rewards conformity, a world filled with lowest-common denominators, Jaguars have always been strikingly individual cars. It has been Jaguar's philosophy to set automotive trends rather than to follow them.

The present Jaguar XJ Series luxury sedans sum up this legendary Jaguar way of doing things. There are two XJs, the XJ6 and the XJ12.

Both XJs share an uncommon degree of luxury. For example, the dashboard of every XJ Series Jaguar is cut from one piece of burled Belgian walnut. The grain is matched by hand and the wood is finished by hand. The air-conditioning system allows you to set your preferred temperature once, after which it regulates itself, regardless of time or season.

And both XJs move and handle with that uncommon blending of quiet comfort and sports car response that is unique to Jaguar. This is because all four wheels have independent suspension and power disc brakes. And because Jaguars steer with a race-inspired and very precise rack and pinion system.

The only difference is under the hood. And there you may choose between two uncommon power-plants.

In the XJ6, you have one of the most famous engines of all time, the Jaguar twin-overhead cam six. The XJ12 is the world's only production V-12 powered sedan. However, this aluminum-alloy engine is not oversized; its displacement is 326 cubic inches as compared to luxury V-8s which displace up to 500 cubic inches.

Uncommon? We are proud to be able to say it.

Drive the very individual Jaguar XJs. For the name of the dealer nearest you, call these numbers toll-free: (800) 447-4700, or, in Illinois, (800) 322-4400.

Jaguar

BRITISH LEYLAND MOTORS INC., LEONIA, N.J. 07605

BRITISH LEYLAND

Don't buy any receiver until you compare its price, power, and specs to these.

Technics' four new stereo receivers. All with impressive specs. And a lot more.

All four have direct coupling. To give you a tighter, cleaner bass.

All with reserve power to float through complex, high-level musical passages without distortion or clipping. Because all have large capacitors, conservatively rated transformers, and bridge rectifiers in the power supplies.

All with a Phase Lock Loop IC and flat group delay filters in the tuner section. For clean, well-separated highs as well as lower distortion on FM. And about 20% less wiring. To reduce hum.

All with Technics' exclusive linear dial scale. For effortless tuning on *both* AM and FM. Negative feedback low distortion tone controls. And all the inputs and outputs you'd expect from Technics.

Whichever Technics receiver you choose, you get all the advantages of Technics' sophisticated engineering, good power, and good specs. And all at a good price.

The concept is simple. The execution is precise. The performance is outstanding. The name is Technics.

SPECIFICATIONS	SA-5150	SA-5250	SA-5350	SA-5550
Price*	$229.95	$299.95	$349.95	$479.95
Min. RMS Power per channel into 8 ohms	16 watts	23 watts	28 watts	58 watts
Bandwidth	40Hz-20kHz	20Hz-20kHz	20Hz-20kHz	20Hz-20kHz
Total Harmonic Distortion (Max.)	0.8%	0.5%	0.5%	0.3%
FM Sensitivity (IHF)	1.9μv	1.9μv	1.9μv	1.8μv
Selectivity (IHF)	70dB	70dB	70dB	70dB
FM Stereo Separation at 1 kHz at 10 kHz	40dB 30dB	40dB 30dB	40dB 30dB	40dB 30dB

*Suggested minimum price, which is the fair trade price in states where Technics products are fair traded.

FOR YOUR TECHNICS DEALER, CALL FREE 800 447-4700.
IN ILLINOIS, 800 322-4400.

Technics
by Panasonic