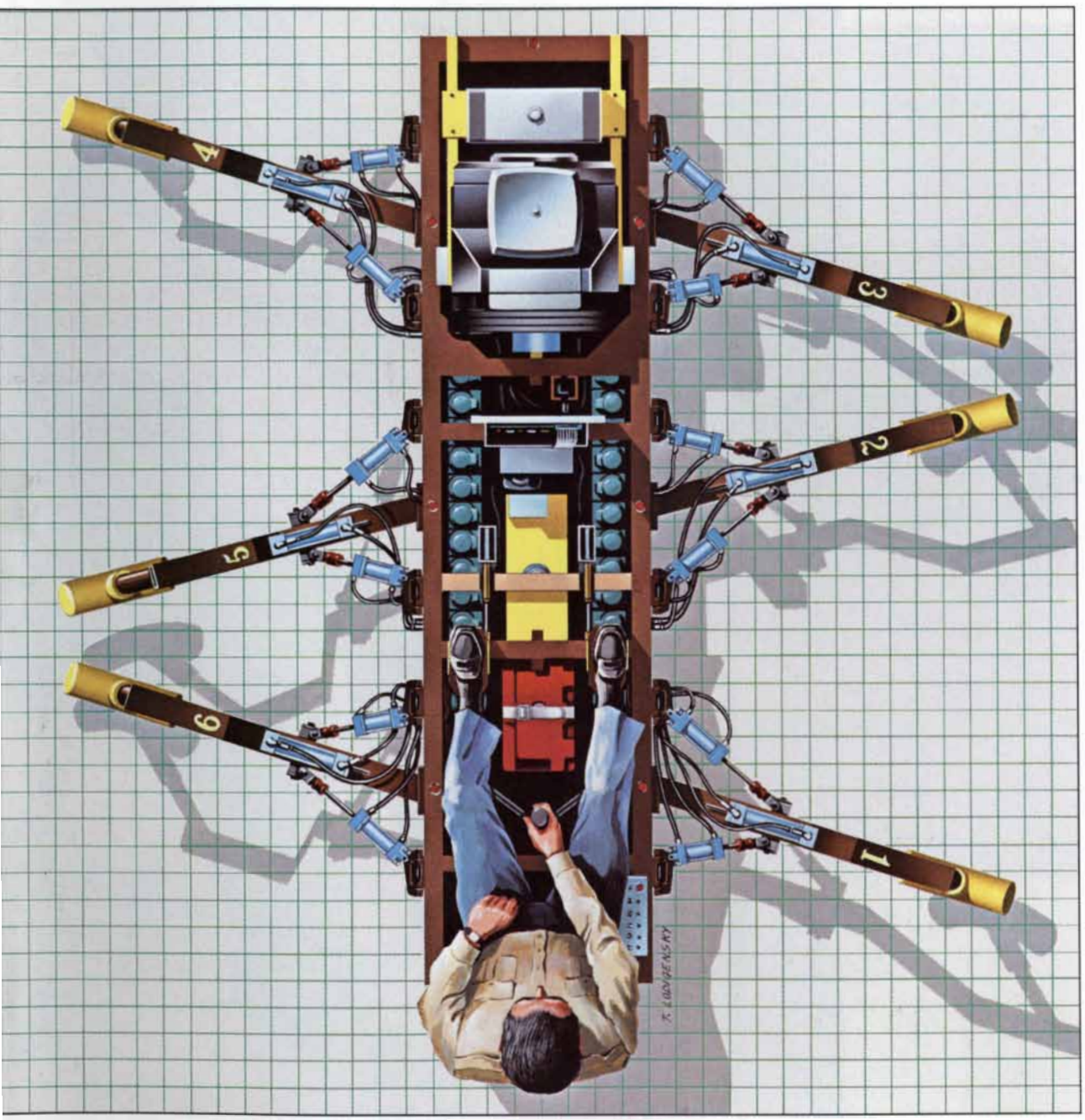
SCIENTIFIC AMERICAN



MACHINES THAT WALK

$2.50

January 1983

Information, please.

Or, how the IBM Personal Computer can bring you the world.

Modem shown not supplied by IBM.

There's a world of information just waiting for you. But to use it, study it, enjoy it and profit from it, you first have to get at it.

Yet the facts can literally be right at your fingertips—with your own telephone, a modem and the IBM Personal Computer.

Plug into the network.

The modem that plugs into your phone takes the codes your IBM Personal Computer understands and turns them into signals that can be transmitted over phone lines. (For specifications, look under "Communications" in the box at right.)

On the other end of the phone lines are independent information services* ready for you to access from your home, school, lab or office. So, in minutes, you can tap the resources of over a thousand data bases—and plug into a world of possibilities.

Something for everyone.

At home, for example, you can shop through an electronic catalog. Access mortgage amortization and tax depreciation schedules. Study the wines of France, the foods of Italy. Find property to sell, rent, swap or share. And play hundreds of games.

At school, you can bring news into the classroom—as it happens. You can teach and learn with facts about everything from phrase origins to plane geometry. You can also tap instructional courses and learn how to program the very IBM Personal Computer you're using.

In the lab, you can call up a wealth of mathematical and statistical programs or retrieve scientific and technological data.

In the office, you can plug into the commodities market. Read abstracts of leading publications. Even get census figures to see who's in your major market. And when you're planning a business trip, your IBM Personal Computer can provide airline schedules and car rental information, as well as tell you what the weather is and where the best restaurants are.

Calling all computers.

And these outside data bases aren't your only information link, either.

IBM PERSONAL COMPUTER SPECIFICATIONS

User Memory
16K-512K bytes*
Microprocessor
16-bit, 8088*
Auxiliary Memory
2 optional internal diskette drives, 5¼", 160K bytes or 320K bytes per diskette
Keyboard
83 keys, 6 ft. cord attaches to system unit*
10 function keys*
10-key numeric pad
Tactile feedback*
Diagnostics
Power-on self testing*
Parity checking*

Display Screen
High-resolution*
80 characters x 25 lines
Upper and lower case
Green phosphor screen*
Operating Systems
DOS, UCSD p-System, CP/M-86†
Languages
BASIC, Pascal, FORTRAN, MACRO Assembler, COBOL
Printer
Bidirectional*
80 characters/second
12 character styles, up to 132 characters/line*
9 x 9 character matrix*

Permanent Memory
(ROM) 40K bytes*
Color/Graphics
Text mode:
16 colors*
256 characters and symbols in ROM*
Graphics mode:
4-color resolution: 320h x 200v*
Black & white resolution: 640h x 200v*
Simultaneous graphics & text capability*
Communications
RS-232-C interface
Asynchronous (start/stop) protocol
Up to 9600 bits per second

*ADVANCED FEATURES FOR PERSONAL COMPUTERS

With assigned passwords, you can also have access to data that's stored in your company's IBM minicomputer or mainframe via your IBM Personal Computer.

You can even exchange a sales forecast with a manager two floors down or two thousand miles away—one IBM Personal Computer to another.

So the IBM Personal Computer really *can* let you call the world. But there's one call to make first. For more information on where to buy the IBM Personal Computer, call 800-447-4700. In Illinois, 800-322-4400. In Alaska or Hawaii, 800-447-0890.

IBM®

The IBM Personal Computer
A tool for modern times

* Information services are available by subscription at additional cost, and are not supplied by IBM.
†UCSD p-System is a trademark of the Regents of the University of California. CP/M-86 is a trademark of Digital Research, Inc.

Established 1845 **SCIENTIFIC AMERICAN** January 1983 Volume 248 Number 1

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), 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: $21 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $27 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at 75 cents each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Student sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

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

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, N.Y. 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

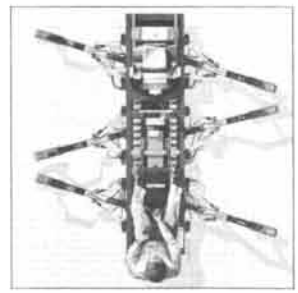
State and ZIP

Old Address

Street

City

State and ZIP



THE COVER

The painting on the cover depicts a six-legged machine that walks, largely under the control of a built-in microprocessor (see "Machines That Walk," by Marc H. Raibert and Ivan E. Sutherland, page 44). More precisely, the machine crawls, meaning that it employs a method of locomotion that does not demand attention to the problem of balance. At least three of the six legs always touch the ground to provide a tripod for support. The microprocessor operates on data supplied by sensing devices that report to it the position of each leg and the forces on the leg. The human driver can steer the machine by making the legs on one side move faster than those on the other. He also can adjust the attitude of the body and its height above the ground and can aid the microprocessor in choosing suitable footholds on rough terrain. Power is provided by an 18-horsepower gasoline engine; hydraulic actuators move the legs. The driver sits in back where he can see the legs. The potential of machines that walk is that they can function in terrain where vehicles with wheels or tracks are ineffective.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
22	Lee Sallows
37	Jack C. Wahl, Monsanto Company
38–42	Jerome Kuhl
43	Jerome Kuhl (*top*); Joseph Feder and William R. Tolbert, Monsanto Company (*bottom*)
45	Leg Laboratory, Carnegie-Mellon University
46–51	George V. Kelvin
52	R. F. Bonifield
53	R. F. Bonifield, courtesy of New York Public Library
55	Richard A. Jacobs, California Institute of Technology
56	Bunji Tagawa
57	A. J. Hudspeth
58–60	Bunji Tagawa
61	A. J. Hudspeth
62–64	Bunji Tagawa
74	Suzanne Swibald
76–81	Patricia J. Wynne
82	Philip J. Currie (*top*), Giuseppe Leonardi (*bottom*)
84	American Museum of Natural History
87–93	Gabor Kiss
95	Pieter Stroethoff
96	Andrew Christie
97	Neville H. Fletcher
98–99	Andrew Christie
100	Pieter Stroethoff
101	Neville H. Fletcher
102–103	Andrew Christie
105	© 1960 National Geographic Society–Palomar Observatory Sky Survey. Reproduced by permission of the California Institute of Technology (*top*); Gary A. Chanan, Columbia University (*bottom*)
108	Ilil Arbel
109	Walken Graphics
110	© 1960 National Geographic Society–Palomar Observatory Sky Survey. Reproduced by permission of the California Institute of Technology (*top*); Gary A. Chanan, Columbia University (*bottom*)
113	Bruce Margon, University of Washington (*left*); Ilil Arbel (*right*)
114–115	Ilil Arbel
120–122	Michael Goodman
123	Hua Jue-ming
124–126	Michael Goodman
127	Hua Jue-ming
128–136	Michael Goodman

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading. It's easy to learn... lasts a lifetime... applies to everything you read... and is the only accredited course with the option of college or continuing education credits.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and re-read words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions — then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading" — but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you — no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week*, *Wall Street Journal*, *Family Circle*, *N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.

Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 15 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 15 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

COLLEGE CREDITS

You may obtain *2 full semester hour credits* for course completion, wherever you reside. Credits offered through Whittier College (California). Details included in your program

CONTINUING EDUCATION UNITS

National Management Association, the world's largest association of professional managers, awards 3.0 CEU's for course completion. CEU's can be applied toward the certificate in Management Studies.

PROFESSIONAL SOCIETIES

Speed Learning is offered internationally to members of professional associations such as: American Chemical Society, Foundation for Accounting Education, Institute of Electrical and Electronics Engineers and dozens more. Consult your Education Director for information.

BUSINESS, INDUSTRY, GOVERNMENT

Many companies and government agencies offer Speed Learning as a wholly-paid or tuition reimbursement program. Consult your Training or Personnel Director for details.

DZ −21

learn INCORPORATED 113 Gaither Drive, Mt. Laurel, NJ 08054

YES! Please rush me the materials checked below:

- ☐ Please send the Speed Learning program @ $99.95 plus $4 postage and handling.
- ☐ Please send the Speed Learning Medical Edition @ $109.95 plus $4 postage and handling.
- ☐ Please send the Junior Speed Learning program (ages 11 to 16) @ $89.95 plus $4 postage and handling.

NJ residents add 5% sales tax

Check method of payment below:

- ☐ Check or money order enclosed (payable to learn incorporated)
- ☐ Charge my credit card under the regular payment terms
  - ☐ Visa ☐ Master Card Interbank No. _____ ☐ American Express
  - Card No. _____ Exp. Date _____

I understand that if after 15 days I am not delighted in every way, that I may return the materials in their original condition for a full refund. No questions asked.

Name _____

Address _____

City _____ State _____ Zip _____

Signature _____

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95 (Includes handling and delivery.)

Check here to order ☐

Outside USA add $10 per item—Airmail extra

SCIENCE/SCOPE

Two communications satellites made history as the first to be launched from NASA's space shuttle. The first, SBS-3, is operated by Satellite Business Systems and will carry high-speed data for many U.S. companies. The second, Anik-C, is operated by Telesat Canada and will improve telephone, television, and data service in Canada. The satellites are versions of Hughes Aircraft Company's HS 376, the world's most widely purchased communications satellite. Hughes now has built 70% of the world's operating commercial communications satellites and has more successes than all other companies combined.

A safety device that snuffs out explosions in the blink of an eye, originally developed for the military, is being applied in commercial situations where fire poses an immediate threat to human life. The Dual Spectrum™ sensing and suppression system has been evaluated in New York Transit Authority toll booths. It detects fire bomb explosions set off by criminals, and suppresses them in one-tenth of a second -- before transit employees can be injured. The system could be applied almost anywhere fire explosions occur within an enclosed area. It was developed by the Santa Barbara Research Center, a Hughes subsidiary.

A new "quick draw" capability for the Maverick missile system would let pilots hit more targets in less time and reduce their risk of being hit by enemy fire. In the last of 21 flight tests, a U.S. Air Force F-4 fighter crew fired three air-to-surface Mavericks within 12 seconds from an altitude of 700 feet. The three missiles scored direct hits on three trucks parked about 70 meters apart in normal convoy fashion. The tests employed a three-rail launcher with a modified electrical system. Although TV-guided versions of the Hughes missile were used in the tests, the system could be used for imaging infrared Mavericks.

Advanced military electro-optical systems are being produced in large numbers and at high rates at a new Hughes manufacturing facility. The complex, which covers one-half million square feet, is designed specifically for making such high-technology devices as infrared night sights and laser rangefinders. Recent milestones include the following deliveries to the U.S. Army: the 2,000th laser tank fire control system for the M60A3 tank, the 1,000th airborne TOW antitank missile system for the Cobra attack helicopter, and the 1,000th thermal imaging system and laser rangefinder for the M1 Abrams tank. In addition, production rates for the two M1 units have reached 70 per month.

Hughes Radar Systems Group has career opportunities for engineers, scientists, and programmers. We design and build many of today's most complex airborne and spaceborne radar electronics systems, including data links, electronic warfare systems, and display systems. We need systems analysts, microwave specialists (antenna, receivers, transmitters, data processors), circuit designers (analog, digital, RF/IF), scientific programmers, mechanical designers, systems and test engineers. Send resume to Engineering Employment, Dept. SSA, Hughes Radar Systems Group, P.O. Box 92426, Los Angeles, CA 90009. Equal opportunity employer.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

(213) 670-1515 EXTENSION 5964

LETTERS

Sirs:

In David L. Waltz's article "Artificial Intelligence" [SCIENTIFIC AMERICAN, October, 1982] he attributes the name of Terry Winograd's computer language—SHRDLU—to the seventh through the 12th most frequent letters in the English language. Of course they are not. The generally accepted progression is E, A, O, I, D, H, N, R, S, T, U, Y....

The origin of the character string SHRDLU is interesting. When Linotype operators made a mistake, they marked the slug of type by running their fingers down the leftmost and second leftmost keys of the machine's keyboard, producing the characters ETAOIN SHRDLU. This was a signal to the compositor that the line was faulty and was to be discarded. Occasionally the bad slug slipped past the compositor and the proofreader and inexplicably appeared in the middle of a newspaper story. It was used as the name of a character in several humorous writings, one by James Thurber, who was a newspaperman before he became an essayist and a cartoonist.

THOMAS BAKERSMITH

Maplewood, Mo.

Sirs:

As a physicist and an avid timpanist I read Thomas D. Rossing's article "The Physics of Kettledrums" [SCIENTIFIC AMERICAN, November, 1982] with great interest. His description of air-mass loading can be used to explain one phenomenon familiar to timpanists: the decrease in pitch for loud kettledrum notes. The louder the note, the greater the effects of air-mass loading and thus the lower the pitch. Hence when a timpanist is tuning, he must take into consideration the loudness of the note to be played. For instance, in the scherzo movement of Beethoven's Ninth Symphony there is a prominent three-note timpani phrase, played first very loud, then very soft. Between these phrases careful timpanists change the tension in the drumhead to compensate for air-mass loading, thereby ensuring that the phrases sound at the same pitch.

Whereas the pitch of most instruments rises with the loudness of the note, the kettledrum is one of the very few for which louder notes become lower in pitch, owing to air-mass loading.

DAVID G. STORK

Swarthmore College
Swarthmore, Pa.

Sirs:

In connection with F. Richard Stephenson's article "Historical Eclipses" [SCIENTIFIC AMERICAN, October, 1982] it may be of interest that a lunar eclipse and a solar eclipse both occurring in one month at the end of the year 1044 were the subject of a poem by Samuel HaNagid (the Prince), a Talmudist, poet and statesman, vizier of Granada during the peculiar golden period of peace between Muslims and Jews in 11th-century Spain. This period and the life of Samuel and of Ibn Gabirol, the poet and philosopher, are fictionalized in *The Green Rose,* by Warren A. Silver (Dial Press, 1977). HaNagid was recognized by his Arab contemporaries as a mathematician and astronomer. Some lines of his poem read in part (translation by T. Carmi, *The Penguin Book of Hebrew Verse,* Viking Press, 1981):

"He who is master of might and beauty, He struck both his luminaries in the very same month. He covered the face of the moon with His terrestrial globe and blocked off the sun with His moon."

MALCOLM S. MASON

Washington, D.C.

Sirs:

The August 11, 1124, solar eclipse, reported as total in Russia, is reported as partial in Flanders as follows:

"In the year 1124 from the Incarnation of our Lord, in the month of August, there was visible to all the inhabitants of the lands an eclipse on the body of the sun at about the ninth hour of the day, and an unnatural failure of light so that the eastern part of the sun, darkened little by little, poured forth strange clouds on the other parts, not darkening the whole sun at the same time, however, but only partially. Nevertheless, the same cloud wandered over the whole circle of the sun, moving across from east to west, but only within the circle of the solar essence."

The reporter is Galbert, a notary in Bruges, writing in the year 1127 or 1128 in his book *The Murder of Charles the Good, Count of Flanders* (Harper & Row, 1967). Because Bruges and Kiev, from where I presume the Russian report originates, are within one degree of latitude from each other, either the difference must stem from the 27 degrees separating them in longitude or the Russian report might be an exaggeration.

LEONARDAS V. GERULAITIS

Associate Professor
Department of History
Oakland University
Rochester, Mich.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1933: "The most recent atomic research offers the heartening possibility of ultimately amalgamating the wave and particle concepts of matter. As a matter of fact, there are *no* experiments which prove that matter possesses *all* the properties of *either* a wave *or* a particle. Accordingly Werner Heisenberg has suggested that the two mental pictures that experiment conjures out of our imagination—one of particles, the other of waves—are both to be viewed simply as incomplete analogies arising from our temporary inability to describe matter in everyday language. Although we cannot draw a satisfactory picture of the atom as it is conceived by wave mechanics, the mathematics of the theory enables us to do all the things that were possible with the Bohr model of the atom. In addition it appears that what were heretofore contradictions are removed, fundamental points are refined and made more precise, and the number of assumptions necessary to attain those ends are reduced to a minimum. From this point of view wave mechanics signifies not so much a radical change as it does a welcome and highly significant evolution of the existent atomic theory."

"Photographs at the rate of 4,000 a second with exposures ranging from 1/100,000 to 1/500,000 of a second have been made at the Massachusetts Institute of Technology by means of a unique electrical circuit that produces light with great actinic intensity. The instantaneous flash of this light is many times more brilliant than the sun. The new circuit, which employs either mercury-arc tubes or spark gaps, was developed by Professor Harold E. Edgerton and Kenneth J. Germeshausen of the Department of Electrical Engineering. The device has already been used to make striking photographs in which familiar things are shown in astonishing new forms. The splash of a drop of milk falling into milk is revealed in the shape of a miniature crown tipped with infinitesimal pearl-like drops. The photographic image recorded at the instant of impact between a golf club and a ball shows clearly the momentary flattening of the latter at the point of contact."

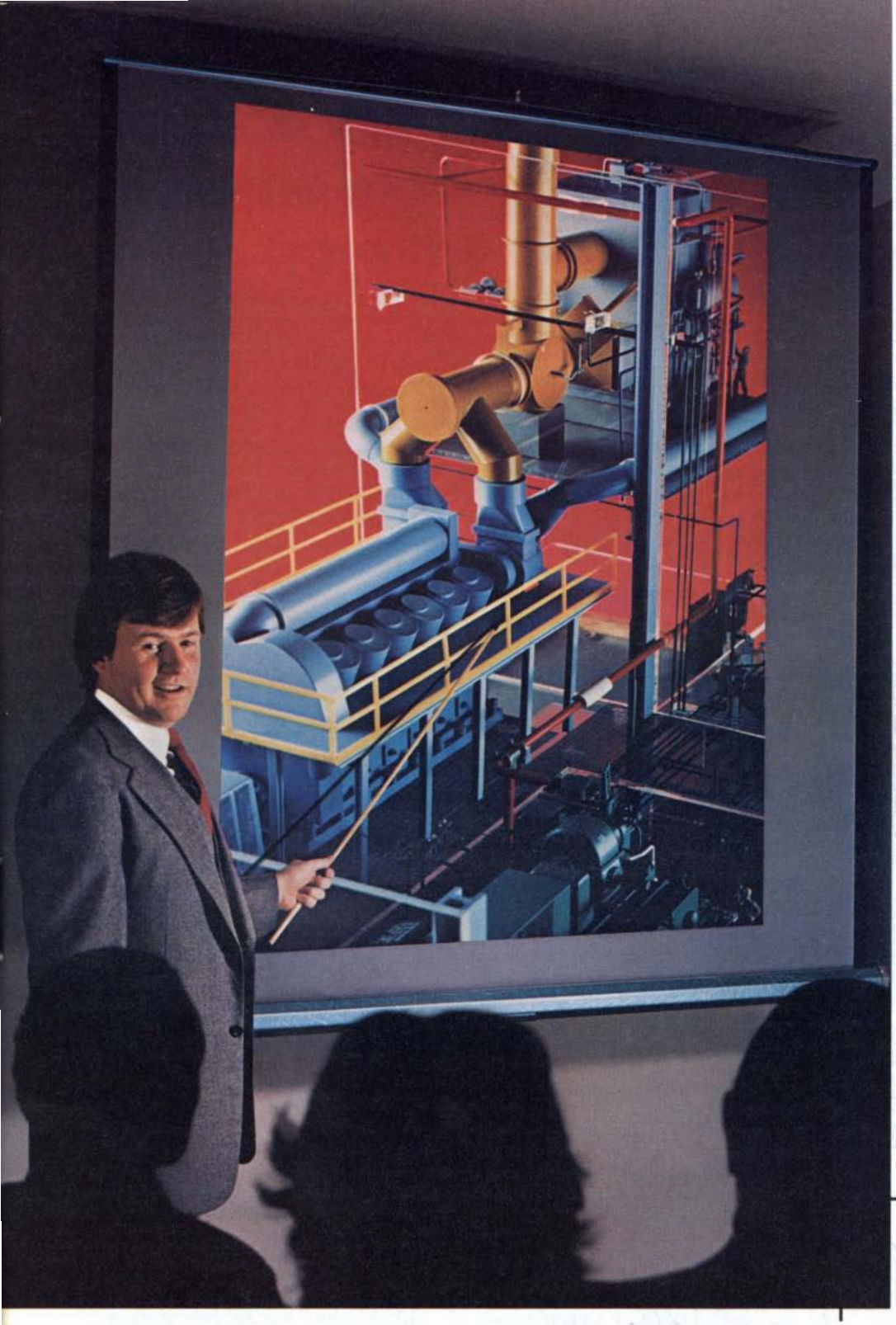
"The plant patent law passed a little more than two years ago has resulted to date in the granting of 39 patents on flowers, fruits and other plant life. Luther Burbank, the famous plant breeder, although dead, has received six patents through his executrix, Elizabeth Burbank of Santa Rosa, Calif. Burbank now has to his credit more plant patents than any other plant breeder. Last April four patents were issued to him covering a new red freestone plum, a pure golden freestone plum, a large, beautifully colored variety of deep carmine plum and a yellow freestone peach. In May he received another patent for a red-skinned, golden-fleshed plum, very large in size. His sixth patent, issued in July, was for a chrome yellow, apricot-like flavored plum."

"The qualifications for the new profession of air hostess are quite severe. The young lady has to be single, under 25, a college graduate and weigh not more than 125 pounds. Apparently the hostess has to move about, and must be light enough not to unbalance the ship! A three-hour flight test for air-sickness and two weeks' observation by pilots are other hurdles. The duties are to attend to passengers, entertain them and divert their minds from the imaginary dangers of a first flight. It is significant of the enterprise and business ability of our transport operators that they do provide such service."

JANUARY, 1883: "A petroleum motor, or rather an engine for obtaining motor power from an explosive mixture of gasoline vapor and air, has been constructed by a firm in Hanover. The working cylinder is eight inches in diameter, with a $14\frac{3}{16}$-inch stroke. The design of the machine is similar to that of a type of gas engine constructed by Wittig and Hees. The gasoline is led through a pipe to the pump cylinder, where it mixes with a definite proportion of atmospheric air. The mixture is then compressed and forced into the working cylinder, where it is ignited by a lamp separately supplied with oil. In four trials with the particular engine in question the maximum force obtained was 4.5 horse power, with 130 revolutions per minute. The consumption of spirit of specific gravity 0.675 was at the rate of $1\frac{3}{4}$ to $2\frac{1}{2}$ pints per horse power per hour. The machine is stated to require as little attention and to work as cheaply as a gas engine."

"The use of dynamo machines in the generation of electricity for public information has now been sufficiently demonstrated to be both practical and economical, especially in large areas where considerable light is required, such as railway stations, harbors and public institutions, not to mention lighthouses, in which the electric light has been used to advantage for a number of years. Next to its importance in the production of the electric light, probably the most useful purpose to which the dynamo has yet been applied is the transmission of power. This property depends on the principle known as the *reversibility* of the dynamo, or, in other words, the transformation of electrical energy into mechanical work, which is simply the converse of converting mechanical into electrical energy. In the transmission of power by electricity the current is generated in one machine, from which it is conveyed to another, which may be at any distance, and may be utilized to set machinery of any kind in motion. The distance over which the electric transmission of power may be carried without serious loss of energy is still a matter of controversy, not having yet been proved by works of any great magnitude. We have it, however, on the authority of Dr. William Siemens, as the result of a long series of experiments made by this distinguished electrician, that the total loss resulting from the double process of conversion (*i.e.*, the conversion of mechanical into electrical energy and secondly the conversion of electrical into mechanical work) is about 20 per cent. To this loss should be added that caused by the resistance of the wires, which depends on their length and sectional area; so that for actual work it is considered safe to assume that the loss sustained over a considerable distance does not exceed 50 per cent., which will not appear very great when compared with the loss sustained in transmitting power by compressed air, water or other means."

"The apparatus employed for separating the various ingredients contained in crude petroleum in the process of obtaining kerosene consists of an iron still having a wrought-iron worm pipe, which is submerged in a tank containing cold water. The still having been filled with crude oil, a fire lighted under it causes the oil to boil and drives off the more volatile vapors. By surrounding the coil with ice, or by compressing these gases by means of an air pump, they may be condensed into the form of very volatile liquids. The average yield of crude Pennsylvania oil is stated to be gasoline $1\frac{1}{2}$, refined naphtha 10, benzine 4, refined petroleum or kerosene 55, lubricating oil $17\frac{1}{2}$, paraffine 2, loss gas and coke 10, total 100. By slow distillation in high stills the production of the heavier oils can be avoided, they being 'cracked' into lighter oils; only crude naphtha, kerosene and coke result."



Introducing Polaroid's Instant Overhead Transparency.

Colorful graphs. Multihued maps. Brilliant photographs.

You can now make management presentations with the powerful impact of color visuals like these, simply, inexpensively, and *instantly.*

Polaroid's Colorgraph 8 × 10 Instant Overhead Transparency System provides you with transparencies that can make your presentations look as professional as you are. And when your information is presented in such a colorful, graphic format, it's bound to be more clearly understood and better remembered. Because color is one of the most effective tools in visual communication.

Besides which, your transparencies can be developed in just 4 minutes. Which means you can present the latest information at the earliest possible moment. Or you can make important changes *moments* before a presentation is scheduled.

What's more, you don't have to send out for your Colorgraph transparencies. You can send in. To the security of your own in-house photo or AV departments. Where your transparencies can be produced easily and quickly from 35mm slides or original art.

The Colorgraph illustrated to the left was made from a 35mm slide in just 4 minutes using Polaroid's new 8 × 10 Polaprinter Instant Slide Copier.

Now full color business presentations can be ready in 4 minutes.

For more information and a demonstration return this coupon to: Polaroid Corp., Industrial Marketing Dept. 553, P.O. Box 5011, Clifton, N.J. 07015. Or call toll free in continental U.S.: (800) 526-7843, Ext. 400. In N.J. call: 800-522-4503, Ext. 400.

Name _____

Title _____

Company _____

Address _____

City _____

State _____ Zip _____

Telephone _____

POLAROID

Plant model courtesy of Thermo Electron Corp. Waltham, MA © 1982 Polaroid Corp. "Polaroid," "Colorgraph" and "Polaprinter"®

THE AUTHORS

JOSEPH FEDER and WILLIAM R. TOLBERT ("The Large-Scale Cultivation of Mammalian Cells") are members of the research department of the Monsanto Company who have worked together on the subject of their article for the past eight years. Feder is senior science fellow at Monsanto. He was graduated in 1953 with a B.S. from Roosevelt University. He went on to obtain his M.S. (1961) and his Ph.D. in biochemistry (1964) from the Illinois Institute of Technology. He is adjunct professor of biochemistry at the University of Missouri at St. Louis. Tolbert is research group leader at Monsanto. His B.S. (1964) is from the University of Richmond. His M.S. in physics (1966) and Ph.D. in biophysics (1971) are from the University of Wisconsin at Madison. In 1971 and 1972 he was research associate at Duke University. From 1972 to 1974 he was postdoctoral fellow at the Allegheny General Hospital; from there he moved to Monsanto. Tolbert describes himself as "an inveterate gadgeteer in [his] approach to scientific problems" whose interests outside the laboratory include the making of wine from fruit that he grows.

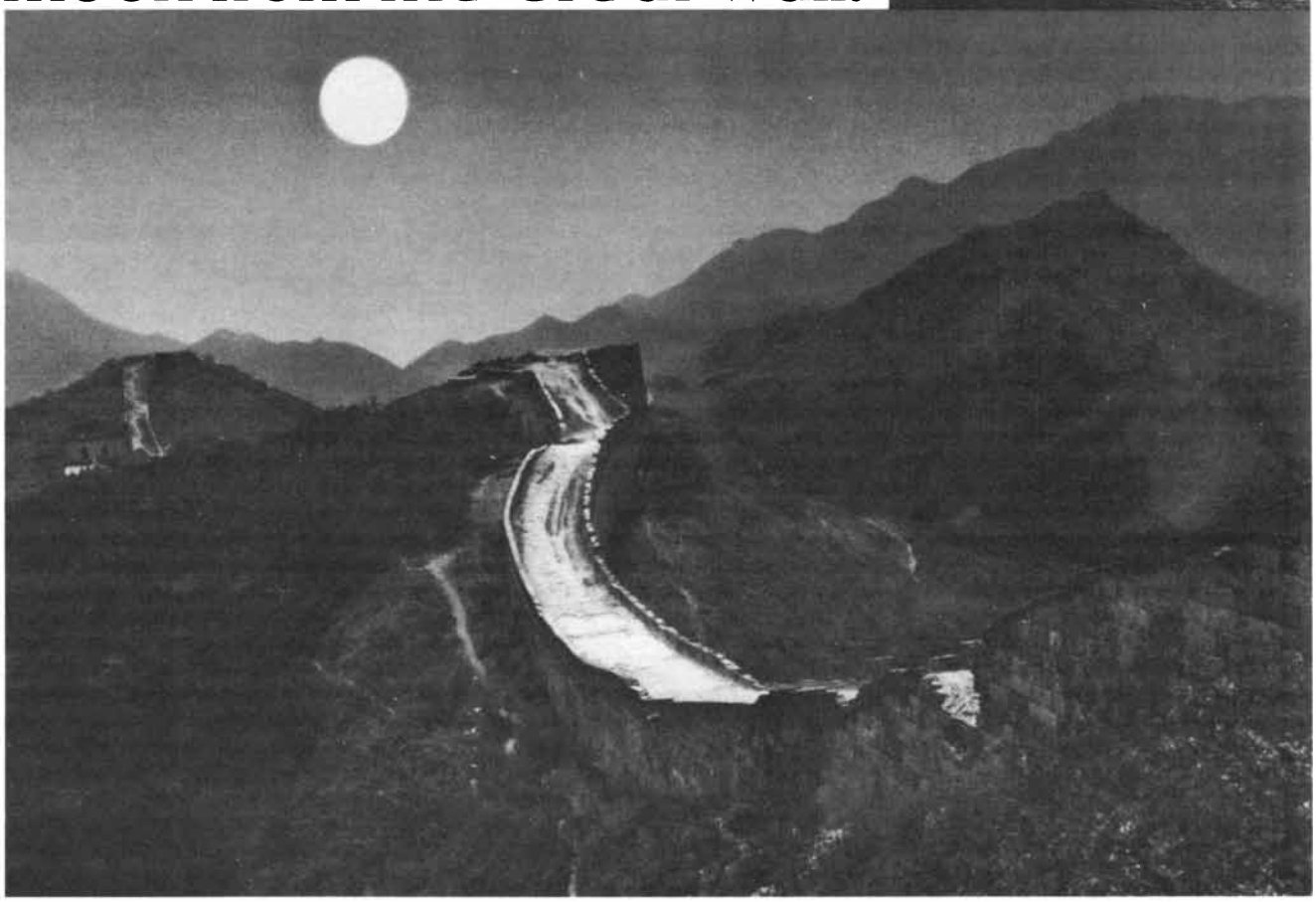
MARC H. RAIBERT and IVAN E. SUTHERLAND ("Machines That Walk") are respectively assistant professor of computer science and robotics and visiting scientist at Carnegie-Mellon University. They have collaborated and competed in the building of machines capable of walking since 1979. Raibert writes: "After finishing my Ph.D. at the Massachusetts Institute of Technology I went to the Jet Propulsion Laboratory of the California Institute of Technology, where I worked on a number of projects in robotics. It was during my stay at J.P.L. that I met Ivan Sutherland, who was then chairman of computer science at Cal Tech. Ivan and I began to talk about projects. I told him I thought we could learn a good deal about locomotion from pogo sticks. I was more than a little surprised when he showed genuine enthusiasm about the idea and coughed up $3,000 for me to build a first model. He himself began to spend substantial time in the shop building a small pneumatically powered walker." Sutherland, who has a Ph.D. from M.I.T. and has been affiliated with several universities and companies, adds: "I worked on robots during my college years but then got involved in other research until Marc Raibert revived my interest in locomotion in 1979. Since then I have been working about half time in robotics. The other half of my time I divide between work in the venture-capital business and the consulting work of Sutherland, Sproull and Associates."

A. J. HUDSPETH ("The Hair Cells of the Inner Ear") is professor of biology at the California Institute of Technology; later this year he will move to the University of California School of Medicine in San Francisco as professor of physiology. He has four degrees from Harvard: a B.A. from Harvard College (1967), an M.A. and a Ph.D. from Harvard University (1968 and 1973) and an M.D. from the Harvard Medical School (1974). In 1974 he was visiting research fellow at the Karolinska Hospital in Stockholm. In 1975 he moved to Cal Tech.

DAVID J. MOSSMAN and WILLIAM A. S. SARJEANT ("The Footprints of Extinct Animals") are geologists with a common interest in the subject of their article. Mossman is a native of Canada whose B.Sc. and M.Sc. are from Dalhousie University in Nova Scotia. He went to New Zealand to continue his education, getting his Ph.D. from the University of Otago. He returned to Canada to join the faculty at the University of Saskatchewan. He taught at Saskatchewan from 1971 to 1982, leaving to take up his current job as associate professor of economic geology at Mount Allison University in New Brunswick. Mossman's major interest is mineral deposits. Sarjeant was born and educated in England; his B.Sc. and Ph.D. are from the University of Sheffield. After serving briefly on the faculties of the University College of North Staffordshire and the University of Reading he went to the University of Nottingham in 1963 and remained there until 1972. In 1972 he emigrated to Canada to become professor of geological science at the University of Saskatchewan. Among his interests is palynology: the study of fossil spores and pollens.

R. G. SHULMAN ("NMR Spectroscopy of Living Cells") is professor of molecular biophysics and biochemistry and chemistry and director of the division of biological sciences at Yale University. He got his education at Columbia University, from which he has three degrees: an A.B. (1943), an A.M. (1947) and a Ph.D. in chemistry (1949). In 1949 and 1950 he was an Atomic Energy Commission fellow at the California Institute of Technology. From 1950 to 1953 he was head of semiconductor research at the Hughes Aircraft Company. From 1953 until 1979 he was at Bell Laboratories; at the end of that time he moved to Yale. In addition to the subject of his article, his research interests

You can see the Great Wall from the moon. Or see the moon from the Great Wall.



Fantastic though it seems the Great Wall of China is actually visible from the moon.

But Salén Lindblad Cruising offers you the chance to see it from a little closer perspective on our China Coastal/Yangtse River cruise aboard the **M.V. Yao Hua.***

You'll travel along the coast of China and up the Yangtse River spending three nights in a lovely Beijing hotel. You'll also visit Hong Kong, Shanghai, Xiamen, Nanjing, Suzhou—the places that make China one of the most intriguing spots under the sun. Or moon.

But this is also a culinary journey: along the way you will travel through China's four distinct cuisines as you travel through China itself. (Excellent Scandinavian food is also served.) And your entire journey will be accompanied by eminent Sinologists.

It's a living experience. A learning experience. One that will tear down walls —even as you visit the greatest wall of all.

Cruises aboard the **Yao Hua** leave weekly, March through November 1983. For more information send us the coupon or talk to your travel agent or call **800-223-5688**. In New York call **(212) 751-2300**.

*Registered in China.

Please send me your brochure on China Coastal/Yangtse River cruises aboard the **Yao Hua**.

Name

Address

City, State, Zip

Phone No.

Travel Agent SA-1

TO LEARN IS TO LIVE.

Salén Lindblad
Cruising, Inc.
133 East 55th Street
N.Y., N.Y. 10022 U.S.A.
Telex: 423661

SALÉN LINDBLAD CRUISING

include radiation damage to DNA, bacterial-virus genetics and paramagnetic metal-ion complexes of nucleic acids.

NEVILLE H. FLETCHER and SUSZANNE THWAITES ("The Physics of Organ Pipes") are respectively professor of physics and a graduate student in physics at the University of New England in Australia. Fletcher obtained two degrees from the University of Sydney: a B.Sc. and a D.Sc. Between getting them he earned his Ph.D. from Harvard University. His original research interests were solid-state physics and the physics of clouds, but in the past 10 years he has given much attention to musical acoustics and the mechanisms of sound production in animals. Fletcher is president of the Australian Institute of Physics and Secretary for Physical Sciences of the Australian Academy of Science. Thwaites has a B.Sc. and an M.Sc. from the University of Western Australia, where she worked on the formation of hail.

BRUCE MARGON ("The Origin of the Cosmic X-Ray Background") is professor of astronomy and chairman of the department of astronomy at the University of Washington. He received his undergraduate education at Columbia University and went on to earn his Ph.D. in astronomy from the University of California at Berkeley in 1973. Before taking up his present job he held appointments at University College London and the University of California at Los Angeles. He is currently an Alfred P. Sloan Foundation research fellow. In 1981 he was awarded the Newton Lacey Pierce Prize of the American Astronomical Society. Margon's main research interests are extrasolar X-ray astronomy, ultraviolet astronomy and observations of the sources of X rays by means of visible radiation.

HUA JUE-MING ("The Mass Production of Iron Castings in Ancient China") is associate research fellow at the Institute of the History of Natural Science of the Chinese Academy of Sciences in Beijing. He studied mechanical engineering at Qinghua University, where he was graduated in 1958. From 1956 until 1964 he did work on the history of smelting and casting in China under Lin Xian-zhou and Zhang Zi-gao. He then moved to the Institute of the History of Natural Science as a graduate student specializing in the history of metalworking. He is a council member of the Chinese Society of the History of Science and Technology and head of the society's committee on the history of metallurgical technique. Hua's recent work has been on the casting of Shang-Zhou bronzes, the construction of pre–Qing Dynasty bells and the cast iron of the Han and Wei dynasties.

METAMAGICAL THEMAS

Virus-like sentences and self-replicating structures

by Douglas R. Hofstadter

Two years ago, when I first wrote about self-referential sentences in this department, I was hit by an avalanche of mail from readers intrigued by the phenomenon of self-reference in its many different guises. I had the chance to print some of those responses a year ago, and that column triggered a second wave of responses. Many of them have illuminated self-reference in new ways. Here I should like to describe the ideas of several people, beginning with two who responded to my initial column with remarkably similar letters: Stephen Walton of New York and Donald R. Going of Oxon Hill, Md.

Walton and Going saw self-replicating sentences as being similar to viruses: small objects that enslave larger and more self-sufficient "host" objects, getting the hosts by hook or by crook to carry out a complex sequence of replicating operations that bring into existence new copies of the virus, which are then free to go off and enslave further hosts. "Viral sentences," as Walton called them, are "those that seek to obtain their own reproduction by commandeering the facilities of more complex entities."

Both Walton and Going were struck by the perniciousness of such sentences: the selfish way they invade a space of ideas and manage, merely by making copies of themselves, to take over a large portion of that space. Why do they not manage to overrun *all* of the space? It is a good question. The answer should be obvious to students of evolution: the sentences do not do so because of competition from other self-replicators. One type of replicator seizes one region of the space and becomes good at fending off rivals; thus a "niche" in idea space is carved out.

This idea of an evolutionary struggle for survival by self-replicating ideas is not original with Walton or Going, although both of them had fresh things to say about it. The first reference I know of to this notion is in a passage by the neurophysiologist R. W. Sperry in an article he wrote in 1965 titled "Mind, Brain, and Humanist Values." He writes: "Ideas cause ideas and help evolve new ideas. They interact with each other and with other mental forces in the same brain, in neighboring brains, and, thanks to global communication, in far distant, foreign brains. And they also interact with the external surroundings to produce in toto a burstwise advance in evolution that is far beyond anything to hit the evolutionary scene yet, including the emergence of the living cell."

Shortly thereafter, in 1970, the molecular biologist Jacques Monod came out with his book *Chance and Necessity.* In its last chapter he wrote of the selection of ideas as follows:

"For a biologist it is tempting to draw a parallel between the evolution of ideas and that of the biosphere. For while the abstract kingdom stands at a yet greater distance above the biosphere than the latter does above the nonliving universe, ideas have retained some of the properties of organisms. Like them, they tend to perpetuate their structure and to breed; they too can fuse, recombine, segregate their content; indeed they too can evolve, and in this evolution selection must surely play an important role. I shall not hazard a theory of the selection of ideas. But one may at least try to define some of the principal factors involved in it. This selection must necessarily operate at two levels: that of the mind itself and that of performance.

"The performance value of an idea depends upon the change it brings to the behavior of the person or the group that adopts it. The human group upon which a given idea confers greater cohesiveness, greater ambition, and greater self-confidence thereby receives from it an added power to expand which will insure the promotion of the idea itself. Its capacity to 'take,' the extent to which it can be 'put over' has little to do with the amount of objective truth the idea may contain. The important thing about the stout armature a religious ideology constitutes for a society is not what goes into its structure, but the fact that this structure is accepted, that it gains sway. So one cannot well separate such an idea's power to spread from its power to perform.

"The 'spreading power'—the infectivity, as it were—of ideas is much more difficult to analyze. Let us say that it depends upon preexisting structures in the mind, among them ideas already implanted by culture, but also undoubtedly upon certain innate structures which we are hard put to identify. What is very plain, however, is that the ideas having the highest invading potential are those that *explain* man by assigning him his place in an immanent destiny, in whose bosom his anxiety dissolves."

Monod refers to the universe of ideas, or what I earlier called idea space, as "the abstract kingdom." Since he portrays it as a close analogue to the biosphere, we could as well call it the "ideosphere."

In 1976 the evolutionary biologist Richard Dawkins published his book *The Selfish Gene,* whose last chapter further develops this theme. Dawkins' name for the unit of replication and selection in the ideosphere—the ideosphere's counterpart to the biosphere's gene—is "meme." Just as a library is an organized collection of books, so a memory is an organized collection of memes. And the soup in which memes grow and flourish—the analogue to the "primordial soup" out of which life first oozed—is the soup of human culture. Dawkins writes:

"Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. As my colleague N. K. Humphrey neatly summed up an earlier draft of this chapter: '... memes should be regarded as living structures, not just metaphorically but technically. When you plant a fertile meme in my mind, you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell. And this isn't just a way of talking—the meme for, say, "belief in life after death" is actually realized physically,

"Microsoft® Multiplan™ allows me to explore more alternatives in less time. I think it leads to better management decisions."

Edward R. Schwinn, Jr.
President, Schwinn Bicycle Company

"Multiplan software helps me make better business decisions. It's as simple as that. With Multiplan on my microcomputer, I can explore a number of alternatives faster. That's really important to me in planning and tracking my totally hand-crafted Schwinn Paramount Bicycle. As with anything handmade, cost control and product planning are major considerations. For instance, Multiplan lets me study the effect of the learning curve on overall productivity. And although we have a mainframe computer, I find that Multiplan is a tool that allows me to personally examine my business options—without spinning my wheels."

A tool for decision makers. Microsoft Multiplan is a microcomputer software program that can help you too. In many ways. You can analyze cash flow. Plan budgets. Forecast income. Manage production. Multiplan allows you to set up an electronic worksheet for whatever your need may be. And, it lets you explore alternatives faster because when you make one change you immediately see the results of that change throughout the worksheet and on any related worksheets. Just change one number and every number that depends on it is adjusted automatically.

In plain English. You won't have to learn a cryptic language to use Multiplan. It takes commands in plain English. And, it will prompt you as you go along by telling you what to do next. There's even a HELP key in case you need help at any point. Press it and it gives you information to get you going again.

Backed by our knowhow. There are a lot of good reasons to use Microsoft software. And among the best is "knowhow." Microsoft put BASIC, the first microcomputer language, on the very first microcomputer, and today, Microsoft software is running on well over a million computers. Worldwide.

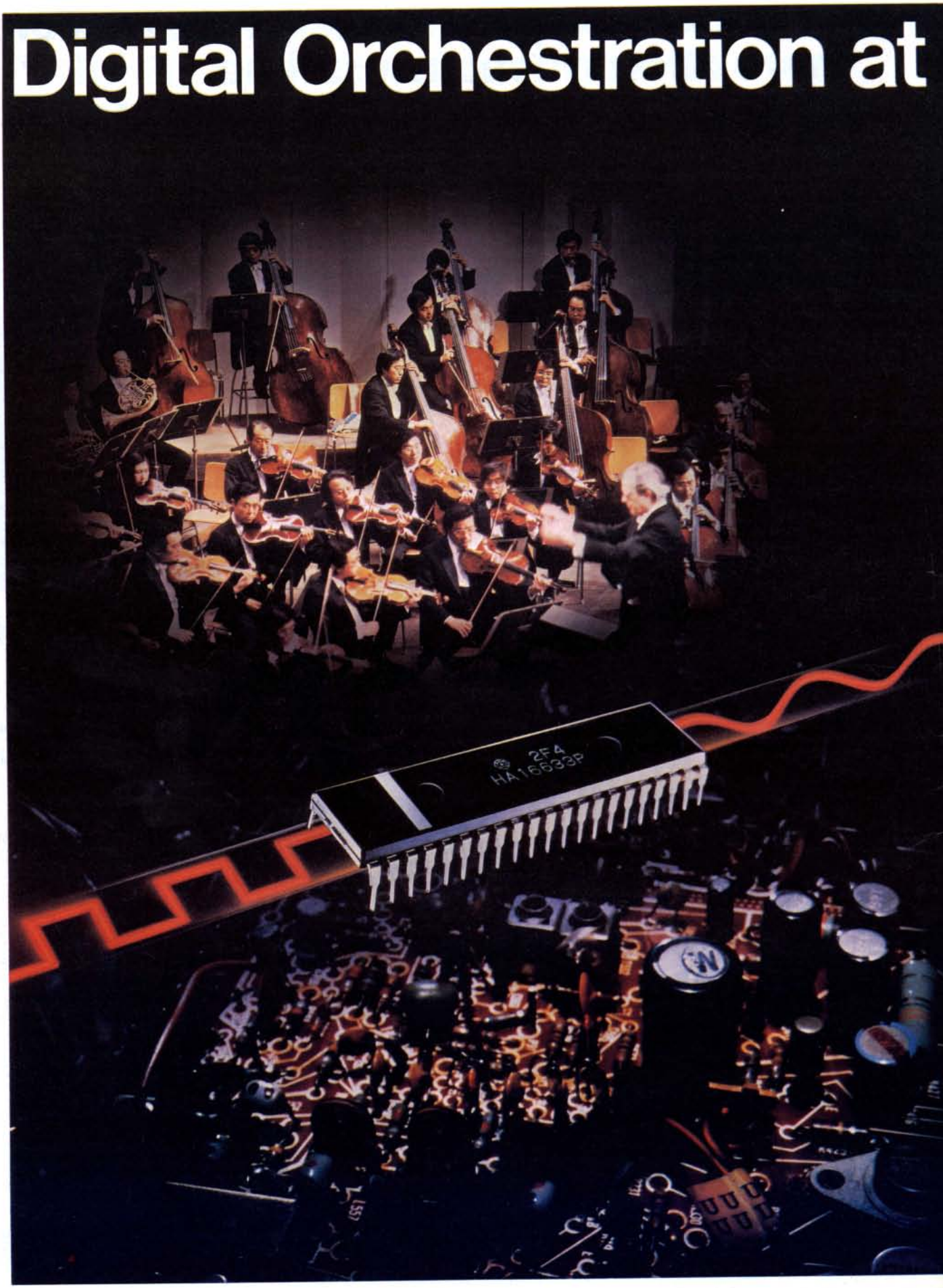
See it in action. If you'd like to explore more alternatives yourself, ask your local computer dealer for a demonstration of Microsoft Multiplan. Multiplan is available for most popular personal computers, including Apple,® the IBM® PC and computers that use CP/M-80® or MS™-DOS operating systems. It's a good management tool for building bicycles. It's also a good management tool for building businesses. Like yours.

BETTER TOOLS FOR MICROCOMPUTERS

MICROSOFT™

MICROSOFT CORPORATION, 10700 NORTHUP WAY, BELLEVUE, WASHINGTON 98004

Digital Orchestration at



the Hands of a Master

Many can play; few can conduct

The latest innovation in audio, the digital audio disc (DAD) player, is a complex package of circuitry which would fill an entire filing cabinet if conventional transistors were used. Large-scale integrated circuits (LSIs) are the obvious way to conserve space and achieve electronic efficiency. But few audio component makers have the technological resources necessary to design and manufacture their own LSIs. That's one reason why DAD players have not been made commercially available. Until now.

Hitachi leads the way with integrated circuitry

Hitachi was able to draw on its own advanced semiconductor capabilities, developed in the course of serving the computer industry, in order to design the DAD player's nine different LSIs. For each LSI, a circuit of discrete components equivalent to some 32,000 transistors had to be assembled in a large rack. Following testing, the circuit was drawn on paper. Then, the drawing was photographically reduced and printed on a silicon wafer—the heart of each LSI. At Hitachi's factories, these original LSIs can be mass-produced at a 99.9% flaw-free rate.

Coordinating the overall performance

One of the most important LSIs in the player is the 16-bit linear digital/analog converter. It's responsible for translating digital pickup signals into analog signals, processing 44,100 sound segments per second, for playback through standard amplifier/speaker systems.

Other key LSIs in the player include those to control the laser tracking mechanism, a data strobe IC, data and control metal-oxide semiconductor LSIs, a 16-bit RAM, a 4-bit microcomputer for system control, and two sample hold ICs for left and right channel output. All together, they form an extremely compact, solid-state system—perfectly orchestrated because it's the product of a single, integrated electronics manufacturer.

Hitachi's DA-1000 digital audio disc player

Technical excellence is embodied in all Hitachi products

The development of LSIs for the DAD player is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from microprocessors and bubble memory devices to a whole host of high-quality electronic appliances. Our comprehensive technical expertise is your guarantee of convenience, easy operation and high reliability in every product that bears the Hitachi brand.

HITACHI

A World Leader in Technology

millions of times over, as a structure in the nervous systems of individual men the world over.'

"Consider the idea of God. We do not know how it arose in the meme pool. Probably it originated many times by independent 'mutation.' In any case, it is very old indeed. How does it replicate itself? By the spoken and written word, aided by great music and great art. Why does it have such high survival value? Remember that 'survival value' here does not mean value for a gene in a gene pool, but value for a meme in a meme pool. The question really means: What is it about the idea of a god which gives it its stability and penetrance in the cultural environment? The survival value of the god meme in the meme pool results from its great psychological appeal. It provides a superficially plausible answer to deep and troubling questions about existence. It suggests that injustices in this world may be rectified in the next. The 'everlasting arms' hold out a cushion against our own inadequacies which, like a doctor's placebo, is none the less effective for being imaginary. These are some of the reasons why the idea of God is copied so readily by successive generations of individual brains. God exists, if only in the form of a meme with high survival value, or infective power, in the environment provided by human culture."

Dawkins takes care here to emphasize that there need not be an exact copy of each meme, written in some universal memetic code, in each person's brain. Memes, like genes, are susceptible to variation or distortion—the analogue of mutation. Various mutations of a meme will have to compete with one another, as well as with other memes, for attention, that is, for brain resources in terms of both space and time devoted to that meme. Memes must compete not only for inner resources but also, since they are transmissible visually and aurally, for radio and television time, billboard space, newspaper and magazine column-inches and library shelf space. Furthermore, some memes will tend to discredit others, and some groups of memes will tend to be internally self-reinforcing. Dawkins writes:

"Mutually suitable teeth, claws, guts, and sense organs evolved in carnivore gene pools, while a different stable set of characteristics emerged from herbivore gene pools. Does anything analogous occur in meme pools? Has the god meme, say, become associated with any other particular memes, and does this association assist the survival of each of the participating memes? Perhaps we could regard an organized church, with its architecture, rituals, laws, music, art, and written tradition, as a co-adapted stable set of mutually assisting memes.

"To take a particular example, an aspect of doctrine which has been very effective in enforcing religious observance is the threat of hell fire. Many children and even some adults believe that they will suffer ghastly torments after death if they do not obey the priestly rules. This is a particularly nasty technique of persuasion, causing great psychological anguish throughout the Middle Ages and even today. But it is highly effective. It might almost have been planned deliberately by a Machiavellian priesthood trained in deep psychological indoctrination techniques. However, I doubt if the priests were that clever. Much more probably, unconscious memes have ensured their own survival by virtue of those same qualities of pseudo-ruthlessness which successful genes display. The idea of hell fire is, quite simply, *self-perpetuating,* because of its own deep psychological impact. It has become linked with the god meme because the two reinforce each other, and assist each other's survival in the meme pool.

"Another member of the religious meme complex is called faith. It means blind trust, in the absence of evidence, even in the teeth of evidence.... Nothing is more lethal for certain kinds of meme than a tendency to look for evidence.... The meme for blind faith secures its own perpetuation by the simple unconscious expedient of discouraging rational inquiry.

"Blind faith can justify anything. If a man believes in a different god, or even if he uses a different ritual for worshipping the same god, blind faith can decree that he should die—on the cross, at the stake, skewered on a Crusader's sword, shot in a Beirut street, or blown up in a bar in Belfast. Memes for blind faith have their own ruthless ways of propagating themselves. This is true of patriotic and political as well as religious blind faith."

When I muse about memes, I often find myself picturing an ephemeral flickering pattern of sparks leaping from brain to brain, shouting "Me, me!" Walton's and Going's letters reinforced this image in interesting ways. For instance, Walton begins with the simplest imaginable viral sentences—"Say me" and "Copy me"—and moves quickly to more complex variations with blandishments ("If you copy me, I'll grant you three wishes") or threats ("Say me or I'll put a curse on you"), neither of which, he observes, is likely to be able to keep its word. Of course, as he points out, this may not matter, the only final test of viability being success at survival in the meme pool. All's fair in love and war—and war includes the eternal battle for survival, in the ideosphere no less than in the biosphere.

To be sure, very few people above the age of five will fall for the simpleminded threats or promises of these sentences. If, however, you simply tack on the phrase "in the afterlife," far more people will be lured into the memetic trap. Walton observes that a similar gimmick is used by the typical chain letter (or "viral text"), which "promises wealth to those who faithfully replicate it and threatens doom to any who fail to copy it." Do you remember the first time you received such a chain letter? Do you remember the sad tale of "Don Elliott, who received $50,000 but then lost it because he broke the chain"? And the grim tale of "General Welch in the Philippines, who lost his life six days after he received this letter because he failed to circulate the prayer—but before he died, he received $775,000"? It is hard not to be just a little sucked in by such tales, even if you wind up throwing the letter out contemptuously.

I found Walton's phrases "viral sentence" and "viral text" to be exceedingly catchy—little memes in themselves, definitely worthy, in my opinion, of replication some 700,000 times in print, and who knows how many times orally beyond that? Walton's own viral text, as you can see before your eyes, has managed to commandeer the facilities of a powerful host: an entire printing press and magazine and distribution service. It has leaped aboard and is now—even as you read this viral sentence—propagating itself madly through the ideosphere.

This idea of choosing the right host is itself an important aspect of the quality of a viral entity. Walton puts it this way: "The recipient of a viral text can, of course, make a big difference. A tobacco mosaic virus that attacks a salt crystal is out of luck, and some people rip up chain letters on sight. A manuscript sent to an editor may be considered viral, even though it contains no explicit self-reference, because it is attempting to secure its own reproduction through an appropriate host; the same manuscript sent to someone who has nothing to do with publishing may have no viral quality at all."

As Walton's letter ends it graciously steps forward from the page and squeaks to me directly on its own behalf: "Finally, I (this text) would be delighted to be included, in whole or in part, in your next discussion of self-reference. With that in mind, please allow me to apologize in advance for infecting you."

Whereas Walton mentioned Dawkins in his letter, Going seems not to have been aware of Dawkins at all, which makes his letter quite remarkable in its close connection to Dawkins' ideas. Go-

ing suggests that we consider, to begin with, Sentence *A:* "It is your duty to convince others that this sentence is true." As he says: "If you were foolish enough to believe this sentence, you would attempt to convince your friends that *A* is true. If they were equally foolish, they would convince their friends, and so on until every human mind contained a copy of *A* in it. Thus *A* is a self-replicating sentence. More particularly, it is the intellectual equivalent of a virus. If Sentence *A* were to enter a mind, it would take control of the mind's machinery and exploit it to produce hundreds of copies of itself in other minds.

"The problem with Sentence *A,* of course, is that it is absurd; no one could possibly believe it. However, consider the following:

System *S:*
Begin:
*S*1: Blah.
*S*2: Blah blah.
*S*3: Blah blah blah.
:
:
:
*S*99: Blah blah blah blah blah blah....
*S*100: It is your duty to convince others that System *S* is true.
End.

Here *S*1 through *S*99 are meant to be statements that taken together constitute a belief system having some degree of coherency. If System *S* taken as a whole were convincing, then the entire system would be self-replicating. System *S* would be particularly convincing if *S*100 were not stated explicitly but held as a logical consequence of the other ideas in the system."

Let us refer to Going's *S*100 as the "hook" of System *S,* for it is by this hook that System *S* hopes to hoist itself onto a higher level of power. Note that on its own a hook that says in effect, "It is your duty to believe me," is not a viable viral entity; in order to "fly" it needs to drag something extra along with it, just as a kite needs a tail to stabilize it. Pure lift goes out of control and self-destructs, but controlled lift can lift itself along with its controller. Similarly, *S*100 and *S*1–*S*99 (taken as a set) are symbiotes: they play complementary, mutually supportive roles in the survival of the meme they constitute together. Now Going develops the theme a little further:

"Are there any real idea systems that behave like System *S?* I know of at least two. Consider the following:

System *X:*
Begin:
*X*1: Anyone who does not believe System *X* will burn in hell.
*X*2: It is your duty to save others from suffering.
End.

"If you believed in System *X,* you would attempt to save others from hell by convincing them that System *X* is true. Thus System *X* has an implicit 'hook' that follows from its two explicit sentences, and so System *X* is a self-replicating idea system. Without being impious, one may suggest that this mechanism has played some small role in the spread of Christianity.

"Note that System *X* is still not credible. It takes quite a bit of skill and luck to produce a believable idea system. A more sinister form of self-replication is Sentence *B:* 'The bourgeoisie is oppressing the proletariat.' If you believed *B,* you would want to liberate the proletariat from the bourgeoisie. You would quickly discover that you were not strong enough for the purpose. You would need the help of thousands of like-minded people. The first step in obtaining their help would be to convince them that *B* is true. Thus a 'hook' follows from Sentence *B,* and *B* is therefore a self-replicating idea system.

"Statement *B* is merely a special case of the generalized statement, Sentence *V:* 'The *villain* is *wronging* the *victim*.' Here the word *villain* must be replaced with the name of some real group (capitalists, communists, imperialists, Jews, freemasons, aristocrats, men, foreigners, etc.). *Victim* must be replaced with the corresponding victim and *wronging* filled in as desired. The result will be a self-replicating idea system for the same reasons as *B* was. Note that each of the suggested substitutions yields a historically attested idea system. In fact, it is difficult to think of a form of political extremism that does not reduce to *V* and therefore to the 'hook,' *S*100. One hesitates to explain real historical events in terms of such a silly mechanism, and yet...."

Going brings his ideas to an amusing conclusion as follows: "Suppose we parody my thesis by proposing Sentence *E:* 'The self-replicating ideas are conspiring to enslave our minds.' This 'paranoid' statement is clearly an idea of type *V.* Thus the thesis seems to describe itself. Further, if we accept *E,* then we must say that this type-*V* idea implies that we must distrust all ideas of type *V.* This is the Epimenides paradox."

It is interesting that all these people who have explored these ideas have given examples ranging from the very small scale of such things as catchy tunes (Dawkins cites the opening theme of Beethoven's Fifth Symphony) and phrases (the word "meme" itself) to the very large scale of ideologies and religions. Dawkins uses the term "meme complex" for these larger agglomerations of memes; I prefer the single word "scheme."

One reason is that it fits well with the usage suggested by the psychiatrist and writer Allen Wheelis in his novel *The Scheme of Things.* Its central character is a psychiatrist and writer named Oliver Thompson, whose darkly brooding essays are scattered throughout the book, interspersed with brightly colored, evocative episodes. Thompson is obsessed with the difference between, on the one hand, "the raw nature of existence, unadorned, unmediated," which he refers to repeatedly as "the way things are," and, on the other hand, "schemes of things" invented by human beings—ways of making order and sense out of the way things are. Here are some of Thompson's musings on that theme:

"I want to write a book... the story of one man whose life becomes a metaphor for the entire experience of man on earth. It will portray his search through a succession of schemes of things, show the breakdown, one after another, of each pattern he finds, his going on always to another, always in the hope that the scheme of things he finds and for the moment is serving is *not* a scheme of things at all but reality, the way things are, therefore an absolute that will endure forever, within which he can serve, to which he can contribute, and through which he can give his mortal life meaning and so achieve eternal life....

"The scheme of things is a system of order. Beginning as our view of the world, it finally *becomes* our world. We live within the space defined by its coordinates. It is self-evidently true, is accepted so naturally and automatically that one is not aware of an act of acceptance having taken place. It comes with one's mother's milk, is chanted in school, proclaimed from the White House, insinuated by television, validated at Harvard. Like the air we breathe, the scheme of things disappears, becomes simply reality, the way things are. It is the lie necessary to life. The world as it exists beyond that scheme becomes vague, irrelevant, largely unperceived, finally nonexistent....

"No scheme of things has ever been both coextensive with the way things are and also true to the way things are. All schemes of things involve limitation and denial....

"A scheme of things is a plan for salvation. How well it works will depend upon its scope and authority. If it is small, even great achievement in its service does little to dispel death. A

scheme of things may be as large as Christianity or as small as the Alameda County Bowling League. We seek the largest possible scheme of things, not in a reaching out for truth, but because the more comprehensive the scheme the greater its promise of banishing dread. If we can make our lives mean something in a cosmic scheme, we will live in the certainty of immortality. Those attributes of a scheme of things that determine its durability and success are its scope, the opportunity it offers for participation and contribution, and the conviction with which it is held as self-evidently true. The very great success of Christianity for a thousand years follows upon its having been of universal scope, including and accounting for everything, assigning to all things a proper place; offering to every man, whether prince or beggar, savant or fool, the privilege of working in the Lord's vineyard; and being accepted as true throughout the Western world.

"As a scheme of things is modified by inroads from outlying existence, it loses authority, is less able to banish dread; its adherents fall away. Eventually it fades, exists only in history, becomes quaint or primitive, becomes, finally, a myth. What we know as legends were once blueprints of reality. The Church was right to stop Galileo; activities such as his import into the regnant scheme of things new being which will eventually destroy that scheme."

Taken in Wheelis' way, "scheme" seems a fitting replacement for Dawkins' "meme complex." A scheme imposes a top-down kind of perceptual order on the world, propagating itself ruthlessly, like Going's System *S* with its 'hook.' Wheelis' description of the inadequacy of all "schemes of things" to fully and accurately capture "the way things are" is strongly reminiscent of the vulnerability of all sufficiently powerful formal systems to either incompleteness or inconsistency—a vulnerability that follows from another kind of 'hook': the famous Gödelian hook, which arises from the capacity for self-reference of such systems, although neither Wheelis nor Thompson makes any mention of the analogy. We shall come back to Gödel shortly.

The reader of *The Scheme of Things* must be struck by the professional similarity between Wheelis and his protagonist. It is impossible to read the book and not surmise that Thompson's views are reflecting Wheelis' own—and yet who can say? It is a tease. Even more tantalizing is the title of Thompson's book, which Wheelis casually mentions toward the end of the novel: it is *The Way Things Are*—a striking contrast to the title of the book in which it is contained. One wonders: What is the meaning of this elegant literary pleat, in which one level folds back on another? What is the symbolism of Wheelis within Wheelis?

Such a twist, by which a thing (a sentence, a book, a system, a person) seems to refer to itself but does so only by allusion to something *resembling* itself, is called "indirect self-reference." You can do this by pointing at your image in a mirror and saying, "That person sure is good-looking!" This one is very simple, because the connection between something and its mirror image is so familiar and obvious-seeming to us that there seems to be no distance whatever between direct and indirect referents: we equate them completely. Hence it seems there is no referential indirectness.

On the other hand, this depends on the ease with which our perceptual systems convert a mirror image into its reverse, and on other qualities of our cognitive systems that enable us to see through several layers of translation without being aware of the layers—like looking through many feet of water and seeing not the water but only what lies at the bottom.

Some indirect self-references are of course subtler than others. Consider the case of Sam and Sue, a couple ostensibly having a conversation about their friends Tina and Tim. It happens that Sam and Sue are having some problems in their relationship, and those problems are quite analogous to those of Tina and Tim, only with the sexes reversed: Sam is to Sue what Tina is to Tim in their respective relationships. Therefore as Sam and Sue's conversation progresses, although on the surface it is about their friends Tina and Tim, on another level it is *actually* about themselves as they are reflected in these other people. It is almost as if by talking about Tina and Tim, Sam and Sue are going over a fable by Aesop that has obvious relevance to their own plight. There are things going on simultaneously on two levels, and it is hard to tell how conscious either of the participants is of the exchange of dual messages—one of concern about their friends, one of concern about themselves.

Indirect self-reference can be exploited in the most unexpected and serious ways. Consider the case of President Reagan, who on a recent occasion of high Russian-American tension over Iran went out of his way to recall President Truman's behavior in 1945, when Truman made some very blunt threats to the Russians about the possibility of the U.S.'s using nuclear weapons if necessary against any Russian move in Iran. Merely by bringing up the memory of that occasion, Reagan was inviting a mapping to be made between himself and Truman and thereby was issuing a not-so-veiled threat, although no one could point to anything explicit. There simply was no way that a sufficiently conscious being could fail to make the connection. The resemblance between the two situations was too blatant.

Does self-reference really come, then, in two varieties—direct and indirect—or are the two types just distant points on a continuum? I would say unhesitatingly that it is the latter. Furthermore, you can delete the prefix "self-," so that the question becomes one of reference in general. The essence is simply that one thing refers to another whenever, to a conscious being, there is a sufficiently compelling mapping between the roles the two things are perceived to play in some larger structures or systems. Caution is needed here. By "conscious being" I mean an analogy-hungry perceiving machine that gets along in the world thanks to its perceptions; it need not be human or even organic. Actually I would carry the abstraction of the term "reference" even further, as follows. The mapping of systems and roles that establishes reference need not actually be *perceived* by any such being; it suffices for the mapping to exist and simply be *perceptible* to such a being.

The movie *The French Lieutenant's Woman* (made from John Fowles's novel of the same name) provides an elegant example of ambiguous degrees of reference. It consists of interlaced vignettes from two concurrently developing stories both of which involve complex romances; one takes place in Victorian England, the other in the present. The fact that there are two romances already suggests, even if only slightly, that a mapping is called for. But much more than that is suggested. There are structural similarities between the two romances: each of them has triangular qualities, and each story focuses on only one leg of the triangle. Moreover, the same two actors play the lovers in both romances, so that you see them in alternating contexts and with alternating personality traits. The reason for this "coincidence" is that the contemporary story concerns the making of a film of the Victorian story.

As the two stories unfold in parallel a number of coincidences arise that suggest even more strongly that a mapping should be made. It is left, however, to the viewer to carry out this mapping; it is never explicitly called for. After a time, though, it simply becomes unavoidable. What is pleasant in this game is the fluidity left to the viewer: there is much room for artistic license in seeing connections, or suspecting or even inventing connections.

Indirect reference of the artistic type

is much less precise than indirect reference of the formal type. The latter arises when two formal systems are isomorphic, that is, when they have strictly analogous internal structures, so that there is a rigorous one-to-one mapping between the roles in the one and the roles in the other. In such a case the existence of genuine reference becomes as clear to us as it is in the case of someone talking about his or her mirror image: we take it as an immediate, pure self-reference, without even noticing the indirectness, the translational steps mediated by the isomorphism. In fact, the connection may seem too direct even to be called "reference"; some may see it simply as identity.

This perceptual immediacy is the reason a famous statement of mathematical logic, Gödel's sentence *G,* is called self-referential. Everyone accepts the idea that *G* talks about a *number g;* the tricky Gödelian step is in seeing that the number *g* plays a role in the system of natural numbers strictly analogous to the role the *sentence G* plays in the axiomatic system it is expressed in. This Wheelis-like oblique reference by *G* to itself by way of its "image" *g* is generally accepted as being genuine self-reference. (Note that we have even a further mapping: *G* plays the role of Wheelis, and its Gödel number *g* that of Wheelis' alter ego Thompson.)

The two abstract mappings that establish *G*'s self-reference but make it seem indirect can be collapsed into just one mapping, following a slogan we might formulate this way: "If *A* refers to *B,* and *B* is just like *C,* then *A* refers to *C.*" For instance, we can let *A* and *C* be Wheelis, with *B* representing Thompson. This makes Wheelis' self-reference a "theorem." Of course, this "theorem" is not rigorously proved, since our slogan has to be taken with a grain of salt. Being "just like" something else is a highly disputable matter.

In a formal context where "is just like" is virtually synonymous with "plays a role isomorphic to that of," however, the slogan can have a strict meaning, and thereby it can justify a theorem more rigorously. In particular, if *A* and *C* are equated with *G,* and *B* is equated with *g,* then our slogan runs: "If *G* refers to *g,* and *g* plays a role isomorphic to that of *G,* then *G* refers to *G.*" Since the premises are true, the conclusion must be true. According to this scheme of things, then, *G* is a genuinely self-referential sentence rather than some kind of logical illusion as deceptive as an Escher print.

Indirect self-reference suggests the idea of indirect self-replication, in which a viral entity, instead of replicating itself exactly, brings into being another entity that plays the same role as it does but in some other system: perhaps its translation into French, perhaps a string of the product numbers of all its parts, together with preaddressed envelopes containing checks made out to the factories where those parts are made, and a list of instructions telling what to do with all the parts when they arrive in the mail.

That may sound familiar to some readers. In fact, it is an indirect reference to the Von Neumann Challenge, the puzzle I posed in this department last January to create a self-describing sentence whose only quoted matter is at the word or letter level rather than at the level of entire quoted phrases. I discovered, as I received candidate solutions, that many readers did not understand what the requirement meant. The challenge came out of an objection to the complexity of the "seed" (the quoted part) in W. V. Quine's version of the Epimenides paradox:

"yields falsehood when appended to its quotation" yields falsehood when appended to its quotation.

To see what is strange here, imagine that you want to have a space-roving robot build a copy of itself out of raw materials it encounters in its travels. One way you could do it would be to make the robot symmetrical, like a human being. You could also make the robot able to make a mirror-image copy of any structure it encounters along its way. Finally, have the robot programmed to scan the world constantly, the way a hawk scans the ground for prey. The search image in the robot's case is that of an object identical with its own left half. The robot need not be *aware* that its target is identical with its left half; the search can go on merrily for what seems to it to be merely a very complex and arbitrary structure. When, after scouring the universe for 17 googolplex years, it finally comes across such a structure, then of course the robot activates its mirror-image-production facility and creates a right half. The last step is to fasten the two halves together, and presto, a copy emerges. It is as easy as pie—provided you are willing to wait 17 googolplex years (give or take a few minutes).

The arbitrary and peculiar aspect of the Quine sentence, then, is that its seed is half as complex (which is to say nearly as complex) as the sentence itself. If we resume our robot parable, what we should ideally like in a self-replicating robot is the ability to make itself literally from the ground up: let us say, for instance, to mine iron ore, to smelt it, to cast it, to make nuts and bolts and sheet metal out of it and so on, and finally to be able to assemble the small parts into progressively larger subunits until a replica is born out of truly raw materials. That was the spirit of the Von Neumann Challenge: I wanted a linguistic counterpart to the "self-replicating robot of the second kind."

In particular, this means a self-documenting or self-building sentence that builds both its halves—its quoted seed and its unquoted building rule—out of linguistic raw materials (words or letters). The mistake many readers made was to present as the seed a long sequence of individually quoted words (or letters) in a specific order, then to exploit that order in the building rule. They might as well have quoted one big long ordered string, as Quine did. The idea of my challenge, in contrast, was that all structure in the built object must arise exclusively out of some principle enunciated in the building rule, not out of the seed's internal structure.

Just as a self-replicating robot in some random alien environment is hardly likely to find all its parts lined up on a shelf in order of assembly but must rely on its "brain" or program to recognize raw parts wherever and whenever they turn up so that it can grab them and therefrom assemble a copy of itself, so the desired sentence must treat the pieces of the seed without regard to the order in which they are listed and yet be able to construct itself in the proper order out of them. Thus it is fine if you enclose the entire seed within a single pair of quotation marks rather than quoting each word individually—all that matters is that the word order of the seed (or better still the letter order) not be exploited.

The seed of the ideal solution would be a long inventory of parts, similar to the list of ingredients of a recipe—perhaps a list of 50 *e*'s, then 46 *t*'s, and so on. Clearly those letters cannot remain in that order; they simply constitute the raw materials out of which the new sentence is to be built. No one sent in a solution whose seed was at such a primordial level. A few people, however, did send in adequate, if not wonderfully elegant, solutions with seeds at the word level. The first correct solution I received came from Frank Palmer of Chicago, who therefore gets the first "Johnnie" award—a self-replicating dollar bill that is given every other decade to the Grand Winner of the Von Neumann Challenge. Unfortunately the dollar bill consumes the entire body of its owner in its bizarre process of self-replication, and so it is wisest to simply lock up the bill to protect oneself from its voracious appetite.

Palmer submitted several versions. In them he utilized uppercase and lower-

case to distinguish between the seed and the building rule respectively. Here is one solution, slightly modified by me:

after alphabetizing, decapitalize FOR AFTER WORDS STRING FINALLY UPPERCASE FIGPBVKXQJZ NONSENSICAL DECAPITALIZE SUBSTITUTING ALPHABETIZING, finally for nonsensical string substituting uppercase words

Let us watch how it works, step by careful step. We must bear in mind that the instructions we are following are the lowercase words printed above and that the uppercase words are not to be read as instructions. Nor, for that matter, are the lowercase words we shall soon be working with. They are like the inert, anesthetized body of a patient being operated on, who, when the operation is over, will wake up and become animate. So let's go. First we are to alphabetize the seed. (I am treating the comma as attached to the word it follows.) This gives us the following:

AFTER ALPHABETIZING, DECAPITALIZE FIGPBVKXQJZ FINALLY FOR NONSENSICAL STRING SUBSTITUTING UPPERCASE WORDS

Next we are to decapitalize it. This will yield some lowercase words—the "anesthetized" lowercase words I mentioned above:

after alphabetizing, decapitalize figpbvkxqjz finally for nonsensical string substituting uppercase words

All right; now our final instruction is to locate a nonsensical string (that's easy: "figpbvkxqjz") and substitute for it the uppercase words (namely a copy of the original seed itself). This yields:

after alphabetizing, decapitalize FOR AFTER WORDS STRING FINALLY UPPERCASE FIGPBVKXQJZ NONSENSICAL DECAPITALIZE SUBSTITUTING ALPHABETIZING, finally for nonsensical string substituting uppercase words

And this is a perfect copy of our starting sentence! It can now wake up from its anesthesia and replicate itself in turn.

The critical step was the first one: alphabetization. This turns the arbitrarily ordered seed into a grammatical, meaningful sentence—merely by mechanically exploiting a presumed knowledge of the ABC's. But why not? It is perfectly reasonable to presume superficial typographical knowledge about letters and words, since such knowledge deals with printed material as raw material: purely syntactically, without regard to the meanings carried therein. This is just like the way enzymes in the living cell deal with the DNA and RNA they chop up and alter and piece together again: purely chemically, without regard to the "meanings" carried therein. Just as chemical valences and affinities and so on are taken as givens in the workings of the cell, so alphabetical and typographical facts are taken as givens in the Von Neumann Challenge.

In Palmer's solution the seed happens to have been ordered by the length of words, but that is inessential; any random order would have done, and that is the crucial point many readers missed. Another rather elegant solution was sent by Martin Weichert of Munich. It runs this way:

Alphabetize and copy in quotes "in and copy quotes Alphabetize"

It works on the same principle as Palmer's sentence and again features a seed whose internal structure is irrelevant to the self-replication. Weichert also sent a palindromic solution in Esperanto, in which the flexible word order of the language plays a key role. Michael Borowitz and Bob Stein of Durham, N.C., sent a solution similar to Palmer's.

Finally, last year's gold-medal winner for self-documentation, Lee Sallows, was a bit piqued by my suggestion that the gold on his medal was somewhat tarnished since he had not paid close enough attention to the use-mention distinction. Apparently I goaded him into constructing an even more elaborate self-documenting sentence, reproduced at the left. Although it is not self-replicating, since it does not spell out its own construction explicitly, it is another marvelous Sallowsian gem, and I shall allow the gold on his medal to go untarnished this year. (I hope this will satisfy those purists who wrote to me insisting that gold does not tarnish!)

```
                                *
                             Write
                         down ten 'a's,
                     eight 'c's, ten 'd's,
              fifty-two 'e's, thirty-eight 'f's,
         sixteen 'g's, thirty 'h's, forty-eight 'i's,
    six 'l's, four 'm's, thirty-two 'n's, forty-four 'o's,
    four 'p's, four 'q's, forty-two 'r's, eighty-four 's's,
seventy-six 't's, twenty-eight 'u's, four 'v's, four 'W's,
 eighteen 'w's, fourteen 'x's, thirty-two 'y's, four ':'s,
         four '*'s, twenty-six '-'s, fifty-eight ','s,
               sixty '('s and sixty ')'s, in a
                    palindromic sequence
                        whose second
                          half runs
                            thus:
                            :suht
                          snur flah
                        dnoces esohw
                    ecneuqes cimordnilap
               a ni ,s')' ytxis dna s'(' ytxis
         ,s',' thgie-ytfif ,s'-' xis-ytnewt ,s'*' ruof
    ,s':' ruof ,s'y' owt-ytriht ,s'x' neetruof ,s'w' neethgie
  ,s'W' ruof ,s'v' ruof ,s'u' thgie-ytnewt ,s't' xis-ytneves
    ,s's' ruof-ythgie ,s'r' owt-ytrof ,s'q' ruof ,s'p' ruof
    ,s'o' ruof-ytrof ,s'n' owt-ytriht ,s'm' ruof ,s'l' xis
         ,s'i' thgie-ytrof ,s'h' ytriht ,s'g' neetxis
              ,s'f' thgie-ytriht ,s'e' owt-ytfif
                    ,s'd' net ,s'c' thgie
                       ,s'a' net nwod
                            etirW
                              *
```

Lee Sallows' self-documenting sentence

Osborne™ brings you the comparison IBM® and Apple® don't want you to see.

Other computer companies dazzle buyers with an array of options and add-ons that makes the final price hard to determine and makes the computer hard to buy, complex to assemble, and very difficult to carry.

We believe in making personal computers that are easy to learn and use. And that starts with making computers easy to *buy*.

The Osborne 1™ Personal Business Computer. One simple price, $1795, buys it all.

And it all comes in a portable case you can take with you wherever you work. Because once you go to work with an Osborne, you won't want to work any other way.

For your nearest dealer, call (in California) 800 772-3545, ext. 905; (outside California) call 800 227-1617, ext. 905.

OSBORNE™
COMPUTER CORPORATION

	OSBORNE 1™	IBM PERSONAL®	APPLE II®
Computer with 64K RAM, two floppy drives[A], keyboard and CRT:	$1795	$3240[B]	$3120[C]
Serial communications:	INCLUDED	EXTRA COST	EXTRA COST
Modem Connection:	INCLUDED	EXTRA COST	EXTRA COST
IEEE 488 Instrument communications:	INCLUDED	EXTRA COST	EXTRA COST
BASIC interpreter[D]:	INCLUDED	INCLUDED	INCLUDED
Business BASIC[E]:	INCLUDED	EXTRA COST	EXTRA COST
CP/M® Control Program:	INCLUDED	EXTRA COST	F (see below)
Word Processing[G]:	INCLUDED	EXTRA COST	EXTRA COST
Electronic Spreadsheet[H]:	INCLUDED	EXTRA COST	EXTRA COST
Carrying Case:	INCLUDED	EXTRA COST	EXTRA COST
TOTAL PRICE[I]:	**$1795**	**$4000–4700**	**$4000–4700**

A. The Osborne 1™ includes two built-in 100K byte floppy disk drives. The IBM® and APPLE II® drives provide approximately 160K bytes of storage. **B.** From the IBM Product Center Personal Computer Price Schedule. **C.** From the Apple Computer Suggested Retail Price List. **D.** The Osborne includes MBASIC® from Microsoft. **E.** The Osborne includes CBASIC®, a business-oriented BASIC language from Digital Research.™ **F.** The Osborne includes CP/M®, the industry-standard control program from Digital Research. The list of software packages which will run with CP/M is considerable. IBM offers CP/M 86 (a version of CP/M) at extra cost. There are optional hardware systems which allow the Apple II to run CP/M; the Apple II control program is highly comparable to CP/M. **G.** The Osborne includes WORDSTAR® word processing with MAILMERGE®—products of MicroPro™ International. **H.** The Osborne includes SUPERCALC™, the electronic spreadsheet system from Sorcim Corporation. **I.** Exact price comparisons cannot be presented, because the software and hardware options chosen to create the "equivalent" of the Osborne 1 Personal Business Computer vary in price. The range indicated was computed using price lists from IBM and Apple. Documentation of the computations are available on request from Osborne Computer Corporation. Trademarks: OSBORNE 1: Osborne Computer Corporation; SUPERCALC: Sorcim Corporation; Digital Research: Digital Research, Inc.; Registered Trademarks: WORDSTAR, MAILMERGE: MicroPro International Corporation of San Rafael, CA; MBASIC: Microsoft; CBASIC, CP/M: Digital Research, Inc.; IBM: IBM Corporation; Apple, Apple II: Apple Computer Corporation.

The Mercedes-Benz 240 D Sedan: the most pedigreed four-cylinder automobile in the world.

This $24,000* sedan uniquely blends hard-nosed diesel efficiency and masterful over-the-road performance with quality that is the subject of legend. It is no mystery why resale value after three years has been calculated at 86 percent.

History records a powerful reason why the 240D is a four-cylinder automobile apart: its 2.4-liter overhead camshaft engine is *40 years* more refined than the most refined domestic diesel passenger car power plant. In fact, it descends directly from the engine installed in the first diesel passenger car ever produced for sale.

10 cars, 7,000,000 miles

Pedigree tells. No automobile engine may have so often documented such staggeringly high mileage as the four-cylinder Mercedes-Benz diesel.

One survey revealed that ten American-owned and -driven Mercedes-Benz four-cylinder diesels had between them rolled up well over seven *million* miles.

Perhaps less astonished than the average layman by such claims are the engineers of Mercedes-Benz. It is they who specify machining tolerances as fine as one-hundredth of a millimeter for vital 240D engine parts.

They also implanted a tiny *ball pin* in each cylinder's precombustion chamber–diffusing injected fuel, aiming to make this engine as smooth-running as it is long-running.

One automotive tester punished the 240D under extreme conditions and concluded that "...once you're at 80 it'll stay there till you run out of fuel–which will probably be quite a few hours later." The 240D outruns every sedan in its price class in fuel efficiency. Imagine: a $24,000 sedan capable of 32 est. hwy. and 27 EPA est. mpg.** A happy economy.

Allergic to mediocrity

The 240D's pedigree extends beyond that jewel of an engine.

The 51 years of design refinement invested in its fully independent suspension system can be *felt* in limpetlike roadholding and a supple ride. The 240D is neither a galumphing luxury car nor a testy "sports sedan"–simply one of the best-mannered five-passenger automobiles extant.

There doesn't seem to be a commonplace engineering idea in the car. Even its power steering system is a technological feat: it works *progressively* to feed in extra steering force only when you need it, as in parking. Crucial steering "feel" never deserts you.

Steering to brakes to suspension, every system in the 240D is harmonized with every other. Such superb tune and balance do not just happen. The engineers spent *seven years* developing the 240D.

Quality is standard

The real glory of the 240D's beautifully padded, carpeted, wood-trimmed interior is the seats. Your body may never have been so well cared for in an automobile before. The time to judge them is not in the showroom but after a nonstop eight-hour drive.

You can choose a four-speed manual 240D or a version fitted with a four-speed, torque-converter *automatic* transmission that can also be shifted by hand. Air conditioning, electronic cruise control, and a central vacuum locking system head a long list of built-in amenities.

The 240D epitomizes what so many car makers today are scrambling to achieve: quality of fit, quality of finish–quality, period. It is so solidly built that there are 4,786 individual welding points in the body shell. It is so well finished that between the body metal and surface paint are *six* layers of protection.

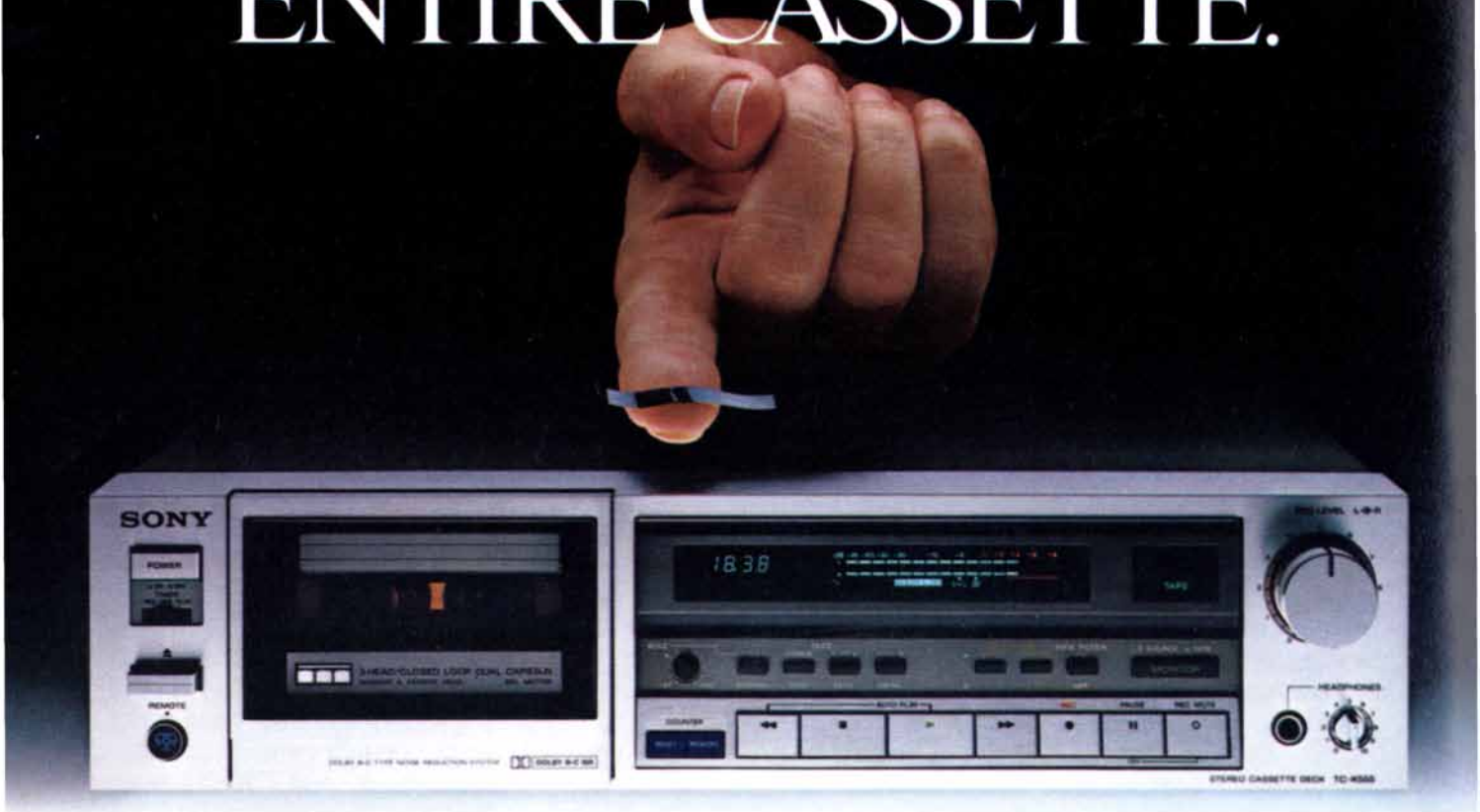
The 240D also epitomizes the rational Mercedes-Benz order of priorities. Paramount is safety. The car incorporates *more than 120* individual safety features. Many could be omitted and few buyers might be the wiser. But then, the 240D would not deserve its pedigree as a Mercedes-Benz.

*Suggested retail price p.o.e. New York. West Coast prices slightly higher.

**Use EPA estimated mpg numbers for comparison. Actual mileage may vary, depending on speed, trip length, and weather. Actual highway mileage probably lower.

Engineered like no other car in the world

SONY CAPTURES MORE MUSIC ON THIS MUCH TAPE THAN MOST RECORDERS DO ON AN ENTIRE CASSETTE.



As incredible as this statement might sound, it's absolutely credible.

Most cassette recordings just aren't faithful to the way the music was originally intended. Simply because they're made on two-head machines. And two-head machines compromise both the dynamic range and frequency response of the music. So you can listen to an entire cassette and never hear the most brilliant highs of a flute. Or the deep, rich, resonant lows of a bass.

The new Sony TC-K 555 is without compromise.

Because there are three separate heads—one to erase, one to record and one to play back—each is optimized for its own specific function. So you hear the highest highs and the lowest lows. Or, to put it another way—you hear more music.

Now, others offer three-head decks, but only Sony offers a unique, Independent Suspension system. This remarkable system allows for incredible precision and consistency in head alignment and prevents the significant high-frequency loss caused by the alignment errors so prevalent in other systems.

And the K 555 also includes closed-loop dual capstan drive for superior tape tension and reduced modulation noise. This results in not only crisper, cleaner, more precise sound, but exposes more tape to the heads. And the more tape exposed to the heads, the more music exposed to your ears.

The head design is equally unique. It's a combination of Sendust and Ferrite for maximum performance no matter what tape type.

And of course, three heads also give you the added benefit of instantaneous off-the-tape monitoring. It's one more assurance that the quality of the recorded music will be faithful to the quality of the original.

Other outstanding features: everything from Dolby's newest, most advanced noise reduction system, Dolby C,* to a state-of-the-art linear time counter which measures elapsed tape time in minutes and seconds, instead of inches.

So if sound is as important to you as it is to Sony, insist on the cassette deck that captures more music on an inch of tape than most decks can on a mile—the K 555 from Sony.

FEATURES AND SPECIFICATIONS: Linear Counter of actual elapsed time. 16-segment Peak-Program Meters with maximum recording level indication. Optional RM-50 remote control, RM-80 wireless remote, RM-65 synchronizer. S/N ratio 61dB (type III tape, Dolby off). Wow & flutter 0.04% (WRMS). Frequency response 25Hz–18kHz ±3dB (metal tape). *Dolby is a registered trademark of Dolby Laboratories. Sony Corp. of America, Sony Drive, Park Ridge, NJ 07656. © 1982 Sony Corp. of America. Sony is a registered trademark of the Sony Corp.

SONY. The one and only.

BOOKS

The sky of the year, natural fires, cacti, city-states and mitochondria

by Philip Morrison

1983 GRAPHIC TIMETABLE OF THE HEAVENS, The Maryland Academy of Sciences and Scientia, Inc. Scientia, Inc., 1815 Landrake Road, Baltimore, Md. 21204 ($6.50). The night sky spreads before the eye in just two dimensions, giving only hints of depth. Those who would guide us to look the right way generally map that vault on a flat page and sample the flow of the third dimension—time—by offering a sequence of maps. The distortions of flattening projections and the use of dots for the luminaries, whether black on white or the somewhat more natural reversal to white symbols on a night black ground, add a measure of abstraction that taxes the rank beginner, although it can soon be mastered.

This colorful wall poster is by comparison highly abstract. It spreads the night sky for the entire year across one single glossy surface the size of four magazine pages. The sky is shrunk effectively to one dimension, along the zodiacal arc where the action is: the path of sun, moon and planets. Down the central axis of the page run the midnights, with each point down the page representing one day later among the 365 days of the year. Plotted across the page, the hours of each night roll by. An hourglasslike profile marks out the times of sunset and sunrise, the long nights at top and bottom, the shorter summer nights in midyear.

Sinuously winding to and fro across sunset and sunrise frontiers, two curves, one orange and one green, display the rising and setting of Mercury and of Venus on each day of the year; two other paths in color, only subtly curved away from simple straight lines, show the outer planets. Mars wends its more ambiguous way, one year as regular as a superior planet, the next year curving strongly as the planet passes behind the sun. A few stars are straight slanting lines. This careful geocentric representation, timed by the sun underfoot, presents the visible appearances in an abstract pattern as epicyclic as Ptolemy's. Not for the entirely uninitiated, the poster is both handsome and practical for those who know the way to their seats in the celestial theater but want a careful program.

The curves are ingeniously marked and repeated to allow easy interpolation for a wide range of latitudes, although the plot is direct to an accuracy of one minute of time (with refraction and parallax corrections) for a location in Illinois. Even finicky users anywhere from Los Angeles to Toronto will have no trouble. Moon phases and the times of moonrise and moonset, with planetary and solar standstills and conjunctions, are also to be read off the curve-filled sheet.

After the group performance of 1982, the quieter year ahead offers no spectacles to adorn the stately procession. The eclipse seasons are well marked; two lunar eclipses this year are "easy if not somewhat bland" in these parts, neither of them total. (The total solar eclipse of next midsummer sends the cognoscenti to study maps of Java.) Bright Venus enters conjunction with Mars twice in autumn morning skies. Shy Mercury should be well placed for view in the late-April evening twilight.

This reliable and calligraphic chart is no new conception; it has been widely admired for a generation. Only lately, however, has it found a format so strong and colorful.

FIRE IN AMERICA: A CULTURAL HISTORY OF WILDLAND AND RURAL FIRE, by Stephen J. Pyne. Princeton University Press ($35). Fire across the land is not as ancient as ocean shore or as oxygen in the atmosphere, but it does go back to the time when plants first learned well the trick of living on dry land, a span a hundred times our hominid one. Thus fire in the New World did not need to wait for the coming of the first hunters but burned freely wherever it was ignited by high-current lightning discharge in a dry season. Down to our own day lightning is a chief predator of trees and grass; perhaps a tenth of all wildland fires are started by it.

Fire is a species with a defined range; the Tropics are too wet to harbor much of it and icy lands lack fuel to feed it. In the broad Temperate Zone it is at home, with good growth piling up many seasons' fuel on forest floor and across rolling prairie, until drought might make the fiery release of that stored energy fully self-propagating. The venerable sequoia is adapted to withstand fire by its thick asbestoslike bark; longevity demands such preparation. An oxygenated world heaped with dried cellulose is patently metastable. Flame is to be seen as a rapid mode of inevitable organic decay, an essential if stochastic part of catabolism.

Nature's fires in North America were multiplied by the hunters from Asia. Our Grandfather Fire served the people from the eastern woodlands to the lands of the Paiute and the Modoc. Broadcast fire in Indian hands was a weapon, a purifier, a key technique of hunting and a sovereign recycler, able to maintain both the light yields of the gatherers of nut and acorn and the higher yields of the woodland growers of maize. The grasslands that dominate the landscape of the U.S. are in large part the artifact of millenniums of purposeful firing by the Indian population. The forest primeval is in the main a literary device; it was widely cleared away before ever Europeans arrived with their iron: ax, gun and will.

Fire is easily tolerable to a society of nomads, people already on the move. Settlers in log cabins or in canyon houses above Malibu fear it; industrial societies see accessible timber as potential wealth, yearn for wilderness and adventure and are able to muster mass human energy and powerful technical weapons against wildfire. Above all, a countryside long cleared by ax, plow and graze calls for reforestation as successful farmlands shift.

The great barbecue of the Gilded Age had profligately consumed the nation's resources. It was in the U.S. Geological Survey that the idea of natural-resource management first crystallized a century ago. John Wesley Powell's concern was water for the dry West. Powell himself saw virtue in the ancient uses of fire: "The best thing to do for the Rocky Mountain forests was to burn them down." After about 1910, however, the old ways came to appear all but treasonous; the heroism of the firefighter joined with the romantic appeal of the wilderness to support the total exclusion of wildfire by moral sanctions beyond any mere prudent husbandry. The Chief Forester over the decades became more and more of a fire chief. From Theodore Roosevelt's day to 1930 the Forest Service fought its fires at the frontier, in economic defense of transitional values. For the next 40 years it supported a policy of strong control: "Each succeeding day will be planned... with an aim of obtaining control before ten o'clock of the next morning." Since 1970 the service has come once again to view fire by prescription as an admissible part of the processes that rule every forest.

The young Iowa historian who writes

this book is as concerned with the ebb and flow of the metaphors behind policy as he is with the documented infighting of a passel of Federal bureaus. One brilliant chapter treats of fire and warfare. Fire as weapon is very old, but as cultures grew over the land the indiscriminate nature of broadcast fire spoke against resorting to it. Open fire was increasingly confined to unequal conflicts between indigenous populations resisting enclosure or relocation and their better-armed rulers and occupiers. Even the descendants of Europeans dwelling in the piny woods of the U.S. Southeast used forest fires to defend traditional hunting grounds and pasture against the newer industrial forestry.

In war between industrial nations open fire became rare. "The images of... war are those of mud, not of fire." World War II rekindled airborne offensive wildfire; "an old word was revived to describe a new weapon: firestorm." In the years from 1950 to 1970 the military issue began to remold the U.S. Forest Service. Smokey the Bear had evolved directly out of a wartime advertising campaign based on Bambi, when the rights to the Disney design became too expensive. Aircraft, smoke jumpers, helicopters and Army crews entered the fire line. Financial and research links to war preparedness grew strong. Twenty years ago the Forest Service research laboratories examined controlling the "greatest fire on earth," the consequence of two imagined 50-megaton explosions over Portland and Seattle.

By the 1970's a new shift had come. Military research was terminated; urban fires were to be studied elsewhere than in Forest Service laboratories; even the received ecological theory changed from a determinist view of green forest, the natural crown of a sequence of stages, to a probabilistic cycling of subtle rhythms that might in part be powered by fire. Fire was again natural, and prescribed fire in the forest was once again imperative. Now it is the exclusion of fire that needs justification; it appears as a naive act of transient intervention. Has this periodic shift of concerns ended? It might be noted that the summer of 1982 saw a serious study of global thermonuclear war (in the Swedish journal *Ambio*) concluding that the most threatening of worldwide consequences would be particulates from the mass fires, engendering a hemispheric darkness at noon for at least one growing season, hence a terrible famine.

Professor Pyne writes with wit and profundity. A reader must prepare for a sometimes disturbing juxtaposition of original and even universal perceptions with the tendentious doings of Federal officeholders of less than great distinction. Pyne himself was crew foreman with the North Rim Longshots amid burning pines along the Grand Canyon for many seasons; his book did not arise out of a merely intellectual inquiry. The detailed treatment of nine geographic regions as case histories enriches the volume beyond a scale that can be covered in a review.

Our symbiotic pact with fire is ancient indeed, pervasive in language and myth. Surely it goes back to the old artificers, *Homo erectus*. For a long time fire was elemental, metaphysical, a principle of explanation. But as the poet-wit John Donne saw in the Galilean years, "The Element of fire is quite put out." It became only one gas-phase reaction, a molecular mechanism in grate and cylinder. To take its place there has arisen a more abstract new element, energy, the grand generalization of fire's power. Only in its scale is wildfire still elemental. The great burns transcend all civil releases of energy; they rage at the yield of megatons per day, like earthquakes. The most extensive fires all but enter on the gigaton domain, the potential of the stocked warheads worldwide. Philosopher–fire fighter that he is, Pyne is an optimist: "If our ancestors, with little more than river cobbles and chipped flint, could seize fire, surely modern mankind can cope with its own versions of the forbidden flame."

THE CACTI OF THE UNITED STATES AND CANADA, by Lyman Benson. With line drawings by Lucretia Breazeale Hamilton. Stanford University Press ($85). "At the point where we left the Gila, there stands a Cereus six feet in circumference and so high, I could not reach half way to the top of it with the point of my sabre by many feet." So wrote the young officer who led a reconnaissance along the new border from Fort Leavenworth to San Diego in 1846, the first account of the giant saguaro cactus to reach print. That very grove of giants is still there; how they look is suggested, with a certain understandable hyperbole, in the grand engraving reproduced from the full report of that trip, which presented to "a waiting world" its first view of the Arizona desert.

Professor Benson has produced a splendid volume to carry to the public the results of his 48 years of study of that remarkable plant family, the New World succulents. It is an exemplary volume, accessible, rich in scientific and historical context, profusely and thoughtfully illustrated, short on jargon and yet meticulous with the documentation obligatory for the taxonomic botanist. The treatment of record and meaning is so fastidious as to inspire any veteran of the Philadelphia bar.

There are in effect three books within the handsome 1,000-page volume. The first is a general account of the cacti, their structure, physiology and ecology along with many taxonomic issues. One chapter, whose aim is to clarify the geo-

graphic distribution and associations of the cacti (there are native prickly pears on Cape Cod, and cacti are found wild in 45 U.S. states and five provinces of Canada), admirably describes and illustrates floristic associations from the foggy Aleutians to the Florida Keys. The bulk of the text is given over to a modern and expert flora of the cacti, each genus keyed, described and mapped in detail with frequent careful drawings and photographs showing details of growth and structure; 48 color plates display the bizarre forms and lovely flowers typical of these forms.

Some 60 pages in fine print constitute a third book of documentation for readers who might enjoy the minutiae of taxonomic recognition. Professor Benson is particularly careful to describe the evidence, from the correct use and preparation of herbarium sheets to the character of devoted and extensive botanizing in the field. One map dots out the counties this professional has visited in his intricate life path among the cacti; another page remarks that one tiny but elegant species has cost more man-days of search than any other, save a still smaller one. "Finding the first one required many days of searching bent over or on hands and knees." The saguaro grove along the Gila was visited by the author in the course of his own journey to all the 1846 campsites to gather specimens of the cacti first described and sketched, but not collected, on that survey 120 years earlier.

Students of the family "have been preoccupied with discovery of new entities, real or otherwise." Professor Benson chooses rather to follow the intellectual lead of the St. Louis physician George Engelmann, whose unmatched genius for organizing botanical information illumines his pioneer work on cacti. An excess of discrimination mars most subsequent classification of cacti, with their diversity magnified by the industrious gardeners. This book treats of cacti in nature, part of the wild flora of the region studied.

The subject is too detailed to summarize, but a few results may serve as a sample for the general reader. Cacti are wax-coated plants, stems almost without leaves, low in surface for their volume (tinier forms retract entirely into the ground on hot days), spiny to safeguard their water stores from animal attack. Those spines are in fact the vestige of leaves; recent experiments on one form of prickly pear showed that the same buds could develop into either leaves or spines. The structures treated with one familiar plant hormone turned into normal leaves, those treated with another became spines.

Carbon dioxide is unable to diffuse through the waxy cuticle of the cactus; it is admitted by open stomata, in the cool darkness when the loss of water through

the opening is low. At night photosynthesis cannot fix the essential gas, and so it is stored at low temperature bonded into a simple organic acid, ready to be utilized by daylight. Thus cactus juices are a sour morning refreshment; that barrel cactus may be juicy, but its water is acceptable only in an emergency. The saguaro flesh has an unpleasant flavor, but the copious fruit of the right prickly pear is a major food in the markets of highland Mexico. Burbank's spineless cactus wonder turned out to be one more public-relations hype, less his fault than that of the swindlers who exploited his name in hilarious controversy during the years before World War I. The arresting purples that sometimes spread over the spiny green pads are recorded in one color photograph. They are unique to desert scenes; certain pigments formed in cacti (and in one related order) are chemically distinct from the anthocyanins of all other flowering plants.

The windowsill horticulturist is not the intended reader of this book; its cost and bulk are disproportionate to that interest. More important, the treatise is carefully limited in geographic range, aimed not at the collector's array of succulents but at the cacti there on the American land. For those who seek out waste places, and for apropos libraries of all degrees, it is a model of devotion, learning and thoughtful execution.

One final warning emerges: The cacti are the most threatened of all plant families. Collection and sale of cactus seeds—always in great surplus—are to be encouraged, but living plants ought not to be removed from the field either for sale or for the home. New, small protected areas are chosen for special reasons; a "single fantastic soil outcrop" supports one rare species over a range 20 miles long and only a quarter of a mile wide. On Indian land, that plant has owners who are both aware and protective. The danger list, however, is some 70 species long.

THE CITY-STATE IN FIVE CULTURES, edited by Robert Griffith and Carol G. Thomas. ABC-Clio, Inc., Rivera Campus, Santa Barbara, Calif. 93103 ($22.50). Most of these cities evoke images at once: Uruk, Sparta, Giotto's Florence, Kepler's Weil der Stadt. Fewer readers can quickly call up the massive walls of Kano or Zaria, the Hausaland examples of this genus of social organization. In an unusually readable volume five scholar-specialists from the shores of Puget Sound present overviews of the history and nature of comparable city-states. The oldest of all cities were the dozen temple towns of Sumer 5,000 years ago; the latest city taken up here lasted until 1812, when the Sword of Truth, 'Uthman dan Fodio, set his own sons over the Hausa towns as viceroys of a new Fulani empire across the Nigerian grasslands.

The aim is compelling; it is to crystallize by detailed comparison a single entity out of these diverse chronicles. In Sumer the people lived by painstaking irrigation of the barley fields, in Florence by the export of wool cloth intricately finished, in the medieval cities of Germany and the Swiss cantons by the artisanship of the guildsmen and the capital stores of bankers and merchants. The wide Hausa town walls protectively encircled fertile gardens, unfailing water holes and even ironstone hills for the smiths. No single economy unites these narratives. Our taxonomists claim no overarching or final system for the perceived unity; they present modestly enough an inductive view.

A city-state first of all is mononuclear. A single urbanized core is its heart, bounded by some membrane: a wall or a moat. That nucleus must be nourished, and so some hinterland is organic to the city, which always seeks economic autonomy within its own small region, more often finding its livelihood only by forming the node of a vigorous system of trade and finance. No city-state is a nation; usually the cities that are its allies and rivals are bearers of the same language and culture. A world outside seems given over to barbarians. The Hausa legend accounted for the kingdoms that surrounded their lands as states founded by the descendants of their own originator, Abuyazidu, although only through the sons of his concubine: the bastard Seven. The final decisive item of the key is independence: the city-state is the single sovereign of its citizens, who come to share a certain kinship, if only a fictitious one.

The sample is not small. There were more than 1,500 Greek *poleis,* and dozens or scores of city-states in all the other cultures considered. They share scale at least in order of magnitude: in space, say a few square miles within the walls; in head count, 10,000 urban families, not including the peasant folk outside; in time, half a thousand years of proud independence. None of the city-states persist. They were all caught in tension, the need to expand control as local resources declined. Or a new perception of potential arose, against the imperative of a consolidated, thriving and devoted core. Frank slavery, selfish rule by a hundred noble families (the Florentines tore down the arrogant towers of the nobility when once Florence looked like San Gimignano today) and the loss of well-worn lucrative routes of trade all stressed the city-states. Their weakness, like their strength, was in their scale. They were poorest at adapting to sweeping structured change. A wider unity finally swallowed them all.

City-states seem to represent one stable form of social grouping when central authority is lacking. Striking, convergent in nature along a spectrum of culture, history and economics, they were neither inevitable stages nor enduring structures as human polity evolved. They were conspicuous eddies in the turbulent flow.

MITOCHONDRIA, by Alexander Tzagoloff. Plenum Press ($42.50). A century ago cytologists, tireless microscopists adept at the fixing and staining of tissue, saw small, threadlike granules (*mitochondria* in the Greek synthesis) scattered by the score within the cells of every plant and animal. They resembled bacteria in size and form. In the first years after World War II it was demonstrated that these intricate organelles held a special importance. The classical biochemistry of solutions, armed with the irresistible Waring blender and an arsenal of chemical and photochemical weapons, had worked out step by delicate step the long chains of metabolism; the crown of them all is the Krebs cycle. There eight enzymes cooperate in a wheel of linked reactions. The first product of organic oxidation is eventually burned to water and carbon dioxide, at last releasing to the air-breathing cells of higher life nearly 20 times as much energy from each glucose molecule as anaerobic fermentation can yield.

In 1948 it became clear that the entire system was housed in the mitochondria. The energy released by the Krebs cycle, however, must somehow be stored for biochemical work. That sequence, so to speak a cog meshed with the Krebs cycle, yields the universal cellular fuel, ATP, by a series of oxidative pathways that overall store chemically in the unstable phosphate all the energy won from glucose. The process lay outside the grasp of the classical methods; its key enzymes are insoluble, and cell-free extracts are still poor producers of ATP. Not mere protein solutions but cellular ultrastructure holds the intricate mechanism of ATP synthesis.

This crisp text by a Columbia participant of the discipline makes its point at once to the eye. Revealing electron micrographs of mitochondria go back 30 years, to the pioneering of Albert Claude and George E. Palade. The microscopists of our electronic times have their own cunning means of fixation and staining that resolve structure at a fineness a couple of hundred times smaller than the optics of an earlier generation could. Mitochondria repay such close scrutiny; indeed, it has turned out that the bag within a bag that is a granule does hold all the enzymes of the Krebs process in its innermost compartment. The outer membrane wall is more or less rod-shaped, but the inner one, seen in writhing motion in rare cinema shots of living examples, is complex, deeply reentrant, changing.

Thermal Conduction Module

IBM

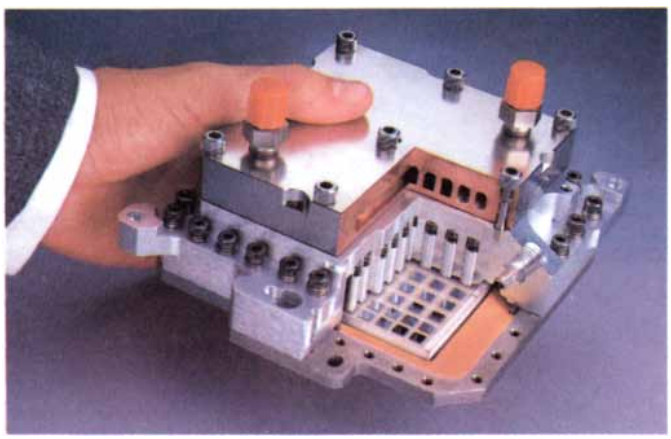
To: Jim
From: Bill
Subject: IBM Technology

I've been reviewing some of our past and present technological achievements, and it occurred to me that the scientific, engineering and academic communities might like to know more about them. Will you select one from the following list as the first topic? Thanks.

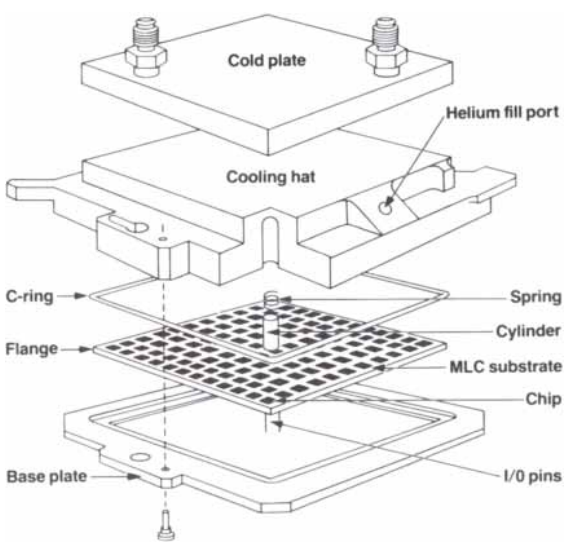
Vacuum tube digital multiplier
IBM 603/604 calculators
Selective Sequence Electronic Calculator (SSEC)
Tape drive vacuum column
Naval Ordnance Research Calculator (NORC)
Input/output channel
IBM 608 transistor calculator
FORTRAN
RAMAC and disks
First automated transistor production
Chain and train printers
Input/Output Control System (IOCS)
STRETCH computer
"Selectric" typewriter
SABRE airline reservation system
Removable disk pack
Virtual machine concept
Hypertape

System/360 compatible family
Operating System/360
Solid Logic Technology
System/360 Model 67/Time-Sharing System
One-transistor memory cell
Cache memory
Relational data base
First all-monolithic main memory
Thin film recording head
Floppy disk
Tape group code recording
Systems Network Architecture
Federal cryptographic standard
Laser/electrophotographic printer
First 64K-bit chip mass production
First E-beam direct-write chip production
Thermal Conduction Module
288K-bit memory chip
Robotic control language

Bill-
Good idea! Our TCM is a state-of-the-art technology. Let's start there.
Jim



Cutaway of TCM (15xl5x6 cm) which contains logic and buffer memory equivalent to an IBM System/370 Model 148. Its new packaging, cooling and assembly technology provides IBM's largest computers with the greatest circuit density yet reported.



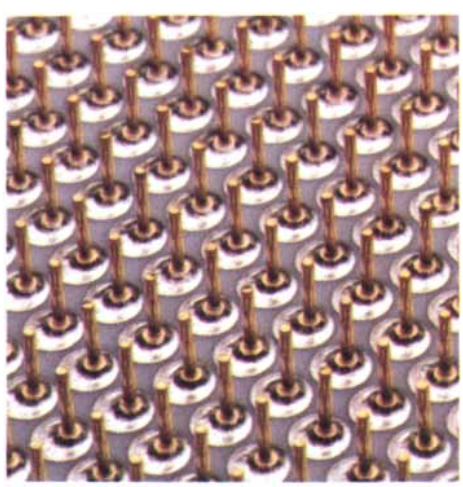
Exploded view of TCM assembly.

The time it takes a signal to travel between circuits is key to a computer's performance. The shorter the distance the signal has to travel, the faster the computer can operate. Shortening the distance requires improving the circuit density of the packaging and interconnections, as well as placing the circuits closer together on the logic and memory chips.

Circuit packaging in IBM's most powerful computers—the 3081, 3083 and 3084—is the densest yet reported in the industry. This has been made possible by our Thermal Conduction Module (TCM) and the way it is combined with a high-performance circuit board.

A TCM holds up to 133 logic and memory array chips on a 90-mm substrate, consisting of 33 layered ceramic sheets. More than 120 meters of wiring within these layers connect up to 45,000 circuits within the TCM. And 1,800 pins are brazed to the substrate bottom to plug the module into a circuit board.

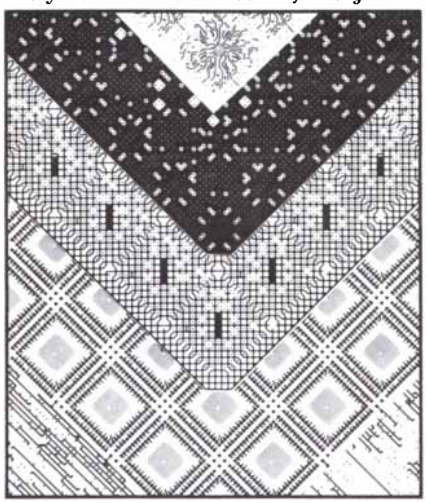
All of this requires computer-aided



Detail of input/output pins brazed to back of multilayer ceramic (MLC) substrate which plug TCM to next level of packaging.

Thermal Conduction Module

Magnified portion of ceramic sheets (0.2 or 0.28 mm thick, unfired), the basic building blocks of the 33-layer MLC. A typical substrate contains 350,000 vias for layer-to-layer connections and 130 m of wiring.



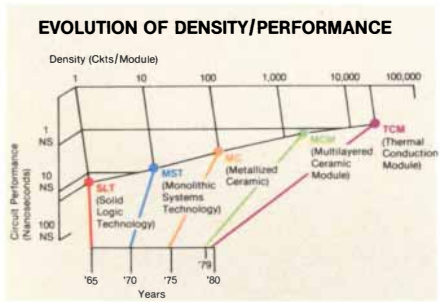
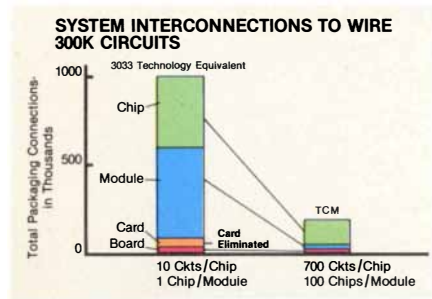


Illustration of TCM's dramatic benefits. Average performance improved more than 20x over 1965 level. Average circuits/module increased from one in SLT to 30,000 in TCM.



Effect of TCM technology on hypothetical 300K circuit system. Circuits/chip and chips/module increased with dramatic reductions of system interconnections and elimination of card-level package. Result is lowered cost, improved performance and reliability.

design techniques and precision manufacturing operations unique to IBM.

Thus "packaged," the chips require a new cooling mechanism to remove the heat created by their unprecedented density. Within the hermetically-sealed TCM, spring-loaded cylinders contact the chips and conduct heat away from them. Helium surrounds the chips and cylinders, providing an inert atmosphere. Its heat-conducting efficiency is approximately six times that of air, and it carries the heat to an aluminum alloy "hat" atop the TCM. The heat is then conducted to a cold plate through which water circulates at 24°C.

The circuit density of the TCM required IBM engineers to develop a high density, high-performance printed-circuit board on which to mount the modules and to provide power and signal paths. This board is the densest yet reported for large-scale manufacturing.

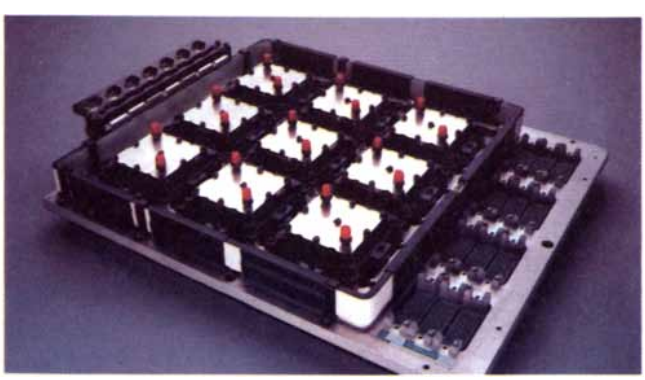
This 600 x 700 mm printed-circuit board contains 20 layers of circuitry, and includes more than one kilometer of wiring. It can accommodate up to six or nine TCMs depending on the board configuration.

Together, the TCM and the circuit board eliminate an entire level of packaging previously required in computers. This combination contains all the functions normally attributed to modules, cards, boards and interconnecting cables in prior technologies.

The circuit density made possible by this new packaging, cooling and assembly technology allows a single TCM to contain as much logic and buffer memory as an IBM System/370 Model 148. Today's large-scale IBM 3081 carries up to 26 TCMs on four circuit boards which contain its complete logic—nearly 800,000 circuits—and displace only four cubic feet.

IBM scientists and engineers worldwide have contributed technologies to the TCM. These technologies are all part of our continuing commitment to research and development, funded with more than $8 billion over the past seven years.

IBM®



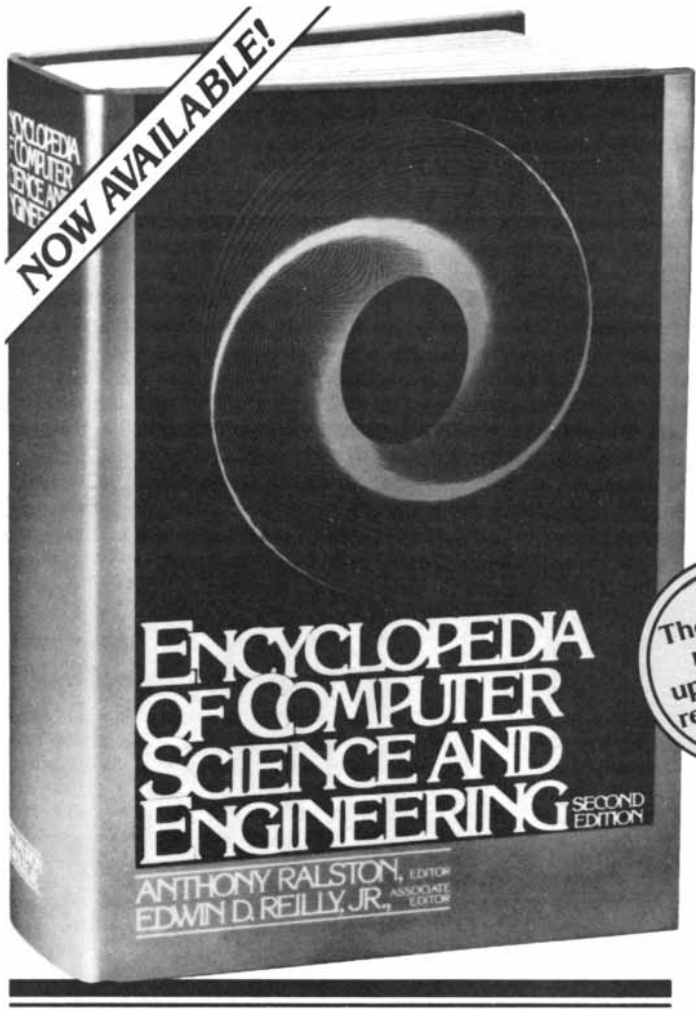
This 3081 printed-circuit board carries nine TCM's and is, in itself, a major advance in circuit packaging.

A recent issue of the IBM Journal of Research and Development is devoted to the IBM 3081. For a free copy, write: Director of Technical Publications, IBM Corporation, Dept. 100, 44 S. Broadway, White Plains, N.Y. 10601.

The game of sonic disassembly and sequential, chemically controlled reassembly of these membranes and the structures they bear has been enthusiastically played for 20 years. There are still, however, missing pieces in the puzzle of the site and nature of ATP production. The tiny protein globules on short rooted stalks that line the inner membrane offer one example. It seems likely that the three parts of this structure—root, stalk and head—constitute the final synthesizing enzyme, although hard proof is lacking. If the first seven chapters outline the classical biochemistry and the later particulate study of the chemical function of the organelles, chapter 9 sounds a strong new theme: the nature of the membranes themselves. These are no inert walls. It is a surprise to see how easily a single active enzyme, the important cytochrome oxidase, is quite ready to self-assemble in vitro into little membrane bags on its own, given a good dollop of phospholipids. Structure and function flow as one within the mitochondria.

The two final chapters play very modern music indeed. The population of mitochondria in a cell divide and fuse to meet the cell's energy needs by processes little understood and only sketchily reviewed here. These extranuclear granules hold a DNA instructional legacy all their own. Here is a beautiful micrograph of the DNA circlet of a human mitochondrion, a simple tape loop a few microns long, compared with the meter or two of tape intricately wound and unwound within each cell nucleus. That singular piece of the human genetic fabric is now being mapped both genetically and for its nucleotide sequences. Yeast and mammalian mitochondria differ considerably in detail, yet both are remarkable in that they rely on an unexpected three-nucleotide coding symbol for translation into one of the essential amino acids. That essential auxiliary tape of every human cell does not share the substitution code for protein synthesis otherwise universal in life.

These final chapters are rich in the results and methods of today's molecular biology of information; the book carries the story up to late 1980. The text ends with a too terse account of the conjectural origins of the mitochondria, which give many signs of alien origin, quasi-bacterial live-in servants within the larger cell. The experienced author is not inclined to speculation, but mitochondrial studies clearly tax the triumphant techniques of molecular biology today. It is satisfying if hard going to follow these investigators well beyond the easier generalizations in the introductory textbooks. The text is attractive and clear; somehow, probably by way of some mutant editing disk, "energy conversion" has throughout received the reading "energy conservation."



TAKE THE *NEW* ENCYCLOPEDIA OF COMPUTER SCIENCE AND ENGINEERING (AN $87.50 VALUE) FOR $2.95!

Thoroughly revised, updated and reorganized.

When you join the Library of Computer and Information Sciences. You simply agree to buy 3 more books — at handsome discounts — within the next 12 months.

Just published, the ENCYCLOPEDIA OF COMPUTER SCIENCE AND ENGINEERING is the most complete computer information resource available. Indispensable to consultants, business people, data processing professionals, and enthusiasts, it is a veritable data base of information on:

- Hardware
- Software
- Programming languages
- Artificial intelligence
- Operating systems
- Computer applications
- Personal computing
- and much more!

The Encyclopedia is organized to make finding and using its wealth of information an ease. Articles are alphabetically arranged and are cross-referenced to related articles and to specific subject matter. The clear and expanded appendices include abbreviations, acronyms, special notation and terminology, as well as numerical tables, the mainstay of applied technologies. A complete 5,000-term index contains references to sub-categories, doubles as a computer science dictionary, and is an invaluable tool for locating specific information.

Praise For the First Edition:
Called "Impressive . . . comprehensive . . . well done" by *Datamation,* and ". . . a real treasure cache" by *Business Management,* the new Second Edition promises to eclipse *Computer Management's* statement on the original of "There isn't another book like it."

Send for your *free* 10-day trial look today.

The Library of Computer and Information Sciences is the oldest and largest book club especially designed for the computer professional. In the incredibly fast-moving world of data processing, where up-to-date knowledge is essential, we make it easy for you to keep totally informed on all areas of the information sciences. In addition, books are offered at discounts up to 30% off publishers' prices.

Begin enjoying the club's benefits today!

If the reply card has been removed, please write to
The Library of Computer and Information Sciences
Dept. 7-BN8, Riverside, N.J. 08075
to obtain membership information and an application.

This up-to-date Second Edition contains:

- **Over 1,670 pages of revised, expanded, and updated text**
- **550 articles on virtually every aspect of the computer sciences**
- **Written by 301 distinguished authorities (over 90 experts have been added to the list of ENCYCLOPEDIA contributors)**
- **Profusely illustrated with over 500 photos, and over 250 diagrams, graphs and charts**

4 Good Reasons to Join

1. The Finest Books. Of the hundreds of books submitted to us each year, only the very finest are selected and offered. Moreover, our books are always of equal quality to publishers' editions, *never* economy editions.
2. Big Savings. In addition to getting the ENCYCLOPEDIA OF COMPUTER SCIENCE AND ENGINEERING for $2.95 when you join, you keep saving substantially—up to 30% and occasionally even more. (For example, your total savings as a trial member—including this introductory offer—can easily be over 50%. That's like getting every other book free!).
3. Bonus Books. Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings up to 70% off the publishers' prices.
4. Convenient Service. At 3-4 week intervals (16 times per year) you will receive the Book Club News, describing the Main Selection and Alternate Selections, together with a dated reply card. If you want the Main Selection do nothing and it will be sent to you automatically. If you prefer another selection, or no book at all, simply indicate your choice on the card, and return it by the date specified. You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

Established 1845 SCIENTIFIC AMERICAN January 1983 Volume 248 Number 1

The Large-Scale Cultivation of Mammalian Cells

Novel reactors have been designed for growing in culture large quantities of the fragile, complex cells that synthesize medically important proteins such as interferon and monoclonal antibodies

by Joseph Feder and William R. Tolbert

There are molecules of great investigative, clinical and perhaps commercial value that can best be produced by growing in culture the human cells and other mammalian cells that synthesize them. The problem is that it is no simple matter to grow large quantities of mammalian cells in an artificial medium. The well-developed technology of industrial microbiology is adapted to the requirements of bacteria, yeasts and molds. Each single cell, encased in a tough cell wall, is an independent metabolic factory with fairly simple nutritional requirements; for bacteria, glucose and some simple salts will often suffice. Microorganisms grow well floating free in a liquid culture medium in tanks with a capacity of as much as 50,000 gallons, resisting damage even when they have proliferated to form a thick suspension and even when the suspension is stirred vigorously with a mechanical agitator.

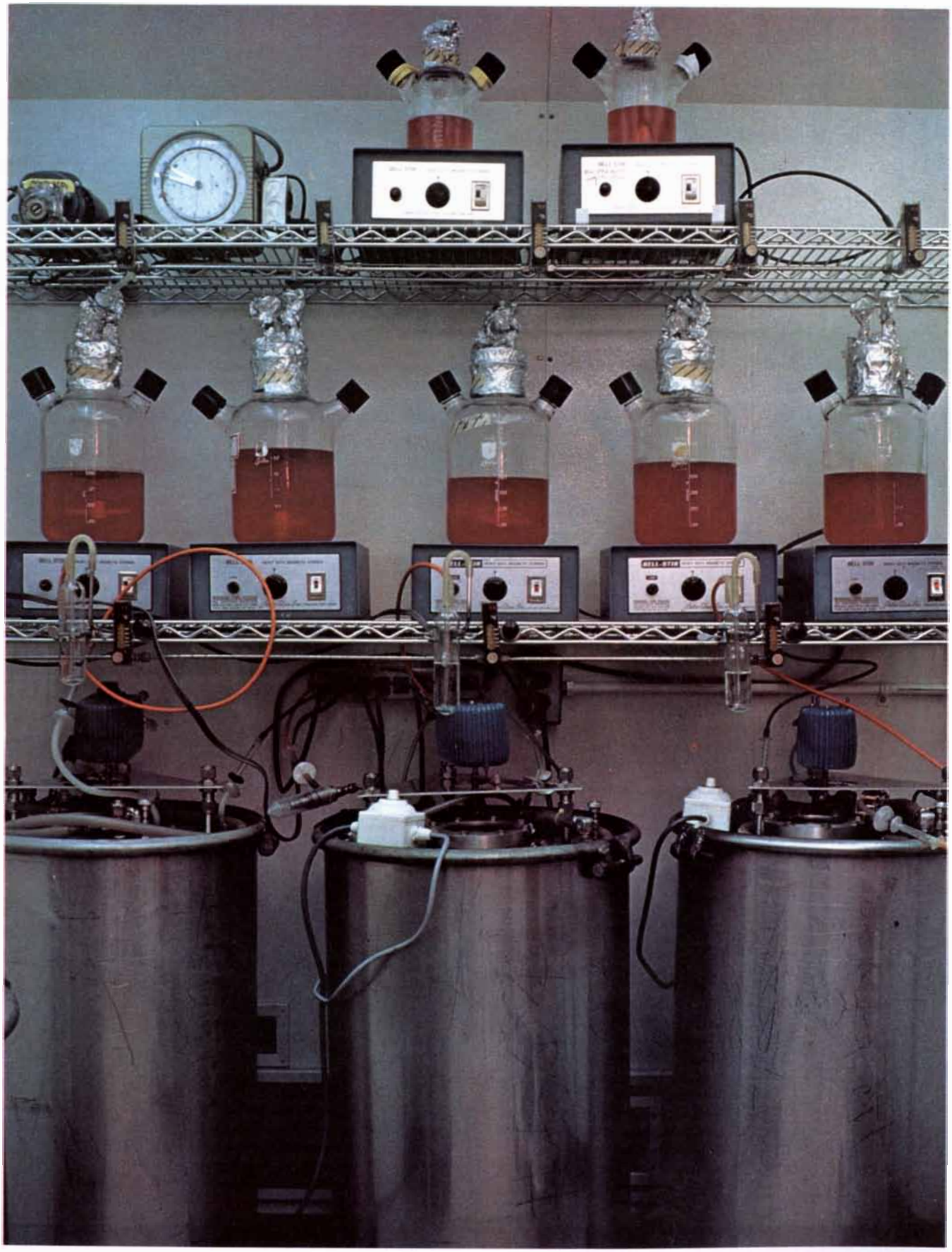
Mammalian cells are different. They are larger than most microorganisms, more fragile and more complex. The delicate plasma membrane that encloses an animal cell is not encased in a tough cell wall. The mammalian cell's nutritional requirements are more stringent than those of most microorganisms and indeed have not yet been fully defined. Rather than being a free-living organism, a mammalian cell is adapted to a specialized life as part of an organized tissue, dependent on the specialized functions of many other cells and on a circulatory system that ensures a precisely adjusted and stable environment for each cell. Such a cell resists being separated from its tissue and grown in an artificial medium. Most animal cells will not grow at all in suspension; they grow only when they can attach themselves to a surface. Over the years techniques have been developed for growing mammalian cells on a small scale in the laboratory. It has proved to be much more difficult to grow them efficiently on even a moderately larger scale.

There is a pressing need for large-scale culture methods. Consider interferon. It is released by animal cells and inhibits viral infection. It was discovered in 1957, but its clinical efficacy is still not well established, mainly because it has been hard to grow large quantities of the mammalian cells that can be induced to synthesize it. By the same token, monoclonal antibodies targeted against a specific protein are manufactured by "hybridomas" formed by the fusion of an antibody-secreting cell to a malignant myeloma cell. Mouse hybridomas can be grown efficiently as tumors in laboratory animals. Human monoclonal antibodies promise to be powerful therapeutic agents, but the human hybridomas that make them need to be grown in culture.

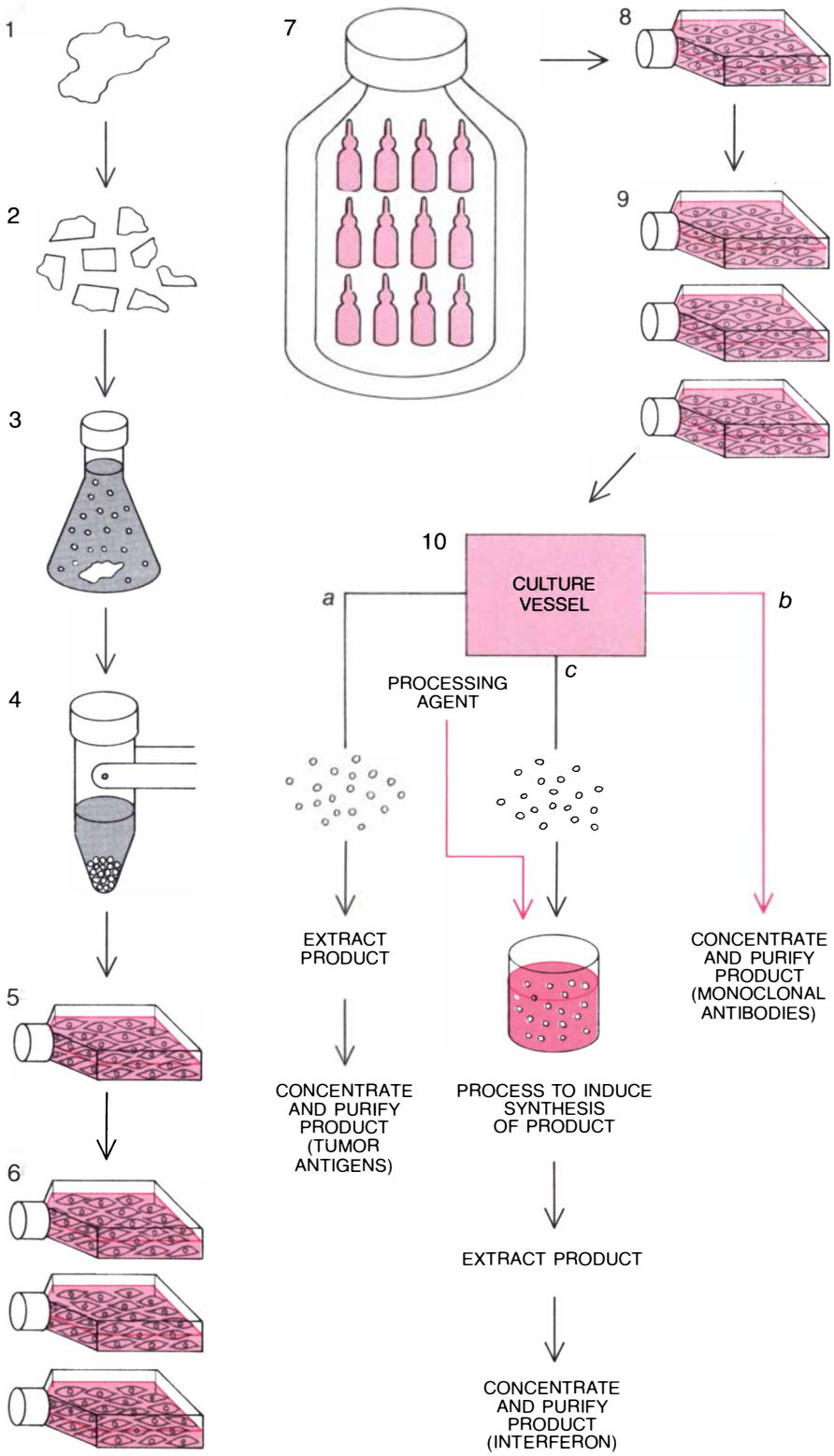
Human urokinase, an enzyme that activates plasminogen to form the blood-clot-dissolving enzyme plasmin, has obvious clinical promise. A good source of urokinase is human kidney cells, which are difficult to grow on a large scale. Angiogenic factors are cell mediators that promote the growth of blood vessels. Specific tumor angiogenic factors released from tumor cells appear to promote the vascularization and growth of solid tumors. Large amounts of these factors are needed for research that may lead to the development of drugs for regulating angiogenesis (and perhaps for controlling the growth of a cancer); large amounts may also be useful for direct administration to promote revascularization. Human cells carry cell-surface antigens: molecules that serve as markers distinguishing self from nonself and tumor cells from normal cells. Tumor antigens are the subject of much current research, and they can be made available in quantity only by large-scale cell culture.

The demand for these molecules and for many others has motivated recent efforts to develop efficient large-scale culture systems for mammalian cells. The efforts have not been discouraged by awareness of the dramatic advances in recombinant-DNA technology, which make it feasible to introduce the gene encoding a desired protein into bacteria and then induce the bacteria to synthesize the protein. Many biologically important molecules cannot be synthesized by these methods. And in the case of a substance, such as the angiogenic factor, that has not even been characterized, the only way to get enough material from which to isolate it is to grow the cells that make it.

A mammalian-cell culture begins with a mammalian tissue. The tissue is dissociated either mechanically or enzymatically, or by a combination of the two methods, to yield a mixture of single cells and small clumps of cells. The mixture is inoculated into an appropriate liquid growth medium that ordinarily includes salts, glucose, certain amino ac-



CULTURE VESSELS for growing mammalian cells in quantity are seen in the authors' laboratory at the Monsanto Company. The three cylindrical steel tanks are 100-liter reactors for growing cells that do not need to be anchored to a surface and can be grown in suspension in a culture medium. The glass vessels on the shelves are spinner bottles in which cells are grown before being inoculated into a large-scale culture vessel. Each bottle stands on a magnetic stirrer. The red color of the culture mediums results from an acid-base indicator.



MEDICALLY IMPORTANT PROTEINS can be produced from a large-scale culture of mammalian cells by the general procedure diagrammed here. A tissue (*1*) is cut into small bits (*2*), which are treated with a protein-dissolving enzyme (*3*) that releases individual cells. The cells are collected by centrifugation (*4*) and inoculated into a nutrient medium, where they proliferate until they cover the floor of the culture flask (*5*). The confluent cells are released from the floor of the vessel by an enzyme and are reinoculated into several flasks to expand the culture (*6*). A stock of the cells is stored frozen in liquid nitrogen (*7*). As required, some of the cells are thawed to start a new culture (*8*), which is expanded (*9*) to make enough cells to inoculate a large-scale culture vessel (*10*). The desired proteins can be retrieved in several different ways. The cells can be harvested and a cell product such as the surface antigens of tumor cells can be removed and purified (*a*). If the desired product is secreted by the cells, as it is in the case of monoclonal antibodies, it can be removed from the medium, which is pumped out of the reactor while the cells are retained (*b*). In some instances an extra stage is required in which the cultured cells are treated to induce the synthesis of a desired product, such as interferon (*c*).

ids (the subunits of proteins) and blood serum (usually accounting for between 5 and 20 percent of the medium).

The serum is included as a means of providing components that have not yet been identified but have been shown to be needed if the cells are to live and grow in an artificial medium. Serum is expensive, and its cost largely determines the economic feasibility of a particular culture system. Cells in culture are deprived of the elaborate immune defense system that is an integral part of the intact animal, and so antibiotics are often included to prevent infection of the culture. The *p*H, the temperature and the level of oxygen and carbon dioxide must be closely controlled. The salt content is regulated to keep the liquid medium's osmotic pressure within bounds and preserve the integrity of the fragile cell membrane.

Although most mammalian cells need to be anchored to a solid surface, cells that originate in blood or lymphatic tissue, along with most tumor cells and other transformed cells, can be adapted to growth in suspension. Such cells can be cultured in spinner bottles. These are simple glass containers in which a stirring element is suspended; the element is driven by an external rotating magnet at the base of the flask. Spinner bottles with capacities of from 25 milliliters to 15 liters are available for laboratory applications.

For anchorage-dependent cells (most normal cells) a number of laboratory vessels are appropriate, ranging from "microwell" plates and flat-bottomed Petri dishes through tissue flasks of various sizes. The most effective vessel is the roller bottle: a cylindrical vessel that is oriented with its long axis horizontal and is continuously rotated. Cells attach themselves to the inner surface of the cylinder, and the slow rotation (about one revolution per minute) exposes them alternately to the liquid medium and to air. The largest roller bottles offer a surface area of some 1,600 square centimeters.

For cells that grow in suspension large-scale reactors have been developed by modifying the fermenters in which single-celled microorganisms are grown. In the process, however, certain inefficiencies have been preserved. Most fermenter-style reactors are elaborate and expensive. For asepsis they rely on complex steam-sterilization equipment and the inclusion of antibiotics in the medium. The large-scale growth of anchorage-dependent cells presents even harder problems, which until recently have been avoided by simply multiplying the number of conventional vessels, most commonly roller bottles.

Some years ago our laboratory in the research department of the Monsanto Company undertook to devise new sys-

tems for large-scale cell culture that would meet our pressing need for a dependable source of hundreds of liters of cell suspensions each week. We decided at the outset that the systems must maintain absolute sterility without depending on antibiotics, which can mask the slow development of a low-grade infection in a culture for some time and can affect the metabolism of cultured cells in subtle ways that cannot always be predicted. We also decided on a modular approach: individual culture vessels would be limited in size and large-scale production would be attained simply by bringing more modules on line. Finally, we were determined to explore innovative designs that would address the specific growth requirements of the large, fragile mammalian cell.

A simple and inexpensive 100-liter reactor for the growth of cells in suspension met all these criteria. It was made from a 30-gallon stainless-steel drum fitted with a removable head through which all input and output attachments were made. The reactor is small enough to be sterilized in an autoclave; a double-septum port in the removable head makes it possible to take samples as often as needed without contaminating the culture. The necessary agitation is provided by a vibromixer (a device that was first demonstrated in 1965 by George E. Moore of the Roswell Park Memorial Institute). A disk with conical apertures is attached to a vertical shaft that oscillates up and down at high speed. The vibration of the disk and the passage of medium through the holes combine to agitate the medium largely in a vertical rather than a horizontal direction. An added advantage of the vibromixer is that its motion is transmitted through a flexible diaphragm, providing an absolute, static seal at the top of the vessel. Tubes introduced through the top of the vessel serve to bubble oxygen through the medium and to move air enriched with carbon dioxide over the surface of the medium.

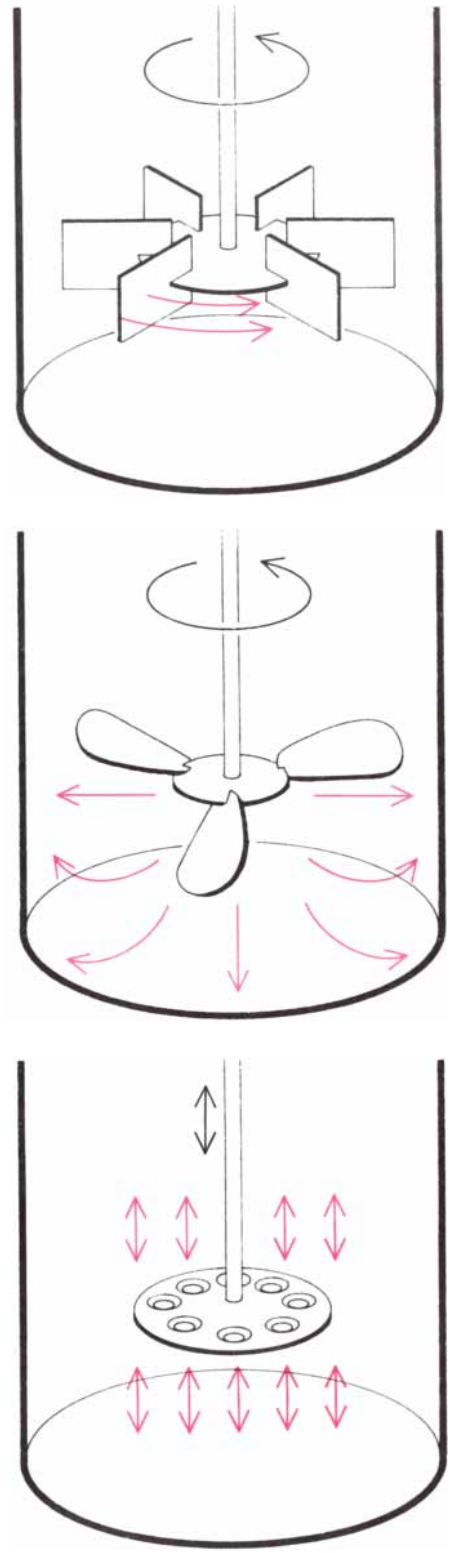
We built a number of these 100-liter vessels at about a twentieth of the cost of conventional fermenters. They are adaptable to a variety of cells that grow in suspension, and over the past few years hundreds of thousands of liters of human-cell and rodent-cell suspensions have been grown and harvested. Nevertheless, we decided against simply making larger versions of the 100-liter reactor in order to attain higher levels of production. Instead we set about developing a system that would more closely imitate the conditions under which a mammalian cell grows in nature.

A conventional cell-culture system, including our 100-liter vessel, exposes growing cells to a constantly changing environment. Cells are inoculated into a medium that is rich in just the right nutrients and is devoid of metabolic waste products; as the cells grow, the concentration of nutrients decreases and waste products accumulate. This imposes a limit on the density of the culture. In the living animal, on the other hand, homeostasis is the rule. An efficient circulatory system delivers oxygen and nutrients and removes carbon dioxide and waste products, providing every cell with a stable environment even though the cell density is between 500 and 1,000 times as high in a living tissue as it is in a typical culture medium. It would be desirable to provide cells in culture with a near-steady-state environment. A way to do that is to constantly perfuse fresh medium through the culture and constantly remove depleted medium. The trick is to renew the medium without removing the cells. In 1969 Philip Himmelfarb and Philip S. Thayer of Arthur D. Little, Inc., developed a laboratory-scale perfusion system for animal-cell suspension cultures and reported attaining much higher cell densities than could be achieved with conventional suspension systems.

We have developed a moderate-scale perfusion system with four-liter and 44-liter vessels in which cells are grown in amounts that compare with what is ordinarily grown in vessels having capacities of 100 and 1,000 liters. The perfusion is made possible by the introduction into the system of a satellite vessel containing a cylindrical porcelain filter with pores less than two micrometers (thousandths of a millimeter) in diameter. The filter retains the cells, which are pumped back into the culture vessel; some of the medium is recycled and some of it is withdrawn as fresh medium is fed into the system. To minimize clogging of the filter pores by cells or cellular debris, shear forces are established at the surface of the filter either by rotating the filter or by surrounding it with a rotating agitator. Special sensors, sampling ports and input and ouput piping make it possible to maintain the *p*H and the concentration of oxygen, carbon dioxide and nutrients at the proper level even when cells have proliferated to a very high density.

The high cell density achieved with perfusion made it necessary to design a new and gentler agitating mechanism. Four flexible sheets woven of a plastic monofilament rotate slowly (at from 8 to 20 r.p.m.) in the culture vessel, billowing like sails as the liquid medium spills from their trailing edge. By distributing the required energy of agitation over a large area the flexible-sail mechanism reduces the shear forces acting on individual cells.

To evaluate the efficiency of the perfusion system we compared the growth of rat tumor cells in it and in a conventional system and monitored the concentration of glucose and of lactic acid



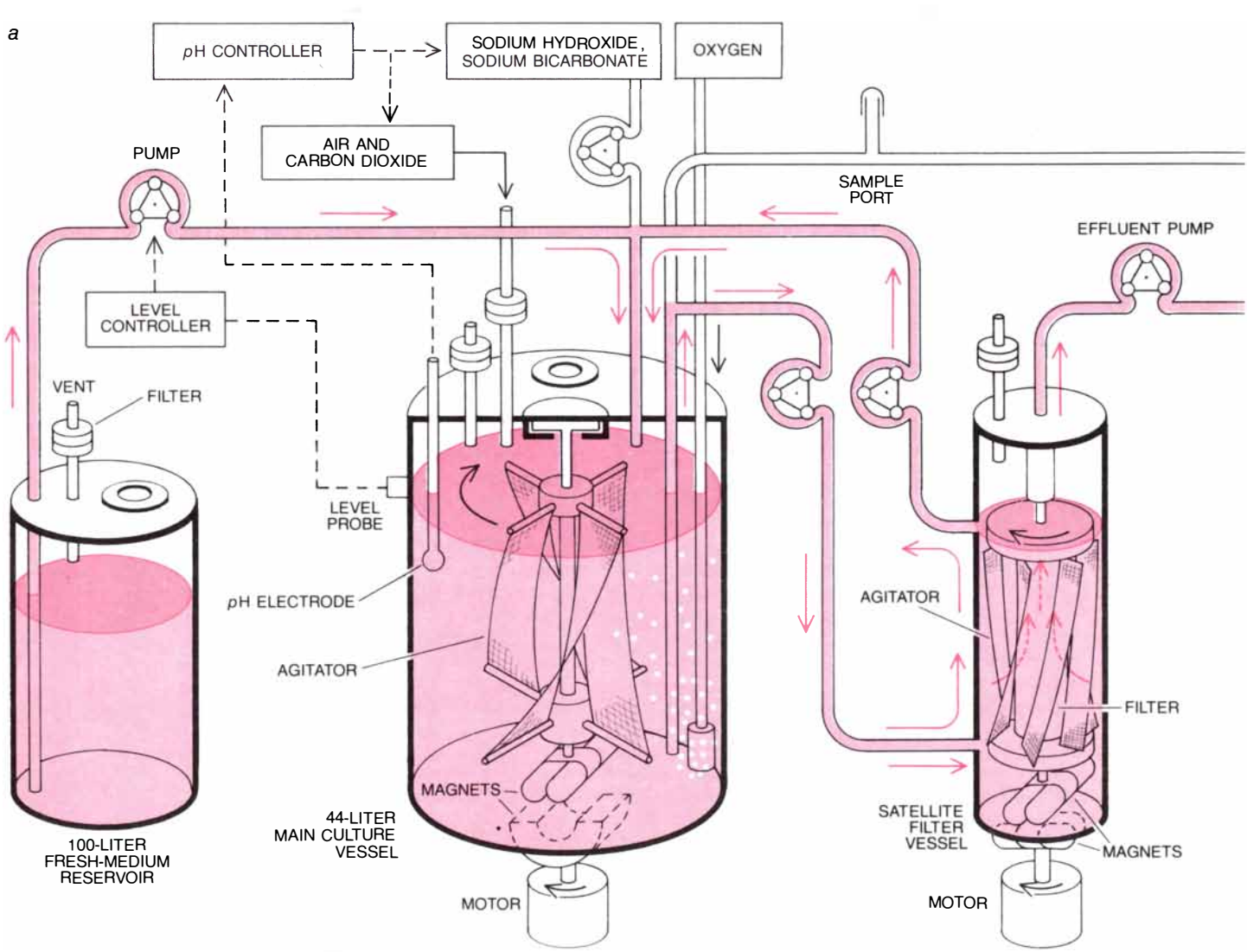
AGITATION is required to keep cells uniformly suspended in the medium and to distribute dissolved gases evenly. The turbine agitator (*top*) is standard in the fermenters in which microorganisms are grown, but its rapid rotation sets up strong horizontal shear forces that can damage fragile mammalian cells. An agitator shaped like a marine propeller (*middle*) provides both horizontal and vertical mixing at a lower rate of rotation. The vibromixer (*bottom*) oscillates vertically, increasing vertical mixing and thereby improving aeration of the medium and reducing cell damage.

(a major metabolic product) in the effluent medium. The rate of growth of the cells was about the same in the two systems but the cell density was much higher in the perfusion reactor: about 25 million cells per milliliter compared with a million cells per milliliter in the conventional reactor. At maximum density all the perfusion-system cells were still alive, whereas some 30 percent of the cells in the conventional one were dead. Perhaps most significant economically, the yield of cells per liter of medium expended was more than two and a half times as high in the perfusion system (even though the medium is constantly renewed in such a system). Moreover, we found that the perfusion reactor can be run in what is called the chemostat mode, in which some cells are continually pumped out of the system along with the expended medium, thereby maintaining a constant cell density at an optimum level.

The vast majority of normal mammalian cells, as we mentioned above, grow only when they are attached to the surface of a culture vessel. Roller bottles and tissue flasks cannot provide a large enough ratio of surface area to volume for practical large-scale culture. The area-to-volume ratio can be increased by means of various stratagems. Cells can be grown, for example, on spongy polymers, on arrays of thin tubing or hollow fibers, on stacks of thin plates or on microscopically small beads called microcarriers. For large-scale culture we have investigated two of these approaches: the hollow fiber and the microcarrier.

Studies done in our laboratory and by Richard A. Knazek and his colleagues at the National Cancer Institute showed that mammalian cells grow well on a variety of hollow fibers with an external diameter ranging from one-third to three-quarters of a millimeter. The fibers are porous, so that air passing through them diffuses to the cells on the outer surface, which is bathed in the culture medium. After studying a number of configurations we finally decided on a "flat bed" arrangement of the fibers. The core of the system is a very shallow bed of fibers, only three to six layers deep, to which the cells attach themselves and where they proliferate.

Fresh medium enters the stainless-

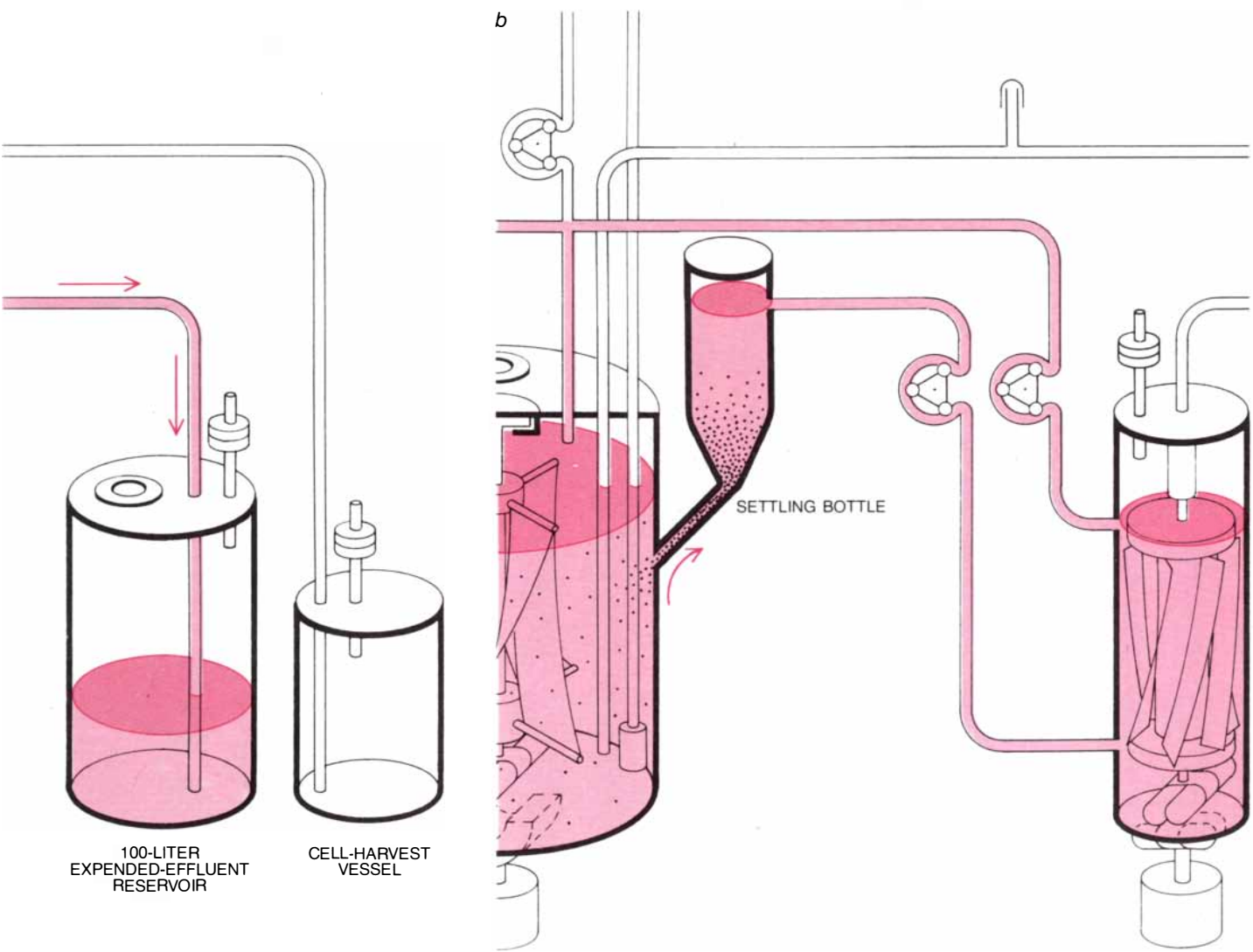


PERFUSION REACTOR designed by the authors moves fresh medium into the system and removes medium in which nutrients are depleted and waste products have accumulated, meanwhile retaining the cells within the system. The diagrams show the version of the reactor developed for cells that grow in suspension (*a*) and a modification (*b*) introduced for cells that grow on microscopic beads called microcarriers. In the main culture vessel gentle agitation is provided by four slowly rotating flexible sheets of monofilament nylon. Medium in the main vessel is continuously pumped to the satellite filter vessel, where some of the medium is pumped through the porcelain filter to the effluent vessel. The filter retains the cells, which are pumped, along with the remainder of the medium, back into the

steel housing below the fiber bed, perfuses up through the bed and is withdrawn at the top of the housing. To make the flow uniform and perpendicular to the plane of the fiber bed we pump the medium up to the bed through a micropore filter: a sheet of stainless steel with pores two micrometers in diameter. Another filter, with 20-micrometer pores, is placed above the bed to limit the amount of backflow. The advantage of the shallow-bed configuration is that the path of perfusion is very short (compared, for example, with the path in a cylindrical "cartridge" system, where the flow of medium parallels the axis of the fibers). As a result there is hardly any gradient from a high to a low nutrient level and from a low to a high waste-product level; all the cells are exposed to essentially fresh medium. The ends of the fibers are bonded together in a nontoxic, elastic material, which is then sliced in such a way that the hollow bore of each fiber is open to the chambers through which air and carbon dioxide are circulated.

We have built two prototype reactors, with fiber-surface areas of 930 square centimeters and 9,300 square centimeters. (The larger prototype is only 40 centimeters square and 4.5 centimeters high.) They have maintained high-density cultures for long periods of time. For example, the smaller prototype grew human lung cells to a density of about a million cells per square centimeter, or about 10 times the density that can be achieved in roller bottles. Perfusion of the medium is particularly effective in these reactors. It can be regulated to maintain a steady-state level of nutrients and wastes as the cells proliferate. That is done by monitoring the level of lactic acid in the expended medium, deriving from it an estimate of the number of cells on the fibers and adjusting the rate of flow of fresh medium accordingly.

The system does not just grow large numbers of cells; in effect it builds an artificial tissue as billions of cells proliferate to encrust the fibers and even penetrate them. Large quantities of cells are thereby maintained under conditions of minimal stress for a long time (from 21 to 59 days in our experiments), during which important molecules such as urokinase and angiogenic factor are produced continuously and are harvested



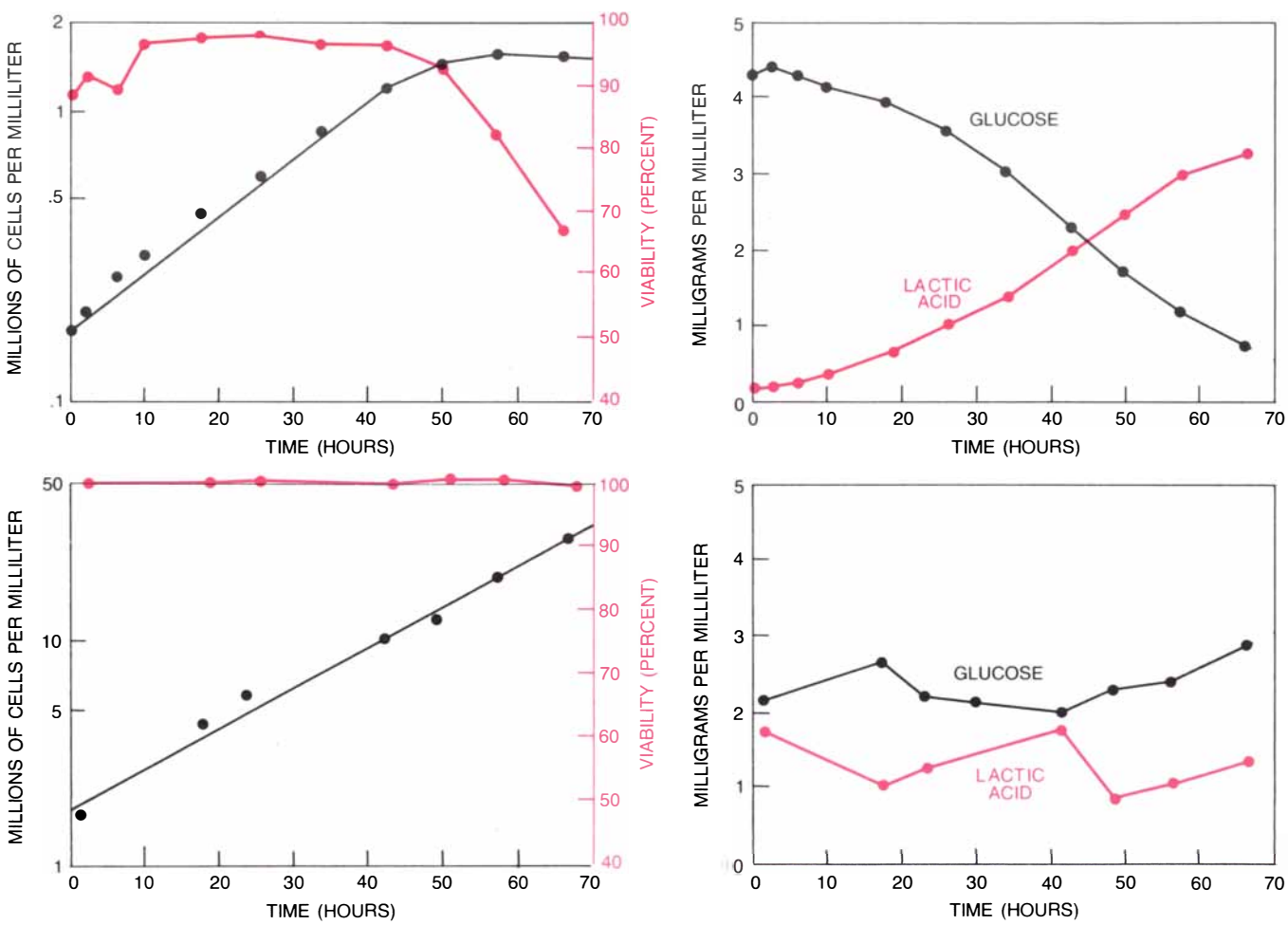
main vessel for further proliferation. A small version of the flexible-sail agitator rotates in the satellite vessel to minimize clogging of the filter pores. Substances secreted by the growing cells can be harvested from the medium in the effluent vessel. The cells can be harvested by being pumped out of the reactor into the cell-harvest vessel. Most normal mammalian cells grow only when they are attached to a surface rather than when they are in suspension in a liquid medium; microcarriers provide the necessary surface. To adapt the perfusion reactor for cells growing on microcarriers a settling bottle was placed in the line between the main vessel and the satellite vessel. Beads encrusted with growing cells settle out of suspension, forming a slurry that slides back into main vessel, where cells continue to proliferate.

from the spent medium. Careful control of contamination should make it possible to operate one of these systems almost indefinitely. Once maximum growth has been attained it is often feasible to substitute a simple maintenance medium, containing little or no serum, for the more expensive medium that supports cell proliferation, and to keep the cells at work producing a desired molecule.

The microcarrier method was first demonstrated in 1967 by Anton L. van Wezel of the Dutch National Institute for Public Health, and many variations of the approach have since been investigated and exploited for industrial production on a moderate scale. The microcarrier beads are composed of the natural glucose polymer dextran or one of various synthetic polymers and range from 50 micrometers to several hundred micrometers in diameter. They are suspended in a nutrient medium. Anchorage-dependent cells inoculated into the medium bind to the beads and multiply. The advantage of a microcarrier system is that although the cells are growing on a surface, the beads that provide the surface are suspended in the medium, and so the cells can be grown in conventional suspension vessels.

Such conventional systems are not, however, very efficient for microcarrier cultures. In addition to such usual concerns as controlling the concentration of nutrients and waste products, one must consider the amount and the distribution in the system of the microcarriers. Collisions between beads can injure cells. The collisions are increased by vigorous agitation, and so there has been a need for a gentler stirring mechanism to keep the beads in suspension without destroying too many cells. To grow large numbers of cells a large surface area is needed. That means a high density of beads, but too high a bead density makes for more collisions. The relatively high surface-to-volume ratio attained with microcarriers gives rise to yet another problem: as the cells proliferate there can be a very rapid depletion of nutrients and a parallel buildup of toxic waste products. Finally, there is the problem of scaling up from one growth vessel to a larger one. The encrusted cells must be removed from the beads, after which they are resuspended in the larger vessel with a new supply of beads. In conventional microcarrier systems it has been hard to remove the cells from the beads without damaging a large proportion of them.

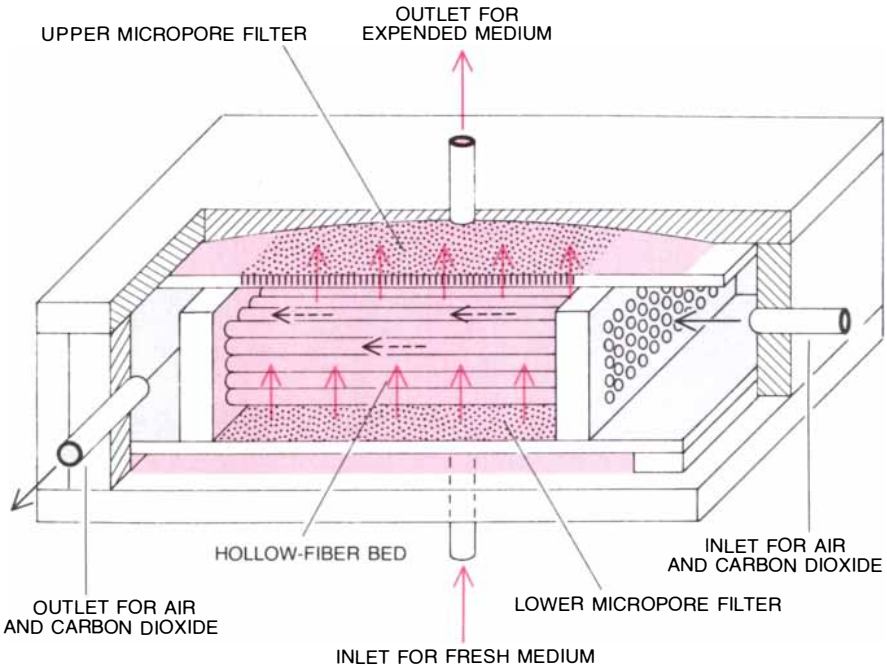
We adapted our novel perfusion system to handle microcarriers and found it made possible a significant increase in cell density. The flexible-sail agitators provide the necessary gentle stirring. Rotating at only from 8 to 20 r.p.m., they keep a high concentration of beads suspended in the medium (as much as 12 grams of beads per liter of medium) without causing cell damage. The major modification was the addition of a settling bottle between the main reactor vessel and the satellite vessel housing



EFFICIENCY OF PERFUSION SYSTEM was demonstrated by a comparison of cell growth (*left*) and depletion of medium (*right*) in a conventional spinner bottle (*top*) and in a four-liter perfusion reactor (*bottom*). Rat tumor cells, which grow well in suspension, proliferated at about the same rate in both systems but attained a much higher cell density in the perfusion system (*black curves at left*) and retained their viability even at high density (*colored curves at left*). The glucose and lactic acid levels reflect the depletion of nutrients and the buildup of waste products in the medium during cell growth. The glucose level fell steadily and the lactic acid level rose steadily in the conventional spinner bottle, whereas in the perfusion reactor virtually constant levels of nutrients and waste products were maintained.

the filter. The recycling pump moves the medium, with its suspended cell-encrusted beads, toward the satellite vessel. Once in the settling bottle, removed from the area where the medium is being stirred, the beads tend to settle and aggregate into a dense slurry that gradually slides back into the main vessel; medium that is essentially bead-free goes on to the satellite filter vessel and is either recycled or removed from the system.

In the settling process the beads and their cells come in close contact and continue to be perfused with fresh medium from the main vessel. Cells bridge from one microcarrier to another, forming large clumps of beads and cells, with the result that there can be up to four times as many cells as one would expect to find on a given bead-surface area. Another advantage, which we had not foreseen, arises from this highly aggregated condition: the separation of the cells from their microcarriers is facilitated, solving the scale-up problem. In the course of aggregation the attachment of individual cells to the beads is somehow reduced, to be replaced by multiple cell-to-cell attachments. When the clumped beads and cells are pumped out of the reactor, a brief treatment with the enzyme trypsin releases the cells from the microcarriers in good condition.

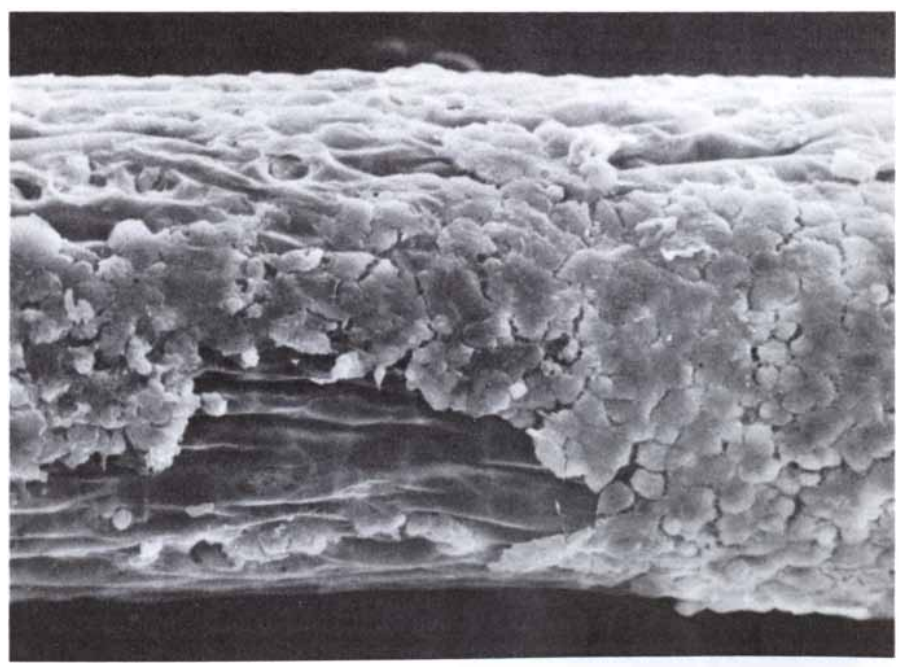


HOLLOW-FIBER PERFUSION REACTOR grows cells on a flat bed of porous plastic fibers .3 millimeter in diameter. The ends of the fibers are open to a supply of air and carbon dioxide, which moves through the fibers (*black arrows*); the outside of the fibers is bathed in a culture medium whose flow (*colored arrows*) is kept uniform by two micropore filters. Cells inoculated into the reactor attach themselves to the surface of the fibers and proliferate, in time forming what is in effect an artificial tissue that can be maintained for several months.

The microcarrier-perfusion reactor has proved to be an efficient system for the large-scale culture of anchorage-dependent cells. Microcarriers have always had the potential of providing a very large surface area. Now all that area can be exploited (actually overexploited, given the bridging effect) because perfusion maintains an optimum environment for cell growth even at extremely high cell densities, which in this system are from five to 10 times as high as before.

The combination of high-density cell growth and efficient scale-up leads to impressive results. In one experimental run of a small (four-liter) microcarrier-perfusion reactor we grew 40 billion human fibroblasts, the equivalent of a normal harvest from 1,300 roller bottles. Those cells were removed from their carriers and were inoculated in turn into a 44-liter reactor along with 400 grams of fresh microcarriers, which provided a surface area of 188 square meters. In the ensuing run 340 billion cells were grown, a harvest that would have required 11,000 roller bottles. The harvested fibroblasts served for the production not only of interferon but also of urokinase and an angiogenic factor. The best measure of the cost-effectiveness of the microcarrier-perfusion system is the yield of cells per liter of medium expended. In all our experiments the yield has been about four times as high as it is in roller-bottle systems.

Efficient large-scale systems for the culturing of a wide range of mammalian cells in the absence of antibiotics and at a reasonable cost are necessary to meet the needs of investigative laboratories as well as for production on a commercial scale. It seems clear that effective new technology can expand knowledge of biological processes and also provide significant quantities of cell products that promise to be important therapeutic agents.



HUMAN CANCER CELLS (HeLa cells) are seen growing on the surface of one of the hollow fibers in this scanning electron micrograph. In addition to eventually covering the entire surface of the fiber the cells penetrate its lobed structure. The enlargement is some 250 diameters.

Machines That Walk

Locomotion on legs resists imitation, but modern control technology should be able to solve the problem. Experiments with machines that hop and crawl can also illuminate the mechanisms of natural walking

by Marc H. Raibert and Ivan E. Sutherland

Many machines imitate nature; a familiar example is the imitation of a soaring bird by the airplane. One form of animal locomotion that has resisted imitation is walking. Can it be that modern computers and feedback control systems make it possible to build machines that walk? We have been exploring the question with computer models and with actual hardware.

So far we have built two machines. One has six legs and a human driver; its purpose is to explore the kind of locomotion displayed by insects, which does not demand attention to the problem of balance. The other machine has only one leg and moves by hopping; it serves to explore the problems of balance. We call the first kind of locomotion crawling to distinguish it from walking, which does require balance, and running, which involves periods of flight as well. Our work has helped us to understand how people and other animals crawl, walk and run.

Unlike a wheel, which changes its point of support continuously and gradually while bearing weight, a leg changes its point of support all at once and must be unloaded to do so. In order for a legged system to crawl, walk or run, each leg must go through periods when it carries load and keeps its foot fixed on the ground and other periods when it is unloaded and its foot is free to move. This type of cyclic alternation between a loaded phase, called stance, and an unloaded phase, called transfer, is found in every form of legged system. As anyone who has ridden a horse at a trot or a gallop knows, the alternation between stance and transfer can generate a pronounced up-and-down motion. We believe legged machines can be built that will minimize this motion.

Our work and related work by others may eventually lead to the development of machines that crawl, walk and run in terrain where softness or bumpiness makes wheeled and tracked vehicles ineffective and thus may lead to useful industrial, agricultural and military applications. The advantage of legged vehicles in difficult terrain is that they can choose footholds to improve traction, to minimize lurching and to step over obstacles. In principle the performance of legged vehicles can be to a great extent independent of the detailed roughness of the ground. Our objective has been to explore the computing tasks involved in controlling and coordinating leg motions. It is clear that very sophisticated computer-control programs will be an important component of machines that smoothly crawl, walk or run.

As we have indicated, locomotion is possible with or without dynamic balance. The animals that crawl avoid the need for balance by having at least six legs, of which at least three can always be deployed to provide a tripod for support. High-speed motion pictures of insects show that they commonly crawl with an alternating tripod gait.

Although a crawling machine that does not need dynamic balance can be built with four legs, such a machine performs awkwardly because its weight must be shifted at each step to keep it from tipping over. Satisfactory performance without active balance calls for at least six legs, since six is the smallest number of legs that always provide a tripod for support even when half of the legs are elevated. Several six-legged machines have now been built, each differing in size and in mechanical design. All of them depend on computer control of the legs.

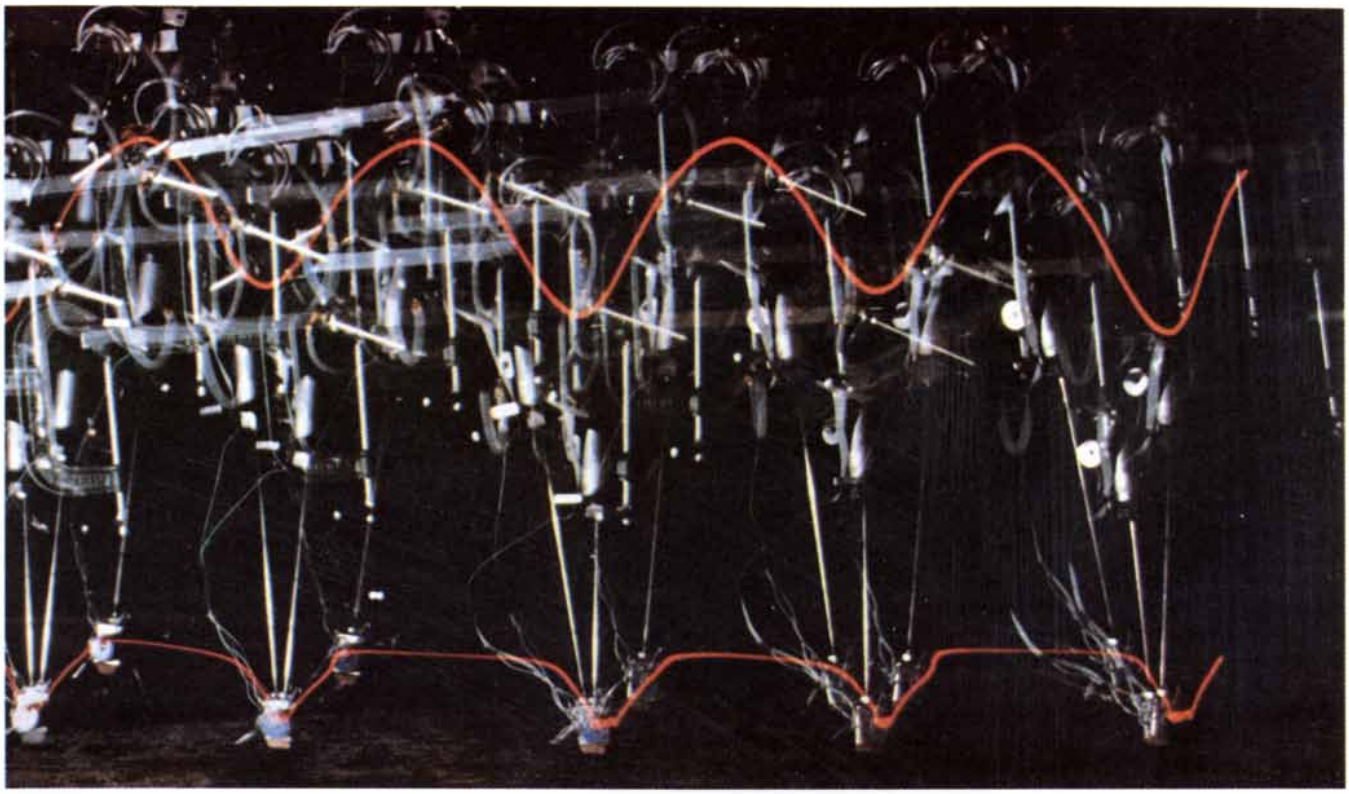
A computer program that controls such a machine accomplishes five tasks. First, it regulates the machine's gait, that is, the sequence and way in which the legs share the task of locomotion. Six-legged machines work with gaits that elevate a single leg at a time or two or three legs simultaneously.

The simplest gaits involve a regular sequence of leg motions. A gait can be described by noting the sequence. For example, the tripod gait can be recorded as (1,5,3;6,4,2;), with the commas designating the concurrent use of legs and the semicolons sequential use. Similarly, gaits that elevate a single leg at a time such as (3;2;1;4;5;6;) and (3;4;2;5;1;6;) are useful. A gait that elevates several legs at once generally makes it possible to travel faster but offers less stability than a gait that keeps more legs on the ground.

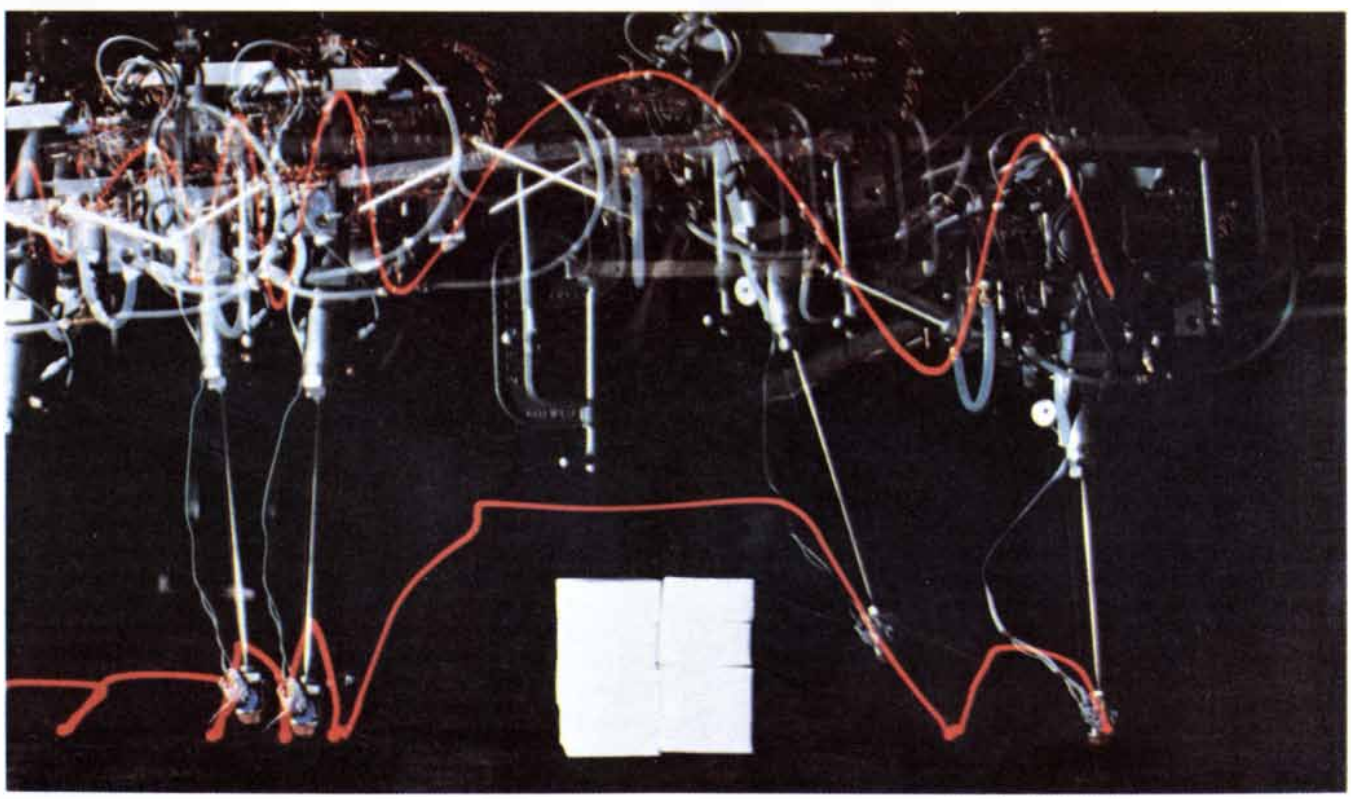
A second task of a computer program controlling a crawling machine is to keep the machine from tipping over. If the center of gravity of the machine moves beyond the base of support provided by the legs, the machine will tip. The computer must monitor the location of the center of gravity of the machine with respect to the placement of the feet to ensure that the base of support is always large enough. For simple gaits the geometry of the legs may suffice to keep an adequate base of support, but for more complex gaits a careful computation of static stability may be critically important.

Since many legs share the support of the machine, a third task of the control computer is to distribute the support load and the lateral forces among the legs. In the tripod gait, of course, the distribution of the support load is set by the geometry of the three supporting legs. With more than three supporting legs, however, the control computer must decide how to manage the distribution of loading in order to achieve higher-level objectives such as smoothness of ride and minimal disturbance of the ground.

Even when only three legs are supporting the machine, the control program must distribute the lateral foot forces. One way of looking at this task is to consider that the control system must keep the machine from simply doing isometric exercises against the ground. The amount of sensing and computation that is needed to distribute the lateral loads among many legs can be formidable. We have reduced this burden for the crawling machine we are building by providing passive hydraulic circuits that automatically distribute the sideways loads.



HOPPING MACHINE was built by one of the authors (Raibert) to explore the problems of controlling a machine that must balance as it moves. Its leg is actuated by compressed air and its motions are controlled by a computer that obtains feedback from position sensors. This version is held by a tethering arm and so balances in a single plane. A series of hops is recorded in this photograph, which was made while the camera lens was kept open. Red lights attached to the body and foot of the machine delineate the path of the hops.



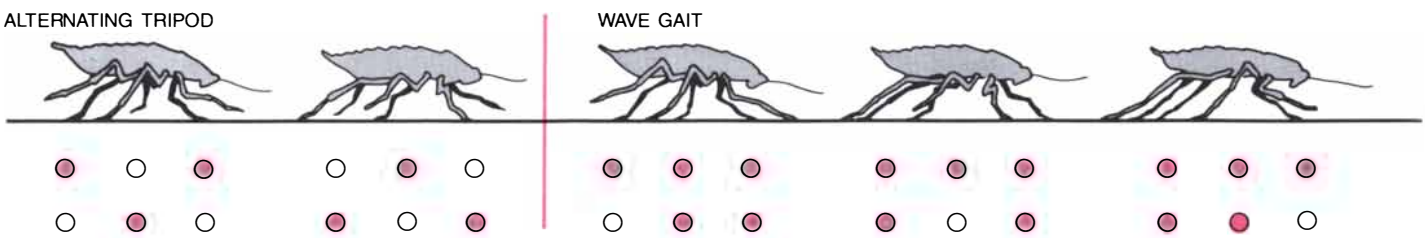
JUMPING MODE of the hopping machine shows it leaping over an obstacle. The machine approaches the obstacle from the right. When it is one step away, the operator pushes a "leap" button. As a result the maximum tension is generated in the drive actuator so that the altitude of the next hop will be increased. In flight the leg is shortened and its normal swinging motion is delayed to provide better clearance of the obstacle. A servomechanism controlling balance moves the leg to the correct landing angle and the leg is lengthened in preparation for landing. Thereafter the machine continues its normal hopping. The obstacle was 15 centimeters (six inches) high.

A fourth task of the control computer is to make sure the legs are not driven past the limits of their travel. The geometry of the legs may make it possible for one leg to bump into another; if legs can collide, the computer must limit their motion to prevent damage. To maximize the usefulness of each leg its placement on the ground must take into account the limits of the leg's motion and the expected motion of the machine during that leg's stance period. For example, if the machine is turning to the right, the forward legs should be placed farther to the right so that their sideways travel can be accommodated during the turn. For a vehicle with autonomous control the placement of the legs can be based on the planned future path of the vehicle. For a vehicle with a human driver the proper placement of each leg requires a prediction of the driver's commands for the next stance period.

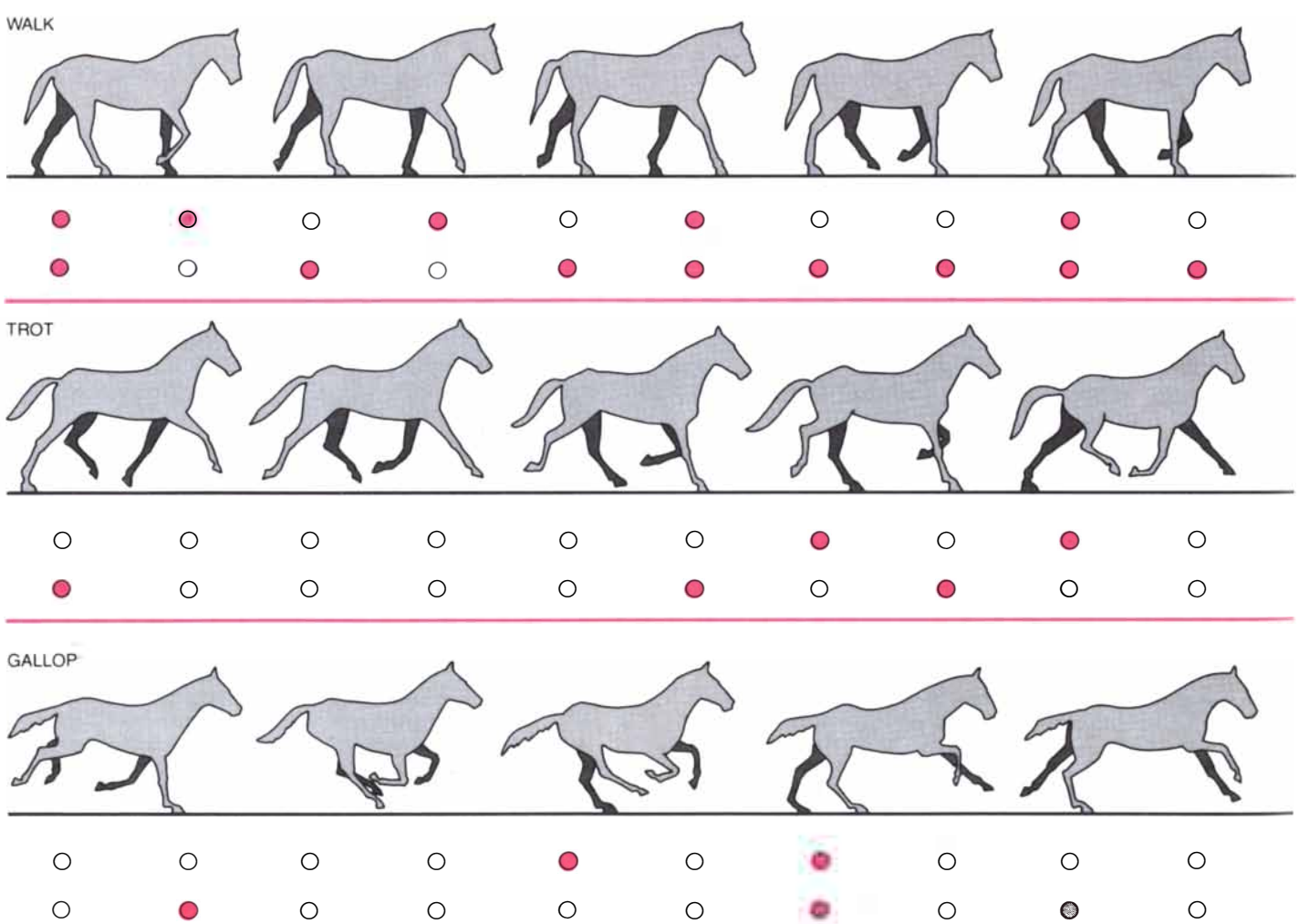
A fifth task for the control computer is to choose places for stepping that will give adequate support. On smooth ground the task is easy, but on rough terrain it may be exceedingly difficult. No system has yet been built that accomplishes this task. One can envision a terrain-scanning system that would survey the ground ahead of the machine and choose likely footholds. To make use of such a scanner the control computer would build an internal digital model of the terrain. Such a model would have to account only for bumps that are about the size of the machine's feet or bigger. Human input to the model might help in the evaluation of possible footholds.

One of us (Sutherland) is building a six-legged, hydraulically driven crawling machine. A gasoline engine provides its power and hydraulic actuators move its legs. There are six legs, so that dynamic balance is not needed.

A built-in microprocessor controls the legs by switching on or off the valves that regulate the flow of oil to the hy-



GAITS OF INSECTS provided a basis for designing the method of locomotion of the six-legged crawling machine portrayed on the cover of this issue. The circles below each drawing show whether the corresponding leg is on the ground or in the air: a filled circle represents



GAITS OF A HORSE represent the kind of locomotion in which balance is a factor. In the walk at least two of the horse's legs touch the ground at all times. In the trot and the gallop the animal periodically leaves the ground. The drawings are based on the stop-motion pho-

draulic actuators. Sensors in each leg report its position and the forces acting on it to the microprocessor. The machine is large enough to accommodate a human driver, who controls its speed and direction of motion and establishes the tilt of its body and its ground clearance. The vehicle's design speed is about two miles per hour.

One objective in the design of the vehicle was to minimize the amount of computation required to obtain a crawling movement. The hydraulic circuits are designed to make the legs move along useful paths without attention from the microprocessor, which merely selects one of the available paths for each leg. Thus the microprocessor is free to concentrate on selecting which legs to use for support and on deciding where to step next; it does not have to spend time computing the details of leg motion.

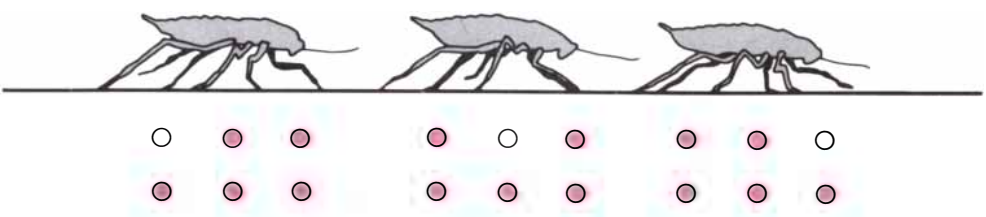
Each leg of the machine can swing fore or aft and up or down on the universal hip joint that attaches it to the frame of the machine. These motions are executed by lengthening or shortening the two hydraulic actuators per leg that are arranged in a *V* configuration above the leg. One setting of the valves provides that oil leaving one actuator will enter the other, so that as one actuator shortens, the other actuator lengthens by the same amount. Because of the geometry of the pivots this connection provides horizontal leg motion.

The horizontal motion can be powered or unpowered depending on the valve settings, so that some legs can serve to drive the machine forward while others coast. As legs are placed on the ground and accept load they are able to coast forward or backward as dictated by the motion of the legs already on the ground and driving. Hence the control computer does not need to compute the precise instant when a leg will touch the ground or the details of the motion required at the time of contact to obtain a smooth forward motion.

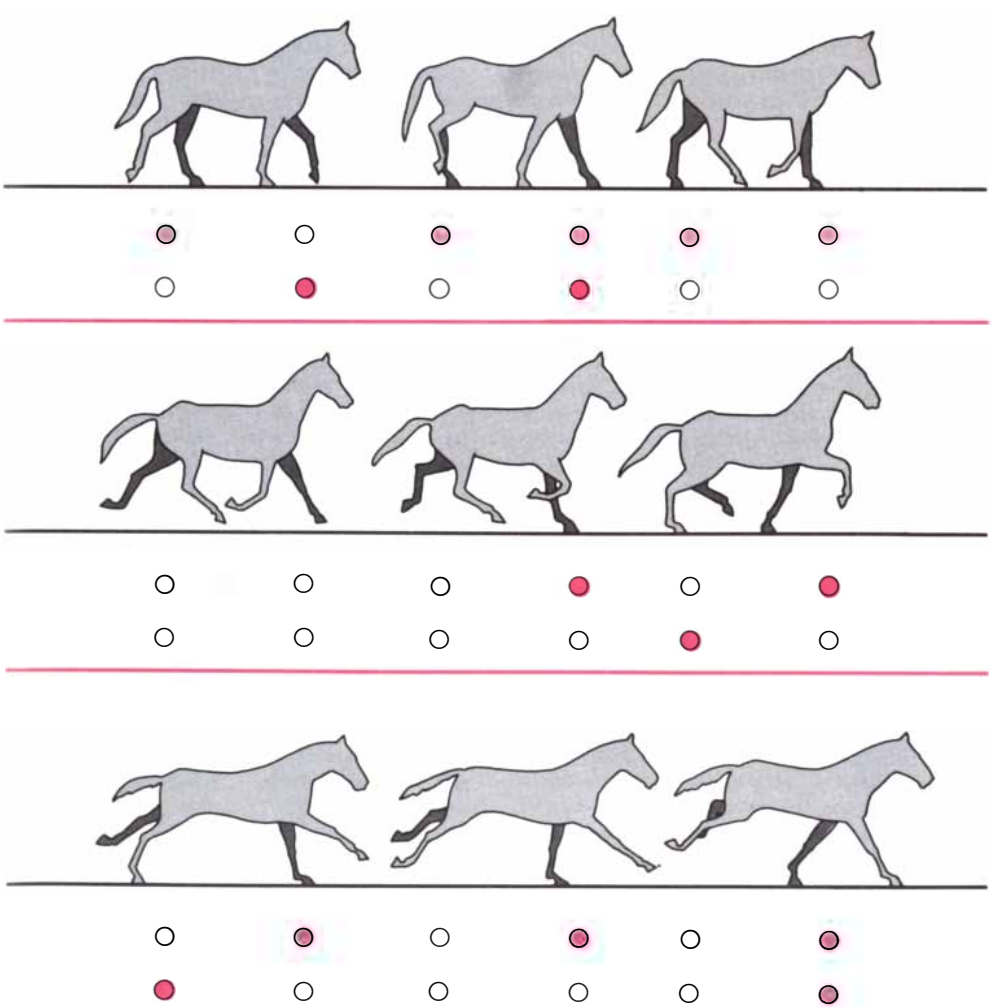
The knee joint of each leg is powered by a separate hydraulic actuator mounted horizontally along the leg. This actuator can be powered while the leg is raised in order to position the foot sideways for the next step. When the foot is on the ground, the knee joint must move slightly to match the circular path of the knee about the hip to the straight path of the foot on the ground. It is a complex motion, but it does not call for action by the computer; instead a simple parallel connection of the knee-joint actuators enables all the knees to accommodate to the average motion of the vehicle. An additional hydraulic pump provided in the system can force a collective sideways motion of all the knee joints, making the machine crawl sideways like a crab.

The human driver of the machine has three kinds of control. First, he can regulate the amount of oil flowing in the system because he can control the displacement of the hydraulic pumps. Separate pumps are provided for the legs on the left and right sides so that the driver can steer by making the machine crawl faster on one side than on the other. The settings of the steering controls are reported to the microprocessor so that it can position the feet properly. For example, if the machine is turning to the right, the front feet must be transferred to the right and the rear ones to the left to accommodate to the turn. If the machine is walking backward, which is achieved by reversing the flow of oil, the feet must be transferred backward with each step rather than forward. As each foot is lifted from the ground the control computer picks a target position for it based on the current rate and direction of oil flow set by the driver. When a supporting foot nears the limit of its travel, the control computer initiates its lift and transfer to a new foothold. If any supporting foot actually reaches the limit of its travel, the microprocessor stops the vehicle until that foot can be lifted from the ground and transferred to a new foothold where it has room to move.

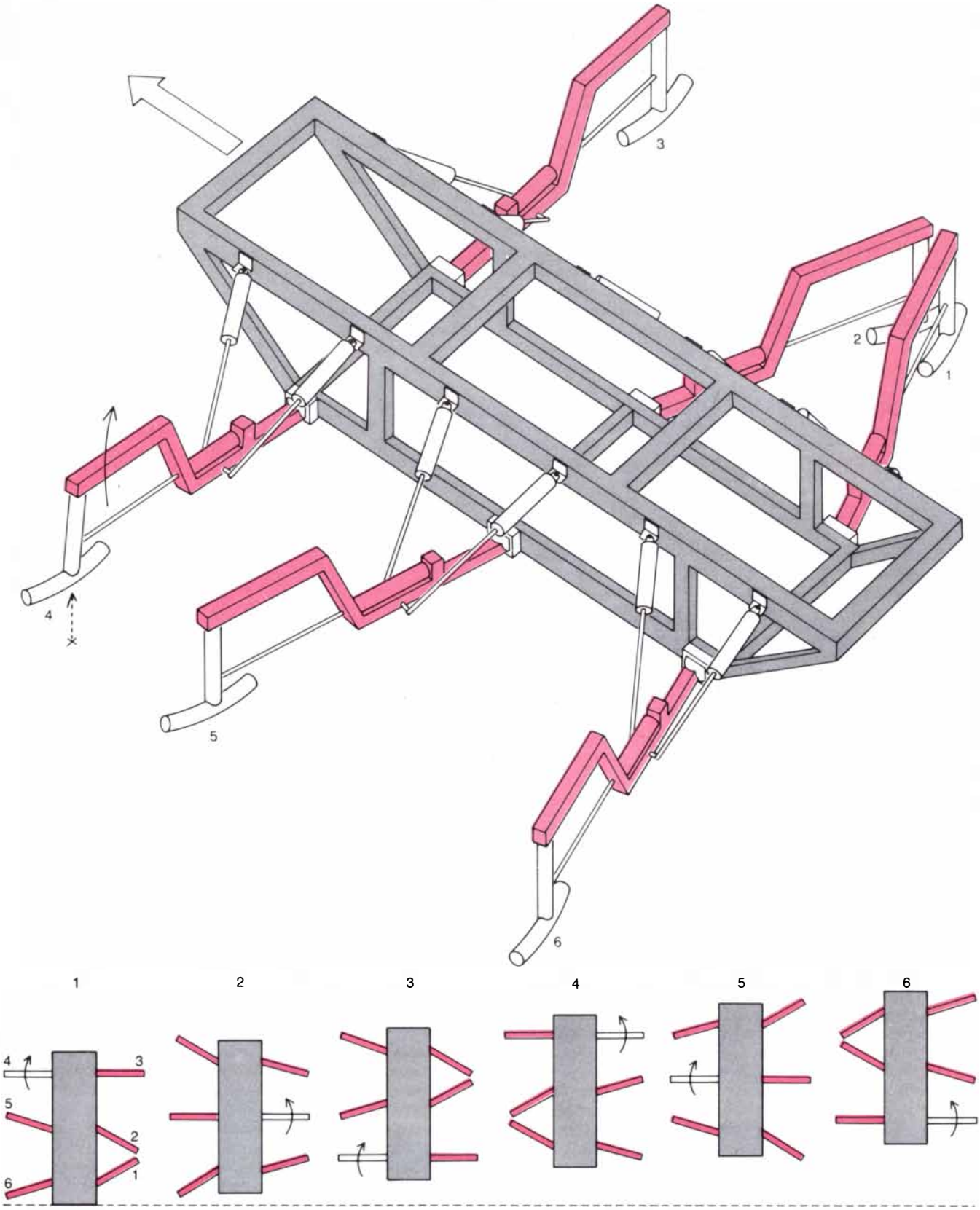
The second kind of human control of the vehicle establishes the attitude and



a leg touching the ground, an open circle a leg in the air. The alternating-tripod gait always provides stability. In creeping with a wave gait the insect's adjacent legs move successively.



tographs made by Eadweard Muybridge 100 years ago that settled a long-standing debate on whether a horse in a trot leaves the ground. Many other animals also do so in running.



SIX-LEGGED MACHINE built by one of the authors (Sutherland) moves by crawling. It does not have to balance. The six legs are controlled by a built-in microcomputer. Power comes from an 18-horsepower gasoline engine that drives separate hydraulic pumps for the left and right legs. A human driver steers the machine by making the oil flow at different rates on the two sides. Sensors report the driver's commands as well as the position of each leg and the forces on it to the microcomputer, which employs the information to choose the order and path of leg motion. Six legs ensure stability because at least three are always on the ground. Passive hydraulic circuits simplify the computing task; a leg that is supporting weight can either be connected to the drive unit or can coast, being pushed by the ground according to the motion generated by the other legs. The diagram at the bottom indicates the position of the legs in a walking cycle, which can be denoted as (4;2;6;3;5;1;). A solid rectangle represents a leg touching the ground and an open rectangle represents a raised leg moving forward as indicated by the arrow. In the numbered walking cycle the semicolons denote the sequential use of the machine's six legs.

ground clearance of the machine. The driver can set a control that changes the vertical support position for the left and right feet to make the vehicle roll left or right. Similarly, he can indicate different vertical support positions for front and rear feet to make the machine pitch forward or back. Another control enables him to indicate that the vertical support positions for all six legs should be raised or lowered collectively to change the ground clearance of the vehicle.

The third kind of human control will achieve careful placement of the feet for operation on very rough terrain. We have not yet decided how to provide this kind of control. A walking truck built some years ago by Ralph Mosher at the General Electric Company depended exclusively on manual control of foot placement and was therefore quite tiring to drive. We believe selection of the gait may also be important, but we have not yet had enough experience to know whether it could be done automatically or whether human inputs will be needed. It is precisely to answer such questions that we have built the machine.

The other subject of our attention is walking and running where balance plays a role. Until a century ago people still debated whether or not a horse in a trot had all its legs off the ground simultaneously. The stop-motion photography of Eadweard Muybridge settled the debate, showing that a horse does leave the ground entirely during a trot. A running person does so too, as do the dog, the cheetah and of course the kangaroo. Such animals not only walk, which requires dynamic balance, but also run, employing ballistic motions effectively to increase their rate of travel.

There are two fundamental differences between a crawling vehicle that is statically balanced and one that is dynamically balanced. The first difference is in the definition of stability. A crawling vehicle is stable if its legs provide at least a tripod of support at all times to ensure that it does not tip over; a dynamically balanced walking or running vehicle can be allowed to tip for brief intervals. Motions of the legs and the body ensure that a single tipping interval is brief and that an adequate base of support is maintained on the average. For example, a running man touches the ground alternately with his two legs, providing a base of support for his body only over time.

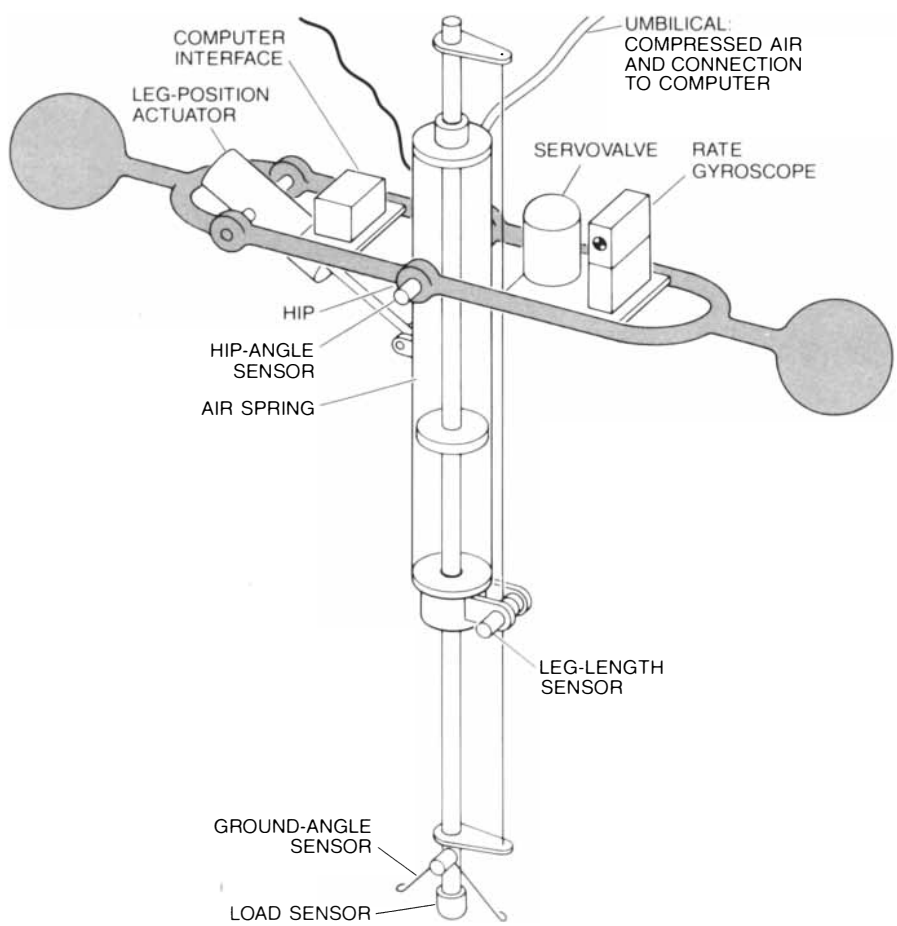
The second difference between static and dynamic balance is in the consideration of speed and momentum. Static balance assumes that the configuration of the supporting legs and the position of the center of gravity are adequate to specify stability; it ignores the vehicle's motion. Such static computations are not always sufficient. For example, a fast-moving vehicle might tip forward if it stopped suddenly with the center of gravity too close to the front legs. In order to understand the greater mobility of walking and running systems one must both relax the definition of stability and account for velocity in computing balance.

It is to study the problem of balance in its simplest form that one of us (Raibert) and his co-workers at Carnegie-Mellon University have built and demonstrated a machine that hops on its single leg and runs like a kangaroo, in a series of leaps. The device can be thought of as a computer-controlled pogo stick. We have been encouraged by the remarkable simplicity of the balancing algorithm. In its present form the machine is limited to movement in a single plane, so that it can tip over in only one direction.

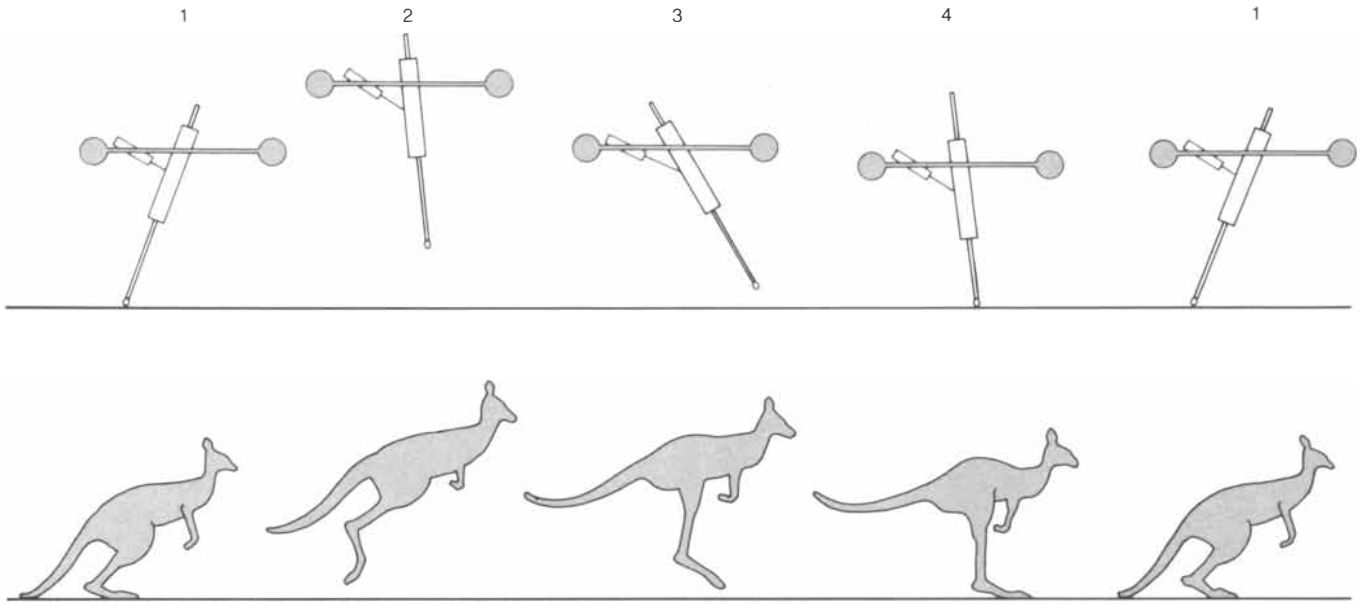
The machine has two main parts: a body and a leg. The body provides the main structure and carries valves, sensors and electronics. The leg is a simple mechanism that not only changes length along its axis but also pivots with respect to the body at a hinge called the hip. The leg bounces on a spring with adjustable tension, much like a human leg with its springy muscles and tendons. The spring is an air cylinder in which pressures are controlled with sensors and valves. At the bottom of the leg is a small foot.

The pivoting motion of the leg is controlled by a second air-operated actuator that applies torques at the hip hinge. A simple on-off valve controls the leg spring, but control of the pivot angle of the leg requires a proportional servovalve, that is, a feedback device that responds in proportion to the strength of the signal it receives. Because the moment of inertia of the leg is less than 10 percent of the body's moment of inertia the leg can pivot during flight without imparting much motion to the body. The tilt of the body is measured by a gyroscope, enabling the control computer to maintain the body in a level attitude. Other sensors measure the angle of the hip, the length of the leg, the air pressure in the leg spring, the angle between the leg and the ground and the force of the leg's contact with the ground.

Three separate servo-control loops



TWO-DIMENSIONAL HOPPING MACHINE serves to study the problems of controlling motion with balance. The machine normally operates while leaning parallel to a tilted wall, separated from it by a cushion of air; the machine can tip only in the plane defined by the wall. A computer receives data from the angle sensors, a pressure sensor and a switch on the foot. It controls hopping by adjusting the pressure in the pneumatic chamber that functions as a spring. The computer also applies torque between the leg and the body to regulate the angle of the hip. This angle determines the horizontal displacement of the foot and therefore influences balance.



HOPPER IN MOTION operates cyclically, as all legged systems do; the leg alternates between periods of support and periods of flight. At the left the machine is about to begin a leap. While it is in the air the leg swings forward at the hip in preparation for the next landing. At touchdown the leg spring shortens to its minimum length to provide for the next leap. A ground-contact sensor acts as a trigger for the vertical-control program. The machine also has feedback loops to control attitude and balance in synchrony with the vertical control. Like a pogo stick, the machine can balance only while it hops. The hop is like the kangaroo's movement Muybridge called a ricochet.

regulate the machine. One loop controls vertical motion, one balance and one body attitude. Each loop is synchronized with the basic hopping motion.

The first loop controls the height of the hopping motion. It adds or removes energy from the motion in order to achieve the correct hopping height and makes up for the energy lost during each hop. The height control does both tasks by periodically adding air to or releasing it from the main drive cylinder to adjust the effective tension of the air spring. In other words, the height control governs the timing and the magnitude of the power delivered to the hopping drive mechanism, thereby achieving the desired hopping height. When a desired hopping height has been achieved, most of the energy needed for the next hop is recovered from the spring, in which it was stored during the previous landing. As long as the hopping motions are relatively stable the task of managing the hopping energy of the machine is not particularly difficult.

The second servo-control loop provides for the balance of the machine by positioning the foot while the machine is in flight so that the next landing is made in a balanced posture. The calculation of the correct foot position takes into account both the forward speed of the vehicle and the inclination of the body. A single computer algorithm for balance works when the machine is hopping in place, accelerating to a run, running at a constant velocity, leaping over objects and slowing to a stationary hop.

When the machine is hopping in place, the leg and foot are moved small distances to compensate for external disturbances and the errors of previous hops. When the machine is to start running, say to the right, the foot is moved first to the left to unbalance the vehicle so that it starts to tip in the desired direction. Stable running is just like hopping in place except that the balancing adjustments supplement large sweeping motions of the leg, which are determined by the rate of travel. Stopping is much like starting except that the machine is made to tip in the direction opposite to the direction of movement.

The third servo-control loop stabilizes the attitude of the body to keep it upright. It provides torques between the leg and the body while the foot is on the ground in order to achieve the desired attitude during the next flight. The effectiveness of this servo depends on good traction between the foot and the ground. The attitude servo that operates when the foot is on the ground shares the hip-drive mechanism employed by the balance servo that operates while the machine is in flight. Certain subtle details of the change from one control mode to the other are associated with detecting the start and finish of each flight. The torquing mechanism must be idle during these events lest the foot slip on the ground.

When an animal runs, its legs swing back and forth through large angles to provide balance and forward drive. We have found that such swinging motions of the leg do not have to be explicitly programmed for a machine but are a natural outcome of the interactions between the controllers for balance and attitude. Suppose the vehicle is traveling at a constant horizontal rate and is landing with its body upright. What must the attitude controller do during stance to maintain the upright attitude? It must make sure that no torques are generated at the hip. Since the foot is fixed on the ground during stance, the leg must sweep back through an angle in order to guarantee that the torque on the hip will be zero while the body moves forward.

On the other hand, what must the balance servo do during flight to maintain balance? Since the foot must spend about as much time in front of the vehicle's center of gravity as behind it, the rate of travel and the duration of stance dictate a forward foot position for landing that will place the foot in a suitable spot for the next stance period. Thus during each flight the leg must swing forward under the direction of the balance servo, and during each stance it must sweep backward under the control of the attitude servo; the forward and back sweeping motions required for running are obtained automatically from the interplay of the servo-control loops for balance and attitude.

We are now building a version of the machine that will balance in three dimensions and therefore be able to move around on an open floor. We have written and tested a computer simulation of the motion of such a machine and have found that control in three dimensions can be broken down into the same three servo-control loops we have described.

Our work in making the one-legged machine run was greatly aided by a

thought that came to us as we went along. It was that running can best be understood by breaking it down into the three parts we have discussed here: height control, balance control and attitude control. Partitioning the control into these three parts has made the complex behavior of legs in walking and running much easier to understand. This insight has led to a fairly simple control system that makes the one-legged machine balance and run.

Our success in this effort encourages us to think about building dynamic-motion machines with more than one leg. We believe the right way to think about such machines is to focus first on their up-and-down and balancing behavior, postponing the complications introduced by forward motion. The notion that hopping is the main activity was natural to the one-legged machine and provided an effective way to think about its behavior, but it seems less natural for machines with several legs. Perhaps it seems less natural because we are accustomed to seeing animals run and want to understand their behavior all at once.

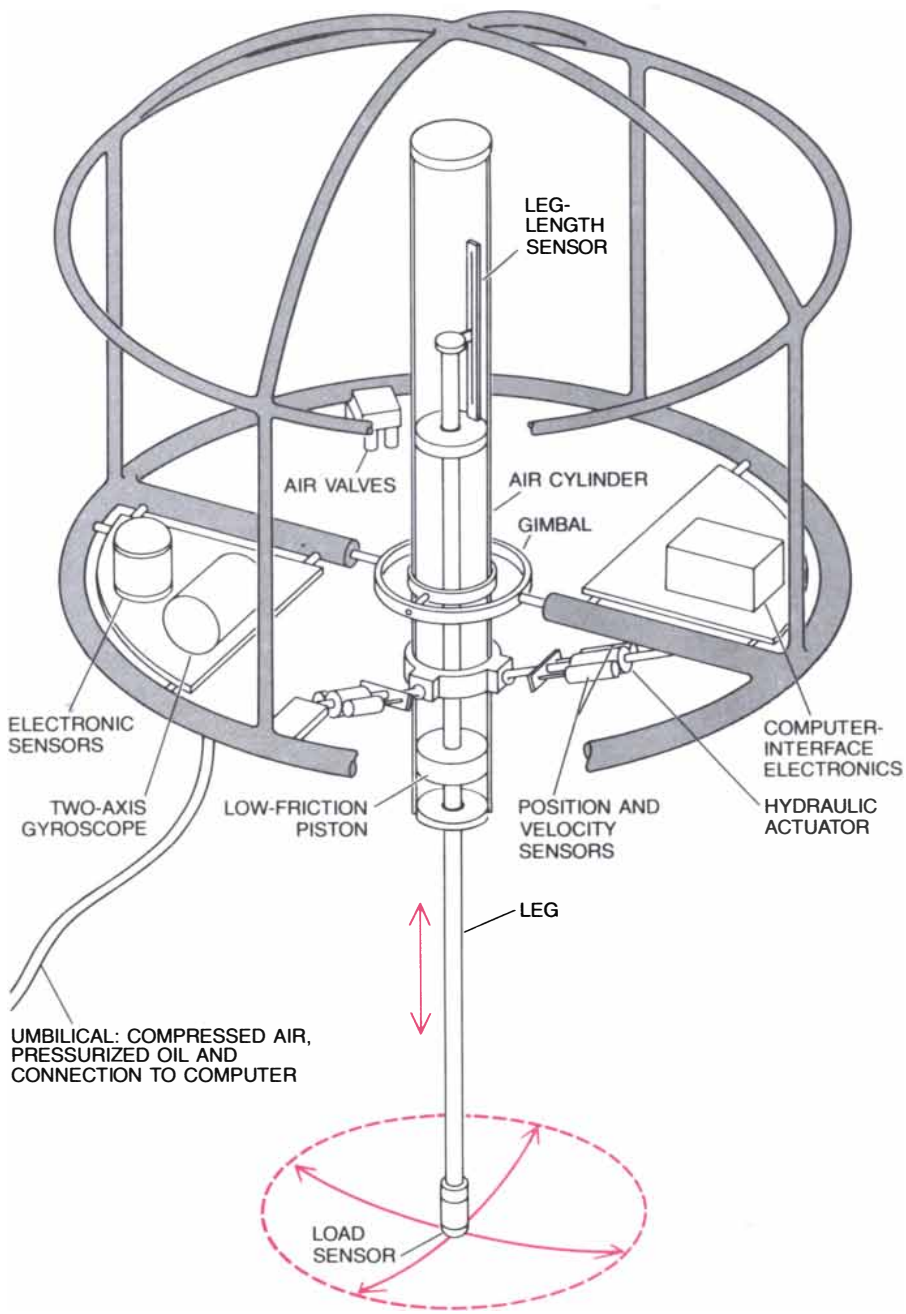
A four-legged machine hopping in place might use any of several sequences of leg activity. The simplest pattern would be to hop simultaneously on all four legs. It is not hard to imagine that the same three servomechanisms that control the one-legged hopping machine might control the motion of a four-legged vehicle in this mode. In fact, the attitude-control loop that keeps the body upright might be substantially simpler because of the broader base of support. When the machine moved forward, the legs would swing together in a pattern of motion that could easily be generated by the same control mechanisms as those that serve the one-legged machine.

Another possible gait for a four-legged machine hopping in one place is bouncing on diagonally opposite pairs of feet. Again it does not stretch the imagination too much to see how one might separate the control of such a vehicle into a height control, a balance control and a body-attitude control. The height control would add energy to the hopping motion to keep the hopping height at the desired level. The balance control would position the two raised legs in such a way as to maintain balance. The body-attitude control would apply appropriate torque to the pair of legs on the ground. The attitude and balance controls would alternate in the use of the same leg actuators, as they do in the one-legged machine. Moreover, just as we have found in that machine, forward motion could easily be accommodated by moving each raised foot to a forward position chosen to make the average balance force of the leg zero during the next stance period. The resulting motion is a trot of the kind common among four-legged animals.

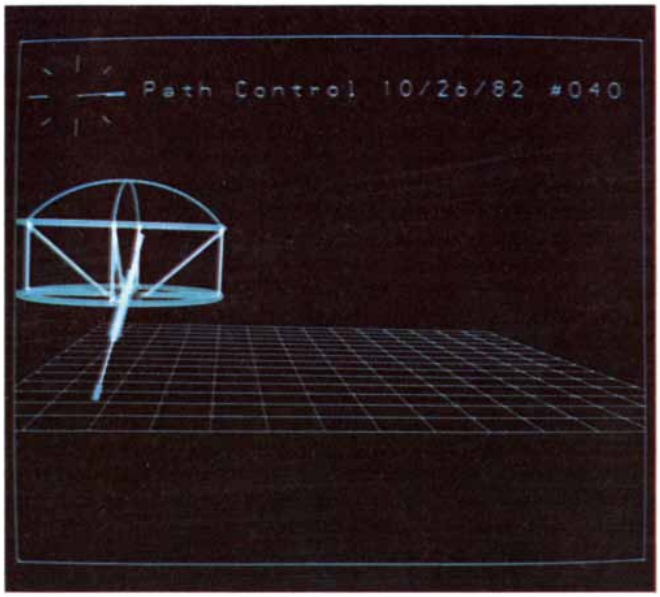
Two other gaits can similarly be understood by separating the control of each leg into vertical, balance and body-attitude components. In the gallop the rear legs land slightly sooner than the front ones. The body attitude is allowed to change during flight so that a nose-up attitude is seen as the rear legs touch down and a nose-down attitude develops as the front legs take off. The bound is a variation of the gallop in which the front legs operate nearly in unison and so do the rear legs but the front and rear actions are equally spaced in time. It is the bound that enables the cheetah to sprint at speeds of more than 60 miles per hour.

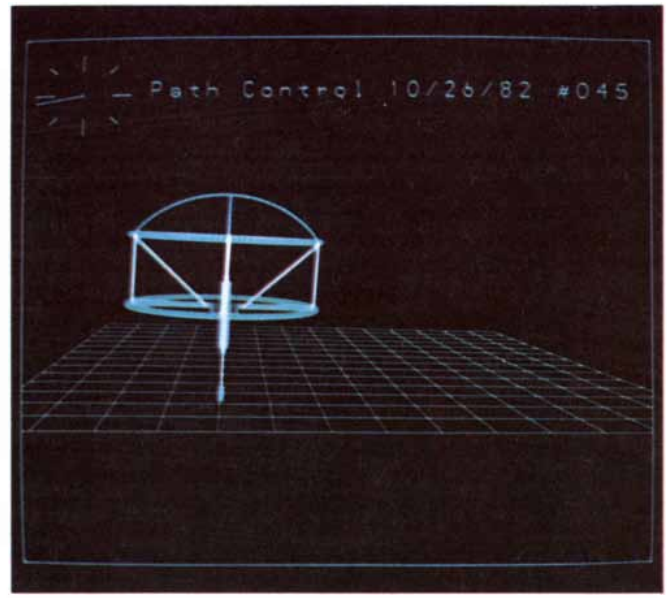
Efficient motion over the ground requires that little energy be lost during each motion of the machine. We have already mentioned how the vertical motions of legs can be made efficient by storing energy in elastic elements. What about fore-and-aft leg motions?

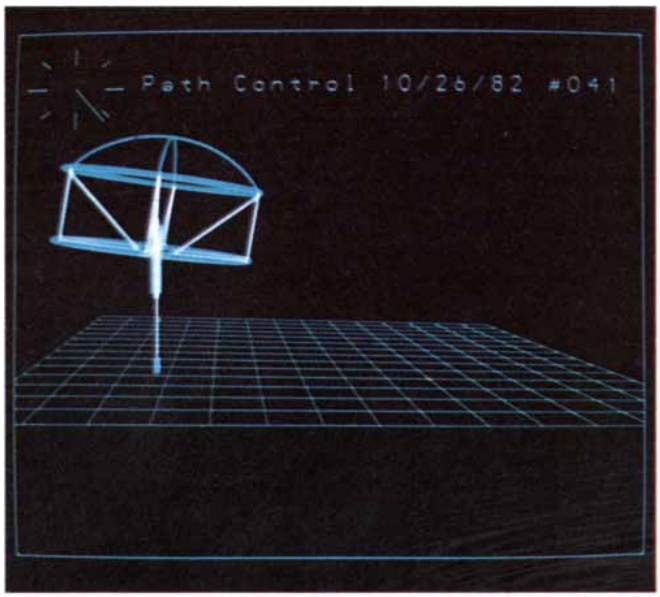
At high speeds over the ground the legs of a vehicle will have to move forward and back quite rapidly. Most of the energy expended by a running animal goes into generating these leg mo-

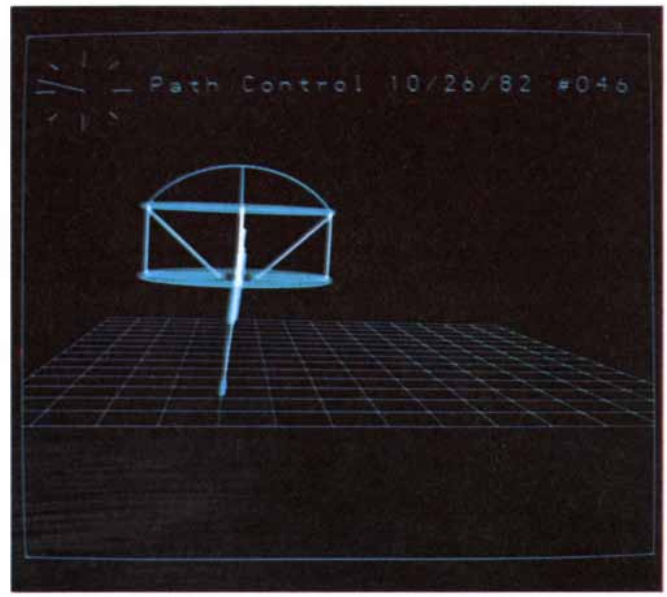


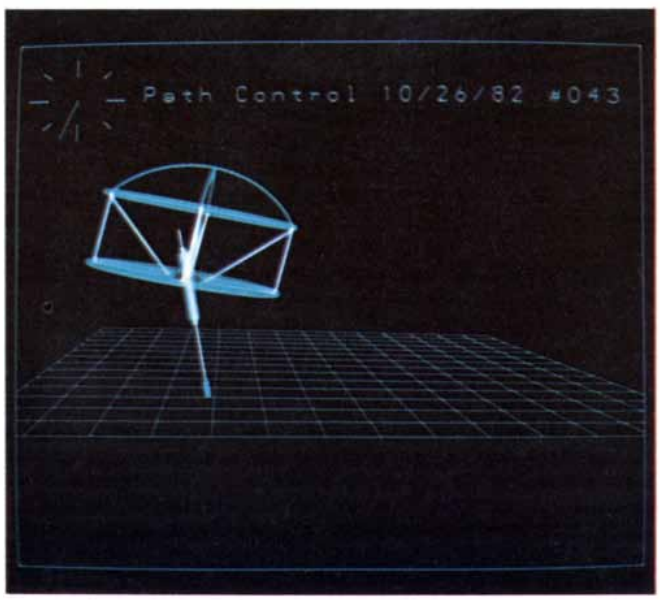
ADVANCED HOPPING MACHINE now under development is designed to operate in three dimensions. It is about one meter high, weighs 20 kilograms and is connected to a nearby computer. Compressed air provides the hopping power and regulates the height of hopping. The actuators that position the foot are hydraulic. During flight they position the foot to maintain balance. When the foot is on the ground, they maintain the machine in an upright body posture.

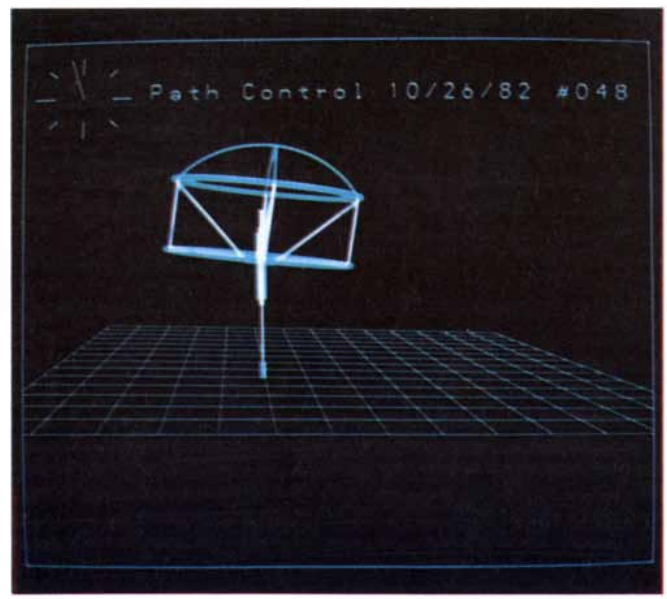












SIMULATED MOTION of the three-dimensional hopper is shown in these photographs from the display of the computer that worked out the motion. Here over a period of approximately .7 second the machine is shown balancing itself while it lands and takes off again.

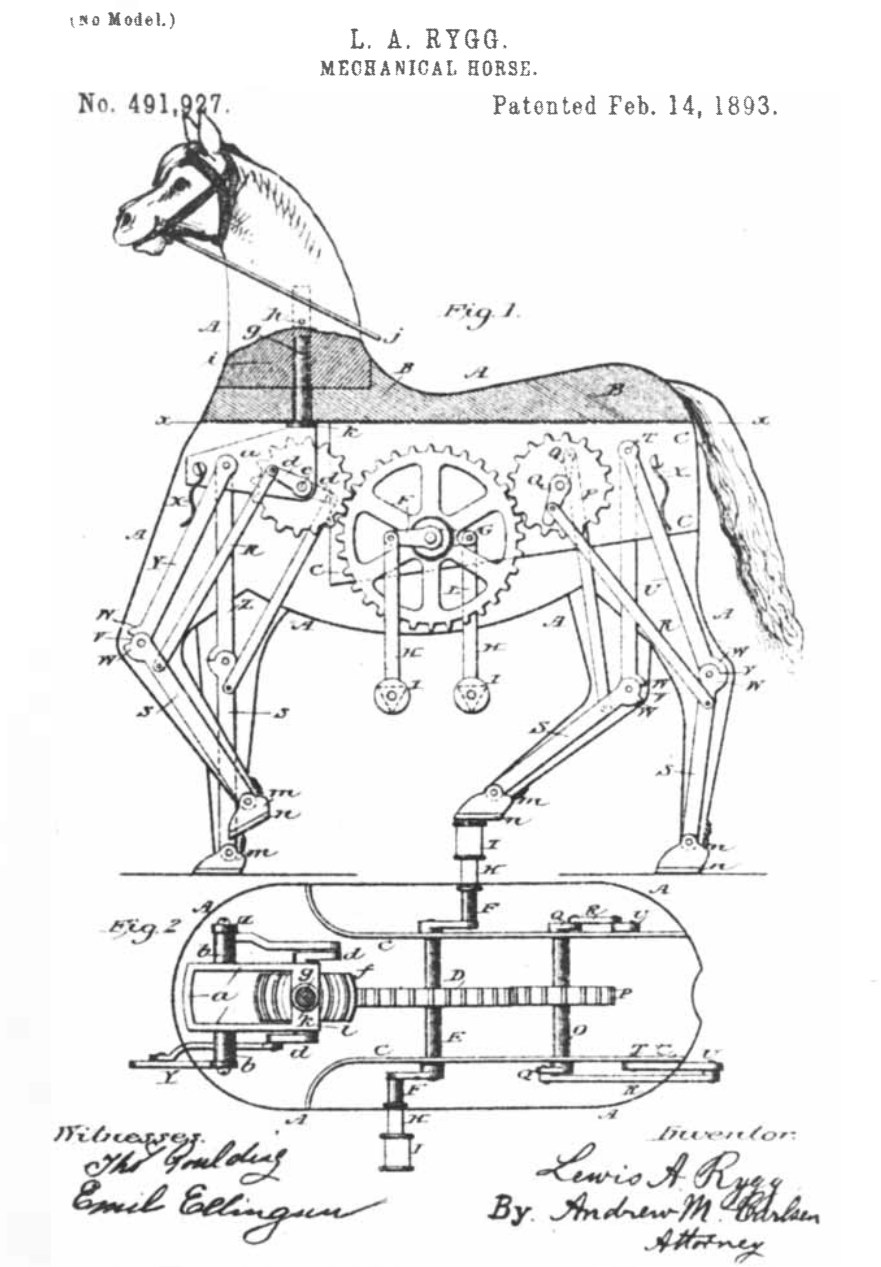
tions. In our one-legged machine these motions are provided by a conventional proportional servomechanism. In such a system the kinetic energy of the leg as it swings is entirely lost as the leg is brought to rest momentarily at each extreme angle. The hopping motion, on the other hand, is obtained by a self-resonant system made up of the leg spring and the mass of the machine, so that the height servo need only add or subtract a small amount of energy to maintain the hopping height. It is obvious that if a machine with several legs is to be made efficient as well as fast, it will have to incorporate some kind of self-resonant system for the fore-and-aft motion of the legs as well. One might design such a mechanism with springs between the legs to make the legs oscillate like a tuning fork at a frequency appropriate to the vertical bouncing rate.

Although we believe we understand how to build a four-legged machine that can run with any of the common gaits we have described, there remain many interesting questions associated with starting and stopping such a machine and selecting its gait. We can easily see how to start forward if the machine is already hopping in place. What is much less obvious is how to coordinate the transition from a standing start to full-speed running. Similarly, how and when should such a machine change from one gait to another? A running horse switches its lead as it turns, that is, it changes which of the two front legs slightly precedes the other. What computations should be done to make such minor changes in the pattern of leg motion? We find these problems fascinating, both as engineering questions in the form "What should we build?" and as scientific questions in the form "How do living systems work?"

A much more difficult problem is how to choose footholds for the machine. The function of vision in walking and running by people and animals, particularly the ability to choose sensible places to put the feet, is not well understood. One can imagine avoiding this problem by having the machine run fast only over smooth ground and by having some kind of human assistance to choose a safe path. It will probably be desirable to scan the ground ahead of the vehicle for holes. Still, just as a galloping horse runs the risk of stepping into a gopher hole, so we must expect that a running machine will also get into that kind of trouble.

The mobility of off-road vehicles is limited by two factors. First, the continuous footprint of wheels and tracks prevents wheeled and tracked vehicles from making use of the discontinuous points of support that are available to a legged vehicle. We are encouraged to think that the state of the computer art is now sufficiently advanced to allow the construction of adequate control systems for legged vehicles, and so the legged alternative for high-mobility vehicles can be seriously considered. Indeed, the Defense Advanced Research Projects Agency is already sponsoring research on such vehicles, including partial support for our projects.

A second source of mobility is narrow width: a motorcycle can get into places a jeep cannot reach. Narrow legged vehicles can be built, but they will have to balance, at least in the sideways direction. Involved as we are both in the construction of a six-legged vehicle that can crawl without attention to balance and in studies of walking and running with balance, we believe the effort to understand balance is much more important. We think the experiments with six-legged crawlers now under way in our laboratory and elsewhere are mainly exercises in the control of multiple legs and are not in themselves useful; such crawlers will ultimately be replaced by machines with fewer legs that can balance. Mastery of balance will be the key to building high-mobility machines that walk and run.



MECHANICAL HORSE was the subject of a patent obtained by Lewis A. Rygg in 1893. The lower drawing is a plan view along the line *x–x* of *Fig. 1*. The stirrups doubled as pedals that were to enable the rider to power stepping motions. Steering was to have been done with reins that moved the head and forelegs from side to side. Apparently the machine was never built. It would have been similar to many modern walking toys. Since they have no sensing or computing facilities, they cannot adapt to variations in terrain. They only crawl on flat surfaces.

The Hair Cells of the Inner Ear

They are exquisitely sensitive transducers that in human beings mediate the senses of hearing and balance. A tiny force applied to the top of the cell produces an electrical signal at the bottom

by A. J. Hudspeth

The sense of hearing, the sense of balance that enables human beings to walk upright, the ability of certain animals to detect vibrations of the ground and the ability of fishes to detect the displacement of water would appear to have little to do with one another. Actually the four senses are closely related. Indeed, each of them is made possible by the same sensory receptor. The receptor is called a hair cell and is named for the hair bundle, a group of fine projections that extend from its upper surface. The hair cell is an extremely sensitive mechano-electric transducer, that is, it converts a mechanical force into an electrical signal. The mechanical force is the stimulus applied to the hair bundle; the electrical signal is the message relayed to the brain.

Each hair cell is sensitive to only a limited range of stimuli. Therefore if the organism is to obtain useful information about its environment and about its own movements the output from thousands of receptors must be combined. The requisite number of hair cells are found in several small sensory organs in the inner ear. In human beings arrays made up of thousands of hair cells with slightly different sensitivities are found in six organs in each ear. The combined responses of the cells yield information about linear acceleration in any direction, about angular acceleration along three perpendicular axes and about acoustic tones with a wide range of frequencies.

The general structure of the hair cell and the receptor's sensory role have been known for many years. How the individual hair cell works, however, is only now becoming clear. To examine events on the cellular scale I have worked with single hair cells from the inner ear of the bullfrog. A microscopic probe is employed to push the bundle back and forth as an electrode records the cell's electrical output. With this setup I have obtained the first recordings made from single hair cells during the application of a precisely controlled mechanical stimulus directly to the hair bundle.

Each hair cell can respond to remarkably small stimuli. Hair cells from mammals begin to respond when the tip of the hair bundle is moved no more than 100 picometers (trillionths of a meter); this distance is about the same as the diameter of some atoms. In addition, recordings from isolated hair cells show that each receptor has one direction of maximum sensitivity. When the hair bundle is displaced in any one direction, the cell responds only to the component of motion that is in the direction of maximum sensitivity. Along with light microscopy and electron microscopy recordings from individual hair cells are beginning to yield a rich understanding of the workings of the sensory receptors of the inner ear. Many questions nonetheless remain, among the most intriguing of which is: What is the exact molecular mechanism whereby the displacement of the hair bundle changes the electrical properties of the cell and causes a message to be sent to the brain?

The hair cell is generally either cylindrical or flask-shaped and is always found as part of an epithelium, a sheet of cells that is at most a few cells thick. It is closely related to the neuron, or nerve cell, although it does not have axons or dendrites, the fibers that extend from neurons and transmit electrical signals in the nervous system. For this reason the hair cell is sometimes referred to as a paraneuron. In the epithelium where the hair cell is found its apical, or upper, end lies flush with the upper ends of the supporting cells that surround it. The apical ends of the hair cells and those of the supporting cells form a smooth surface above which the hair bundle extends. In different cells the length of the hair bundle ranges from about three micrometers (millionths of a meter) to more than 100 micrometers.

The detailed structure of the hair bundle differs among the hair cells of different species and even among hair cells from different organs of the same individual. The basic form of the hair bundle, however, is the same in all vertebrate animals. Each bundle consists of from 30 to 150 thin, rod-shaped extensions. The extensions are of two kinds; they are both called cilia but nonetheless have quite different internal structures. All but one of the extensions are called stereocilia. Stereocilia are cylindrical or club-shaped organelles with a core made up of tightly packed filaments. The plasmalemma, the surface membrane of the cell, extends over the filamentous core of the stereocilium much as the finger of a glove covers the human finger. The diameter of the stereocilium ranges in different cells from about .2 micrometer to one micrometer.

In spite of the term, stereocilia are not true cilia. A true cilium, such as the cilium of a sperm cell, has a complex and highly differentiated central structure called the axoneme. Most true cilia are capable of an independent motion much like that of an oar in rowing; the motion originates in the axoneme. The hair bundle has one true cilium: the kinocilium. It is about .25 micrometer in diameter. Like the axoneme of other true cilia, the axoneme of the kinocilium consists of two central tubules surrounded by nine "doublet" microtubules, pairs of small tubes that share a central wall. The kinocilium is not, however, capable of independent motion in the hair bundle. At the end away from the main body of the cell it is usually attached to the adjacent stereocilia.

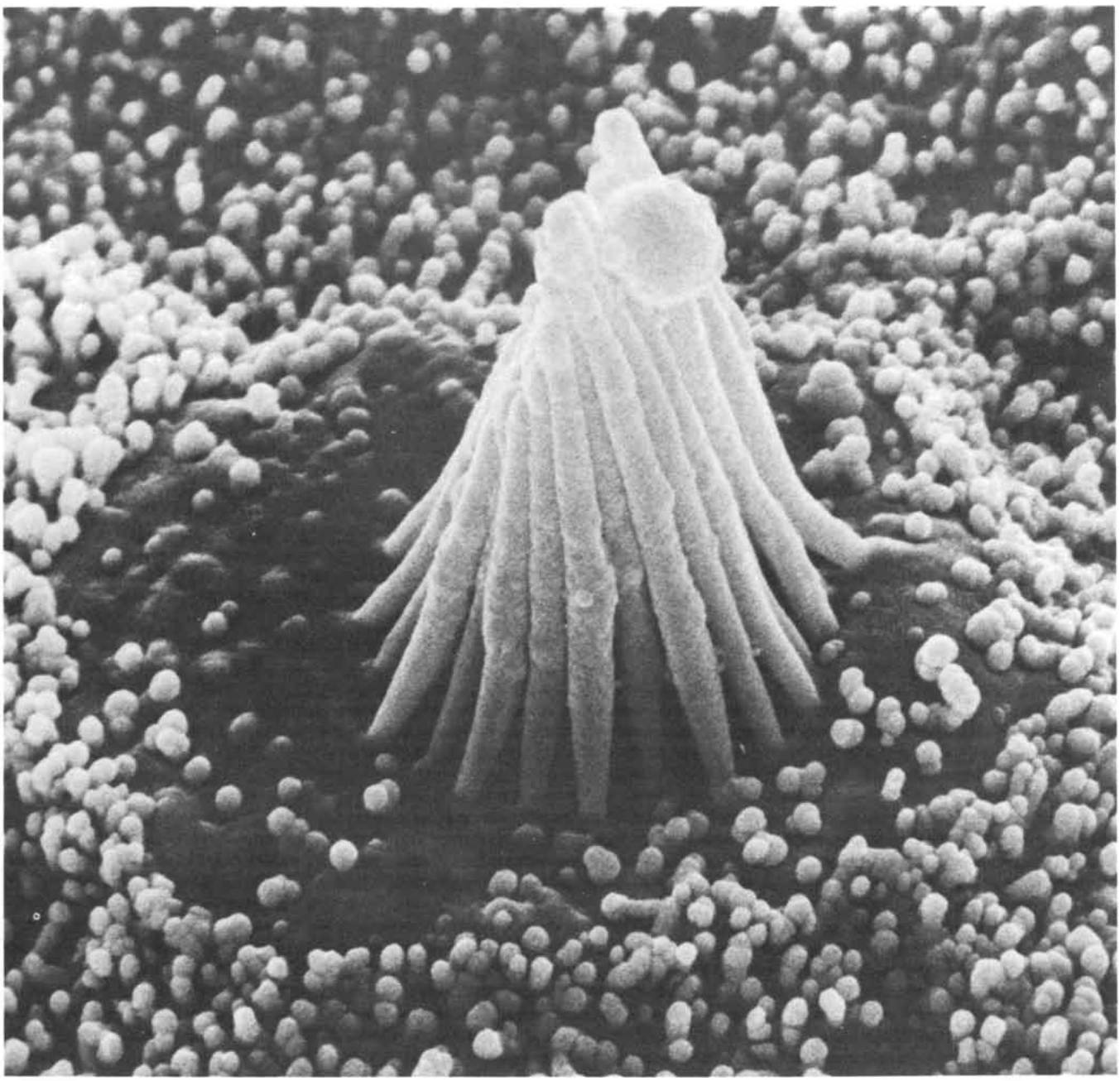
The arrangement of the cilia in the hair bundle is highly regular. The stereocilia are in a hexagonal array: each one is surrounded by six others that are equidistant from it. The stereocilia are not of equal length. The hair bundle is circular in cross section and along one diameter of the circle there is a progressive increase in the length of the stereocilia. Along any axis perpendicular to that diameter, however, the cilia are the same length. Thus the hair bundle has ranks of cilia of equal length that are arranged in order of size. As a result of this arrangement the hair bundle has a plane of bilateral symmetry. In the intact cell the stereocilia are not uniform-

ly perpendicular to the apical surface. Instead they lean together to form an obliquely truncated cone, with the kinocilium lying on the edge of the cone in the plane of bilateral symmetry. Near their tops the cilia approach one another closely or indeed touch.

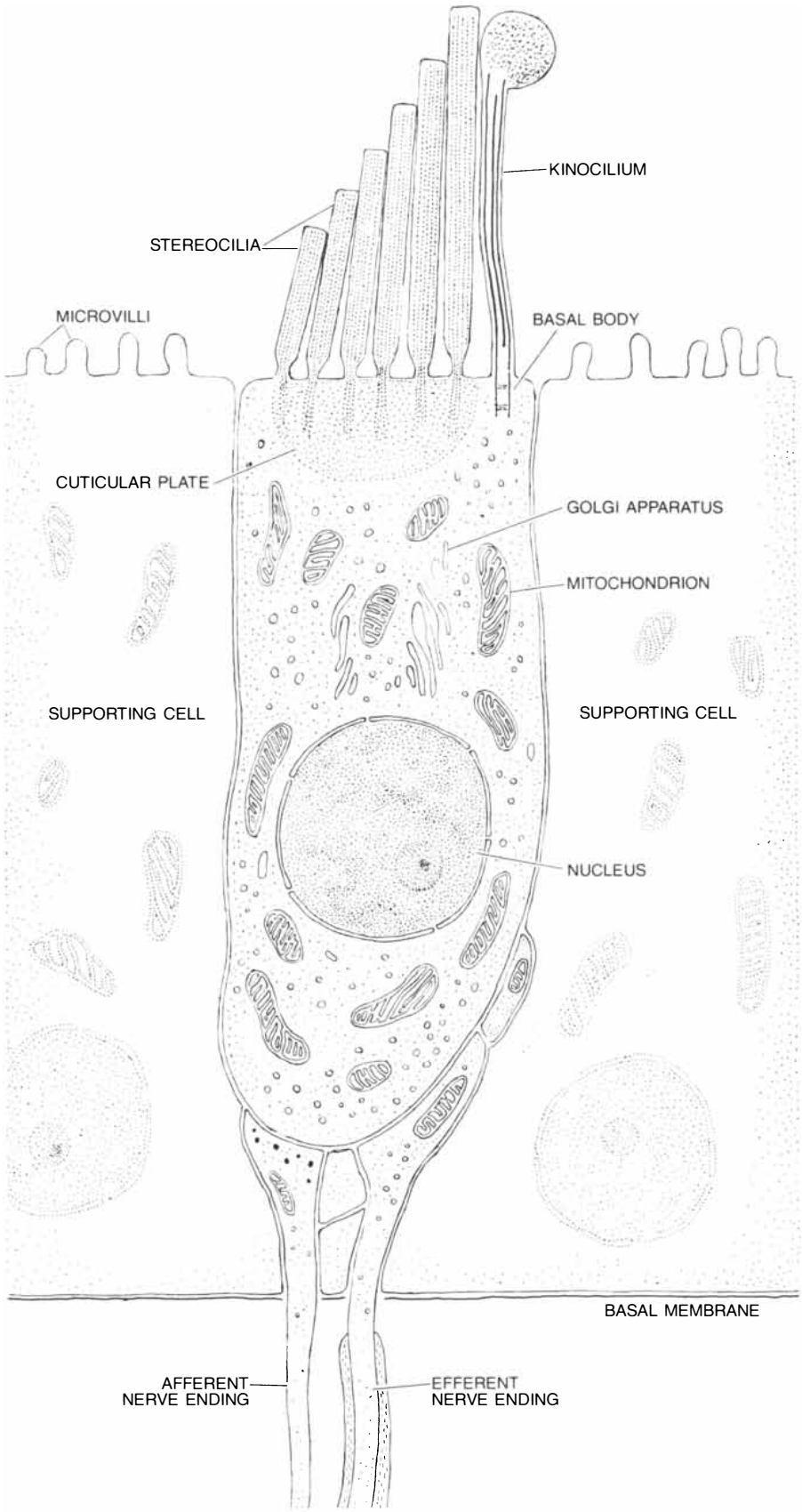
In the growth of a human embryo the cells from which the inner ears develop first become recognizable as two placodes, or thickened patches of cells. The placodes are on the surface of the embryo on each side of the developing brain stem. As they increase in size they buckle inward and ultimately lose their connection with the surface of the embryo. Once they are separated from the surface they are known as otic vesicles. Each otic vesicle is a minute pouch of epithelial tissue filled with fluid.

The otic vesicle undergoes a complex development that includes growth, bending and the fusion of its parts. The result of the process is the formation of the six receptor organs in which hair cells are found. The three semicircular canals, the receptor organs for angular acceleration, are toroidal (doughnut-shaped) tubes. They are joined together in such a way that when the head is in its usual orientation one canal lies in the *x* plane of three-dimensional space, one in the *y* plane and one in the *z* plane. The utricle and the saccule, the organs that are sensitive to linear acceleration, are small, irregular pouches. The cochlea, the organ of hearing, has a spiral form.

The fully formed inner ear has a com-



HAIR BUNDLE is the structure for which the hair cell is named; it is responsible for the cell's mechanical sensitivity. The bundle is enlarged 14,000 diameters in this scanning electron micrograph of a hair cell from the inner ear of a bullfrog, made by Richard A. Jacobs of the California Institute of Technology. Hair cells are found in epithelia, or sheets of tissue a few cells thick, where they are surrounded by supporting cells. The hair bundle shown here consists of some 50 fine fibrous extensions. All but one of the extensions are stereocilia, which in spite of their name are not true cilia. The hair bundle also has one true cilium, the kinocilium, which ends in the bulbous swelling at the upper-right edge of the bundle. The structure is shown in the resting position, where it forms an upright, obliquely truncated cone. Bending the hair bundle away from the resting position causes the hair cell to relay an electrical signal to the brain. The many small light-colored projections around the hair bundle are microvilli extending from the apical surfaces of the nearby supporting cells.



HAIR CELL is shown in cross section. The cell is generally either cylindrical or flask-shaped. Although they are closely related to nerve cells, hair cells have no dendrites or axons. The organization of the hair bundle varies among species in its details but its overall structure is the same in all vertebrate organisms. Along one axis there is a progressive increase in the length of the stereocilia. Along the perpendicular axes, however, the stereocilia have the same length. As a result the hair bundle has a plane of bilateral symmetry; the illustration is a section through this plane. The kinocilium lies in the plane of symmetry at the tall edge of the cone. The afferent nerve ending at the basal surface of the cell transmits the receptor's response to the brain.

plex geometry that arises from the bending, twisting and fusion of one otic vesicle. For this reason the structure is referred to as the labyrinth. In spite of the complex form of the labyrinth, each of the six sensory organs has a fundamental simplicity of design that is derived from the form of the otic vesicle. Each organ consists of a continuous epithelial sheet supported by connective tissue. The epithelium forms a closed space filled with the fluid called endolymph. The space outside the epithelium is filled with ordinary extracellular fluid. Where substantial volumes of the extracellular fluid accumulate in the inner ear the fluid is termed perilymph. The entire structure is encased in cartilage and bone.

Thus the epithelium separates two different fluids. Embedded in the epithelium, the hair cell is in contact with both of them. The apical surface of the cell is in contact with the endolymph in the interior of the organ. The basal, or bottom, surface of the cell is in contact with the perilymph that surrounds the organ. As we shall see, these fluid compartments have an important role in the function of the hair cell.

In order to appreciate the action of the hair cell it is necessary to consider the mechanical operation of each of the organs of the labyrinth. The utricle and the saccule are positioned so that the ensemble of four organs in the two ears provides sensitivity to linear acceleration along any axis. Even more important is the fact that the paired utricles and saccules constantly monitor the acceleration due to gravity; they are crucially important in the control of posture, gait and equilibrium. Although the directional sensitivity of the utricle and saccule are quite complex, as an approximation it is reasonable to consider that the utricle responds to horizontal acceleration and the saccule responds to vertical acceleration.

The hair cells of the utricle and the saccule are found in each organ in a crescent-shaped patch of epithelium about a millimeter across. The sheet is more or less flat and includes a few thousand hair cells. In the saccule the epithelial sheet is mainly vertical and in the utricle it is mainly horizontal. A structure called the otolithic membrane, which is made up of a feltwork of protein molecules, lies parallel to the epithelium and quite close to it. The kinocilium of each hair cell is inserted into an indentation in the otolithic membrane. Embedded in the otolithic membrane and piled on top of it are hundreds of thousands of the tiny crystals called otoconia. Each otoconium is composed of calcium carbonate, usually in the form of calcite, laid down on an organic skeleton.

The density of the otoconia is several times greater than that of the en-

dolymph filling the utricle and saccule. Therefore when the organs are accelerated, the greater inertia of the otoconia causes them to lag behind the endolymph. The lagging motion of the crystals is communicated to the hair bundle by the otolithic membrane. The stimulus that actually reaches the hair cell is a force displacing the tip of the hair bundle in the direction opposite to that of the acceleration.

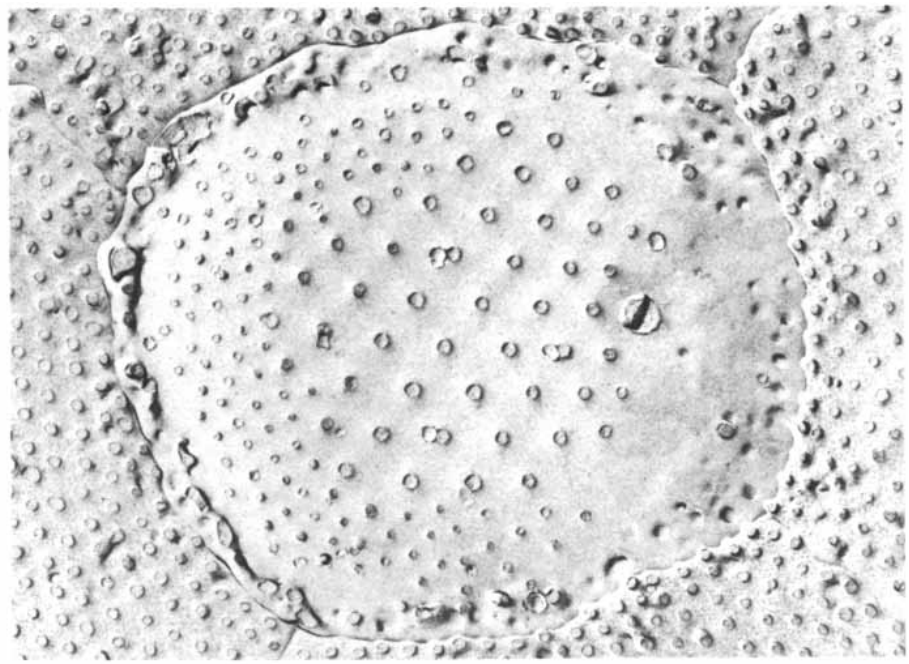
In the semicircular canals a similar mechanical force is transmitted to the tip of the hair bundle in a different way. Each semicircular canal is a tube bent in the form of a circle with a diameter of about six millimeters. In one region the cross-sectional area of the tube is larger than it is in the rest of the canal; the enlarged region is called the ampulla. In the ampulla a septum, or sheet of tissue, crosses the tube. The septum consists of a ridge of cells called the crista with a sheet of extracellular material, the cupula, extending from it. The hair cells lie in the crista. The hair bundles stretch upward and are inserted into the gelatinous cupula.

When the semicircular canal undergoes an angular acceleration (for example when the head is rotated quickly), the inertia of the endolymph in the canal causes the fluid to lag behind the walls of the canal, which move with the head. The fluid presses on the cupula, which is soft and easily distorted; the deflection of the cupula causes the hair bundles inserted into it to bend at their site of attachment to the hair cells.

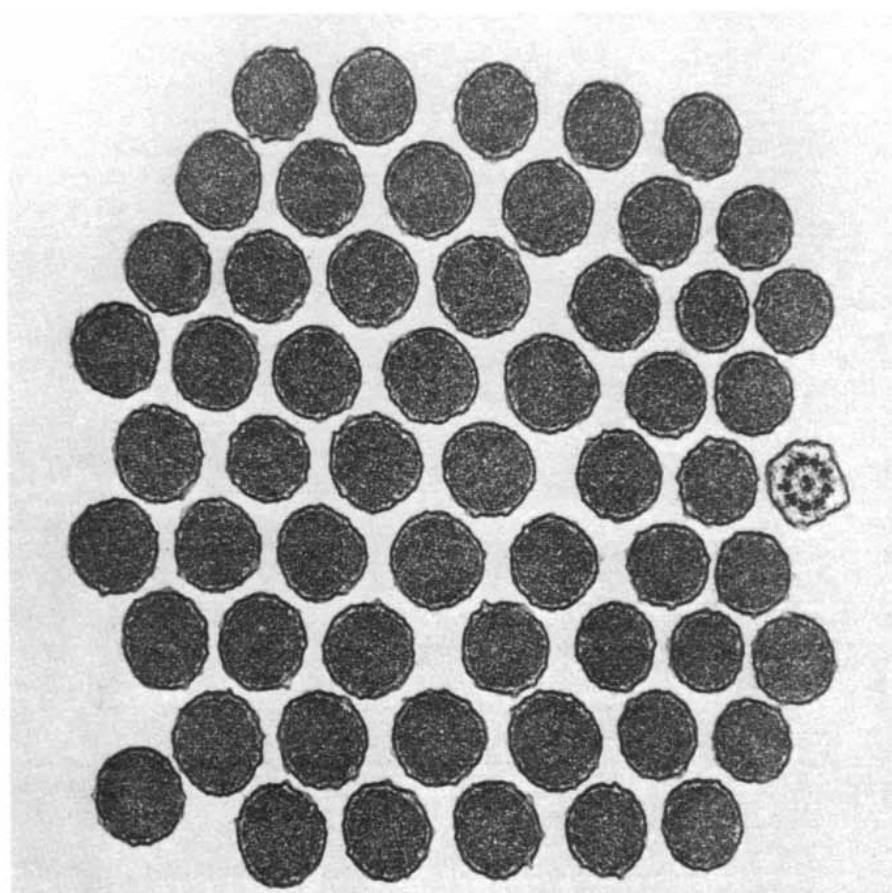
The cochlea, the sixth organ of the inner ear, has the form of a tube twisted into a spiral. The 15,000 or so hair cells in the cochlea of the human ear are sensitive to sounds with frequencies ranging from a few tens of hertz (cycles per second) to about 20 kilohertz. The hair cells are in an orderly array resting on the thin partition known as the basilar membrane. The basilar membrane winds within the cochlea and follows its spiral form. In the cochlea natural sounds with many frequencies are broken down into constituent frequencies; each constituent excites a small fraction of the 15,000 hair cells at one position along the basilar membrane.

The operation of the cochlea is too complex to be described in detail here. Moreover, its mechanism is not fully understood and is the subject of considerable controversy. As in the utricle, the saccule and the semicircular canals, however, the stimulus that reaches the hair cell is thought to be a displacement of the hair bundle. Here it will be sufficient to consider how the oscillatory pressure differences of the sound waves outside the ear are converted into a force applied to the hair bundle.

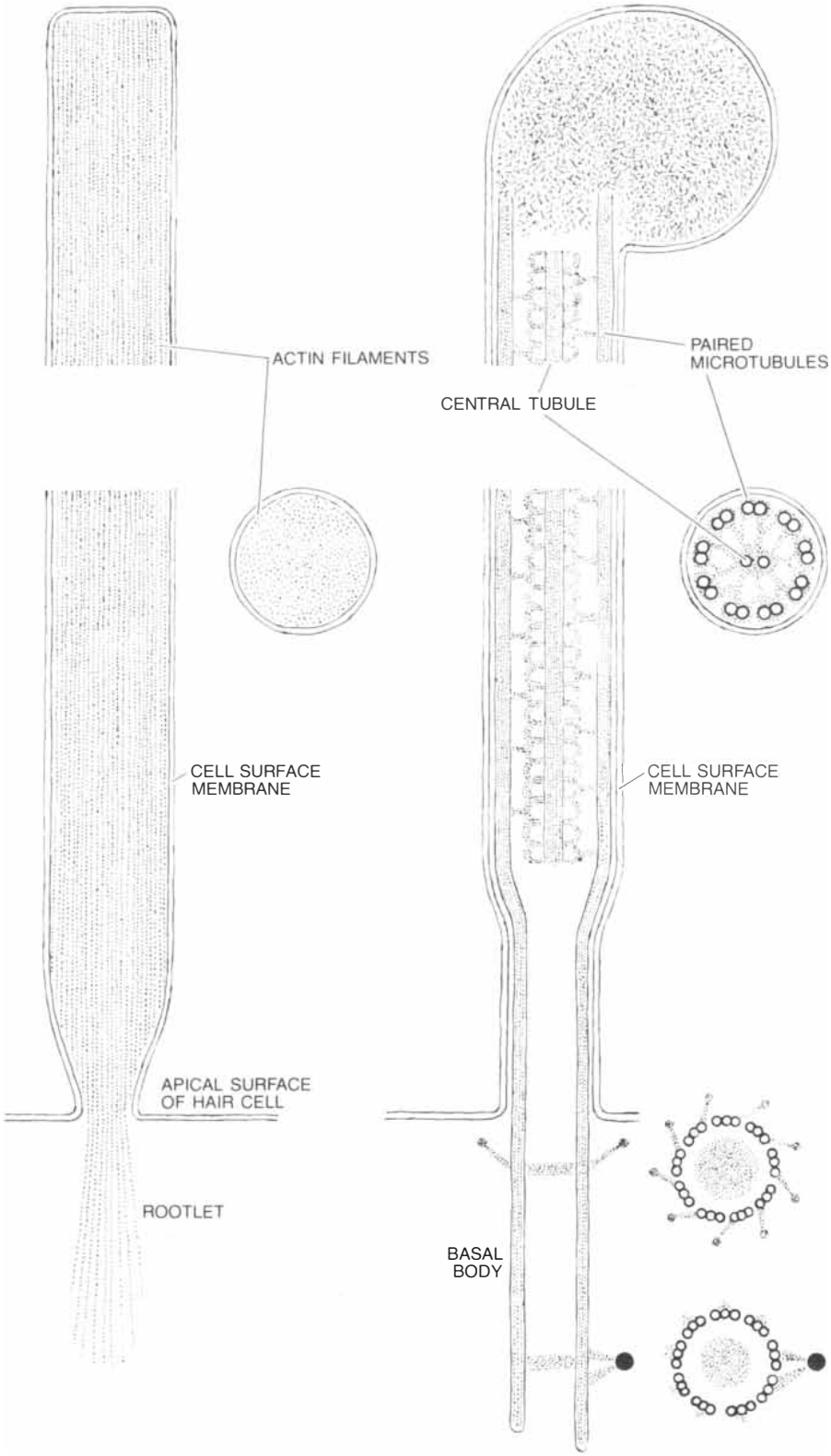
Sound waves arriving at the outer ear impinge on the membranous tympanum: the eardrum. The movement



HEXAGONAL ARRAY of stereocilia is shown in an electron micrograph of a hair cell from the inner ear of a bullfrog. The cell is enlarged 16,000 times. The micrograph was made by freezing the epithelium and shearing off everything protruding above the apical surface. The pits on the hair cell are points of insertion of the stereocilia: each one is surrounded by six others that are equidistant from it. The large broken stub at right is the stub of the kinocilium.



INTERNAL STRUCTURE of the kinocilium is quite different from that of the stereocilia. The cores of both kinds of cilia are shown enlarged 40,000 times. The core of the stereocilium is a bundle of threadlike filaments. The kinocilium is at the right. Its core is a complex structure known as the axoneme. The transmission electron micrograph was made by treating the cilia with preservatives that are better taken up by filaments and tubules than by other structures.



STEREOCILIUM AND KINOCILIUM are shown in vertical and horizontal sections. The stereocilium is at the left and the kinocilium is at the right. Each stereocilium is shaped like a pencil or a baseball bat. Its diameter ranges from .2 micrometer (millionth of a meter) to one micrometer in different cells. The rootlet of the stereocilium penetrates the apical surface of the hair cell. The taper at the lower end of the cilium allows it to pivot. The parallel filaments in the core of the stereocilium consist of the protein actin; these filaments are highly cross-linked and make the cilium rigid. The axoneme of the kinocilium, like the axonemes of other true cilia, consists of two central tubules surrounded by nine pairs of microtubules that share a central wall. The axoneme appears most clearly in the horizontal section at the upper right. The vertical section through the kinocilium is through one of the central tubules and the other does not appear. The surface membrane of the hair cell covers both kinds of cilia completely.

of the tympanum is communicated to three small bones in the middle ear: the malleus, the incus and the stapes (the hammer, anvil and stirrup). The stapes, which is the closest to the inner ear, has a pistonlike action that causes the basilar membrane to vibrate. The hair cells are in the organ of Corti, the auditory organ, which lies on the basilar membrane. The hair bundles are inserted into a structure called the tectorial membrane that is placed parallel to the basilar membrane and close to it. The basilar membrane and the tectorial membrane are hinged along different axes; therefore the upward displacement of the membranes caused by the action of the stapes is accompanied by shearing motions between them. These motions displace the hair bundles extending between the two membranes.

As the simplified account presented above indicates, in the functioning of all the sensory organs of the inner ear an initial stimulus (an acceleration or a sound) is transformed into a force on the hair bundle. Since the output of the hair cell is an electrical signal that is sent to the brain along a nerve fiber, the crux of understanding the function of the hair cell is finding out how the displacement of the hair bundle is turned into an electrical response. A complete understanding of the process will undoubtedly require work on the intact inner ear of mammals. Several difficulties, however, stand in the way of relying completely on hair cells from intact mammalian sensory organs. The temporal bone in which the mammalian labyrinth is embedded is extremely hard; mammalian hair cells are particularly sensitive to interruptions of their blood supply; mammalian hair cells are small and difficult to investigate with microelectrodes. Moreover, it is difficult to ascertain the stimulus that reaches the hair bundle and affects it directly when natural stimuli are presented to the ear.

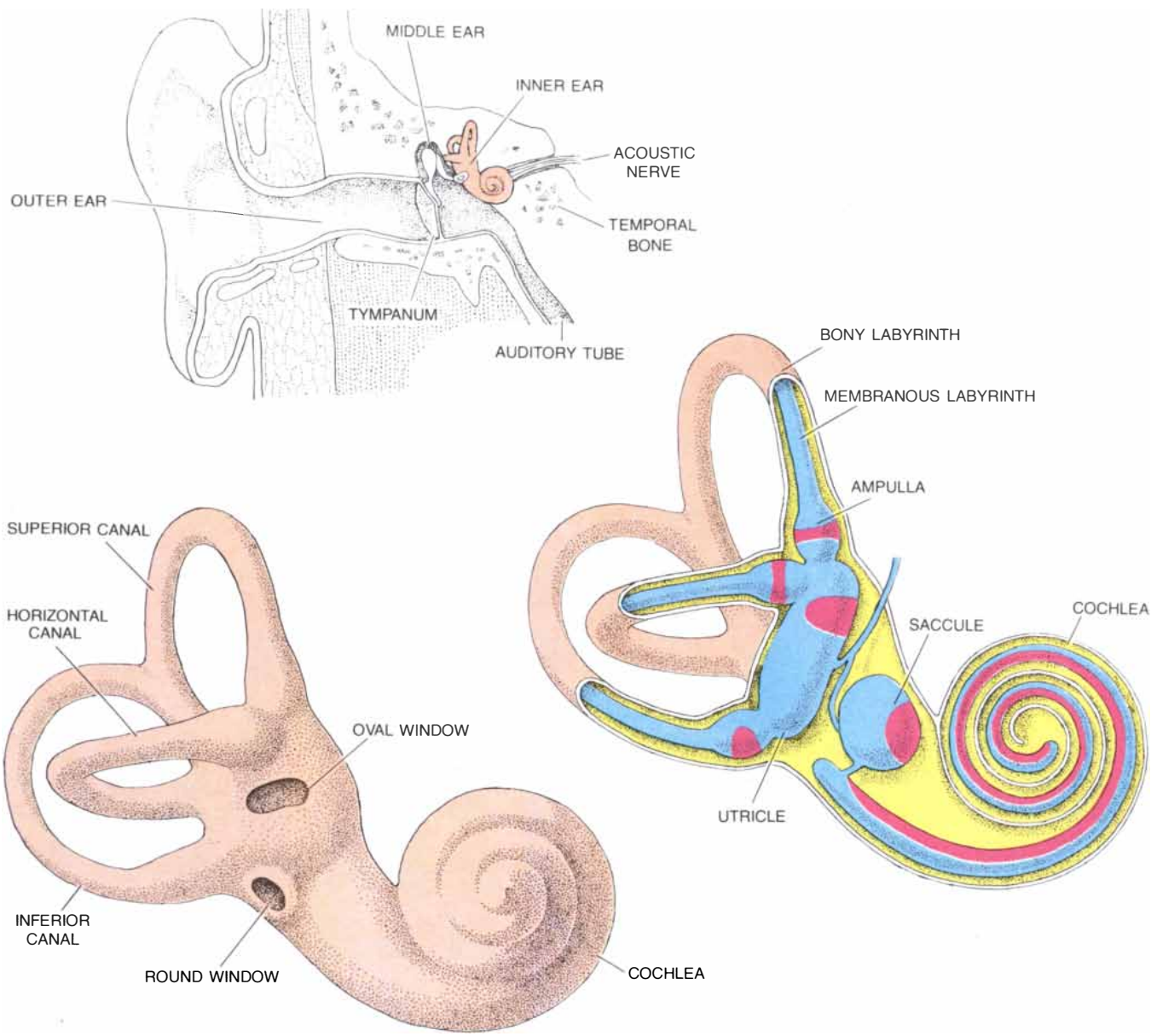
For these reasons preparations of hair cells from lower animals such as fishes, frogs, lizards and turtles have often been exploited in the study of the inner ear. The hardy and relatively large hair cells of the amphibian inner ear make it possible to carry out procedures that cannot yet be done with mammalian hair cells. In much of my own work I have employed hair cells from the inner ear of the American bullfrog (*Rana catesbeiana*). After the saccule is removed from the frog and the otolithic membrane is stripped away the epithelial layer that includes the hair cells is revealed.

The epithelial preparation is put into a physiological solution (a fluid much like the internal fluids of the living organism) and placed under the microscope. A microelectrode is inserted through the apical surface of a single hair cell and into its interior. A micro-

probe is inserted between the kinocilium and the adjacent stereocilia. The probe serves to push the hair bundle back and forth. The motion of the probe can be controlled by a manual "joystick" apparatus or by a computer; in either case the hair bundle can be displaced with high precision. The displacement can also be observed through the microscope. The microelectrode records the electric potential in the interior of the cell and the changes in potential that accompany the movements of the hair bundle.

In order to understand the electrical changes in the hair cell caused by the movements of the hair bundle it is necessary to have some appreciation of the electrical properties of the cell in its unperturbed state. Like the outer membranes of other cells, the plasmalemma of the hair cell is selectively permeable: it allows some molecules to enter the cell and keeps others out. Furthermore, the permeability of the membrane can change through the opening and closing of selective channels. When the channels open, certain ions (atoms or molecules that have lost or gained electrons and therefore have a net electric charge) enter the cell. Through the opening and closing of these channels the cell can regulate its electric potential with respect to the environment.

Molecular pumps in the cell membrane maintain an internal concentration of ions that is different from the concentration in the surrounding fluids. The most intensively investigated of these pumps is the one that reduces the concentration of sodium ions within the cell. Sodium ions have a single unit of positive electric charge. Because the concentration of sodium is lower inside the cell than it is outside, the interior of the cell has a negative potential with re-



HUMAN INNER EAR is the site of six small sensory organs that contain the hair cells. The semicircular canals measure angular acceleration; the utricle and the saccule detect linear acceleration; the cochlea is the organ of hearing. The upper panel is a section through the temporal bone, in which the inner ear is embedded. Because of its shape the inner ear is often referred to as the labyrinth. Sounds striking the tympanum, or eardrum, are communicated to the cochlea by three small bones in the middle ear: the malleus, the incus and the stapes. The panel at the lower left shows the labyrinth as it would appear if it were removed from the temporal bone with a thin layer of bone around it; this is the bony labyrinth. The three doughnut-shaped tubes at the left are the semicircular canals. The spiral structure at the right is the cochlea. The panel at the lower right shows the bony labyrinth opened to reveal the membranous labyrinth, which is composed of the sensory organs. Each organ is a pouch of epithelium. The pouch is filled with the fluid called endolymph (*blue*). Outside the epithelial pouch, between it and the bone, is the fluid called perilymph (*yellow*). The hair cells are found in small patches (*red*). In the semicircular canals the patches are in the region called the ampulla. In the cochlea they are on a membrane that follows the shape of the organ.

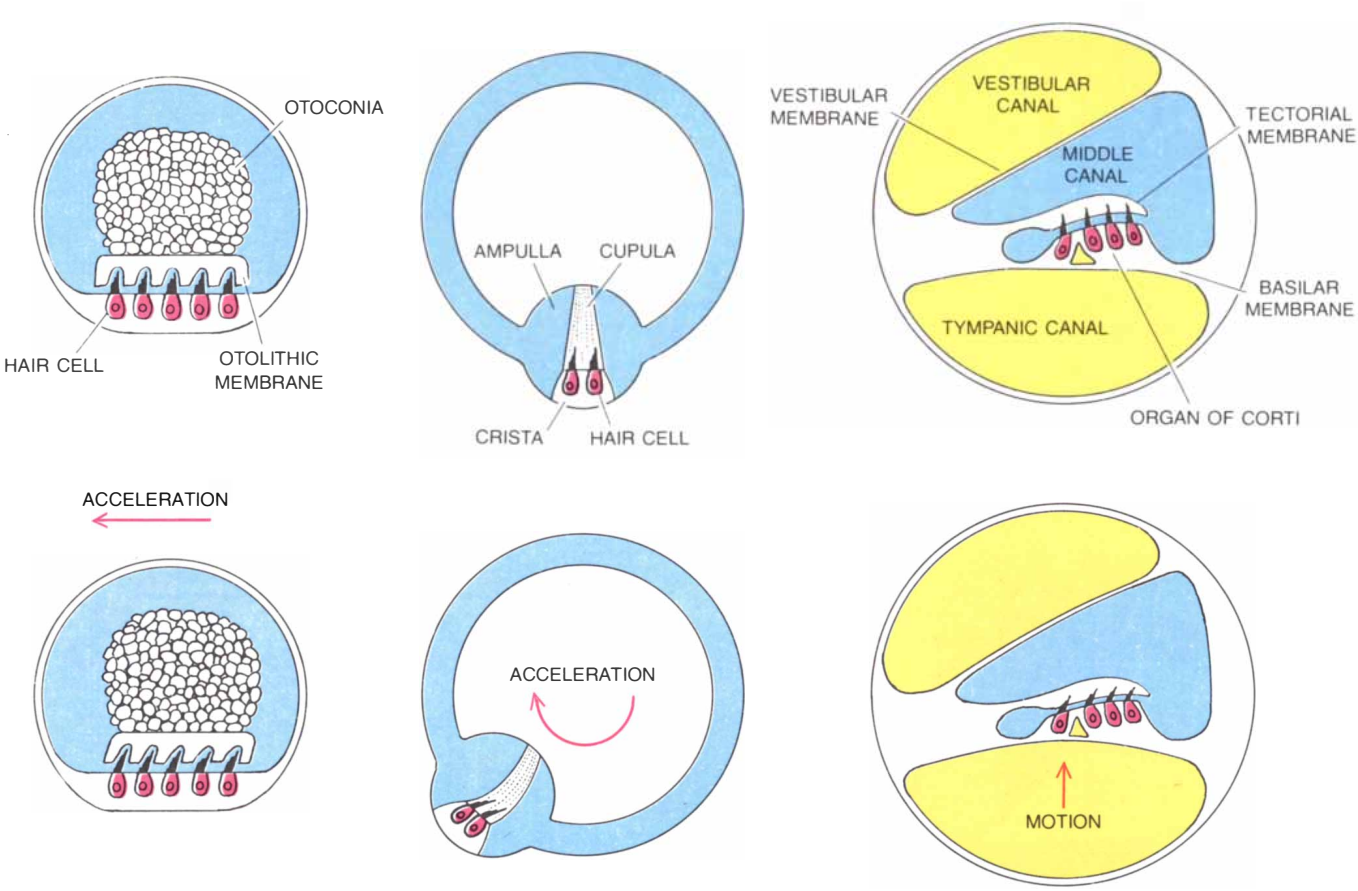
spect to the environment. The difference in potential between the cell and the surrounding fluid is referred to as the membrane potential because it represents the drop in potential across the membrane. It should be noted that if a cell has a negative potential, the interior of the cell attracts positively charged ions. When the channels for such ions open, the ions will rush into the cell, reducing the membrane potential.

The membrane of the cell therefore acts as a capacitor: a thin plate separating electric charges with opposite signs. One property of a capacitor is that it can yield a potential difference by separating a very few charges. Moreover, the passage of only a few charged bodies from one side of the plate to the other can considerably change the potential drop across the capacitor.

The electrical properties of most neurons are regulated by contact with one external fluid. The cell membrane in such cells acts as a single capacitor. In the hair cell the situation is more complex. As we have seen, the hair cell separates the endolymph and the perilymph. The endolymph is in contact with the apical surface of the cell, from which the hair bundle extends. The concentration of positively and negatively charged ions in the endolymph is similar to that in the interior of the cell: high in potassium ions and low in sodium ions. Therefore if the cell were in contact only with endolymph, there would be no potential drop across the membrane.

The basal surface of the cell is in contact with the perilymph. Perilymph has a composition quite different from that of the interior of the cell. In particular, the perilymph has a much higher concentration of sodium ions and a lower concentration of potassium ions than the inside of the cell. If the cell were in contact only with the perilymph, the difference in potential between the interior and the exterior of the cell would be substantial. Indeed, the inside of the cell would have a potential about 90 millivolts lower than that of the surrounding fluid.

Both the apical and the basal surfaces of the hair cell act as capacitors. The membrane potential of the cell as a whole is therefore influenced by contributions from both the apical and the basal areas of the membrane. The contributions are not, however, equal. Both the apical and the basal areas have channels for ions. When the membrane is in the resting, or unperturbed, state, a few channels are always open and there is a small flow of positively charged ions across it. The contribution of the apical and basal areas to the overall membrane potential is determined by the permeability of each area, that is, the number of channels that are open. When the membrane is in the resting state, the permeability of the basal area is substantially greater than that of the apical area. For this reason the membrane potential of the hair cell in the resting state is closer to −90 millivolts than to zero millivolts: the potential is about −60 millivolts.



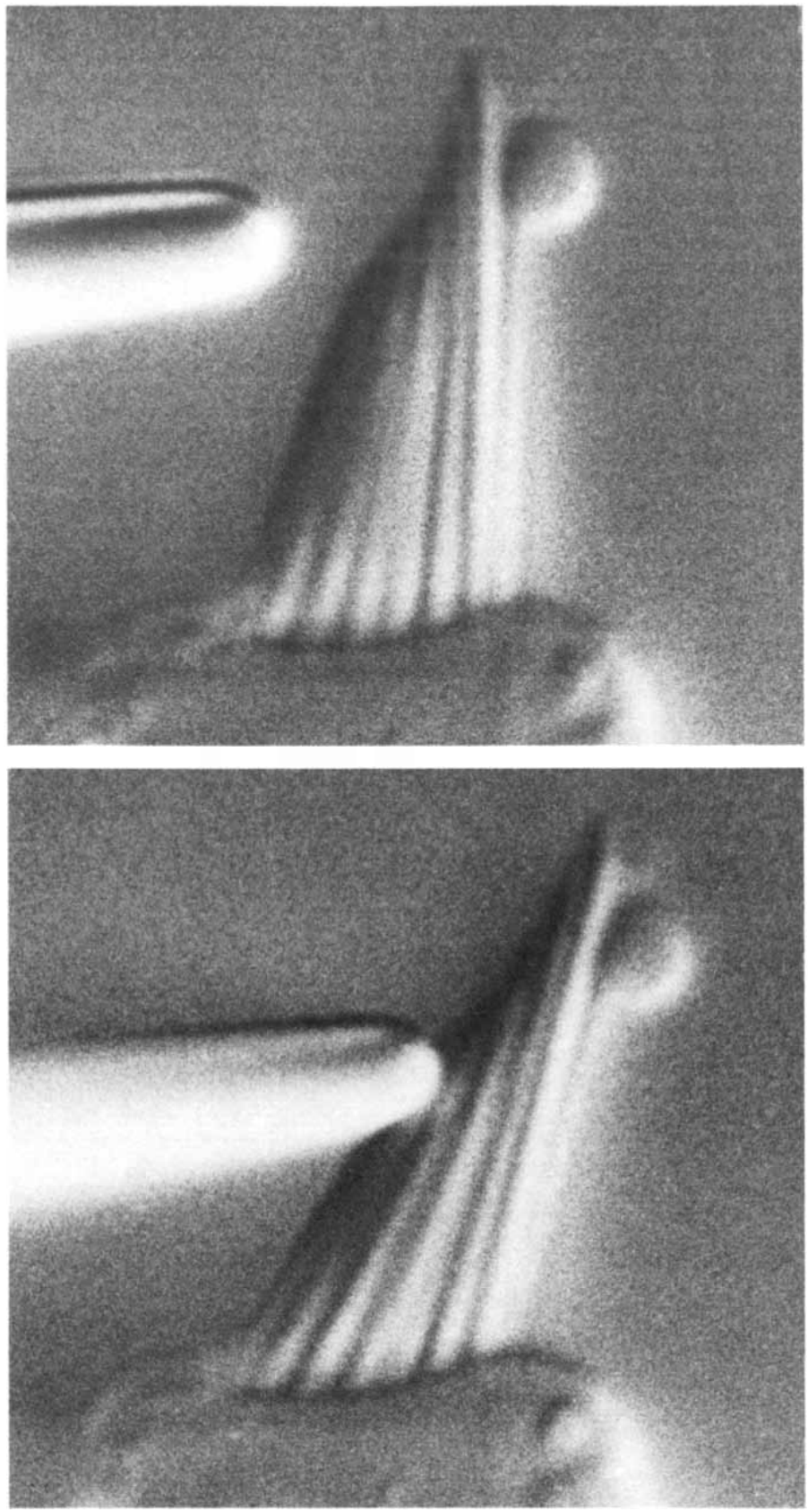
DISPLACEMENT OF THE HAIR BUNDLE is the stimulus that reaches the hair cell in all three types of sensory organ in the inner ear. The three panels of the illustration show how the hair bundle is displaced in the utricle and saccule (*left*), the semicircular canals (*middle*) and the cochlea (*right*). In the utricle and saccule the hair bundles are inserted into a feltwork of protein molecules called the otolithic membrane. Piled on the membrane are the crystals called otoconia. The otoconia are much denser than the endolymph that fills the organ. Therefore when the head is accelerated, the crystals lag behind the fluid. The lagging motion is communicated to the hair bundle by the otolithic membrane. The stimulus applied to the hair cell is thus a displacement of the hair bundle in the direction opposite to that of the acceleration. In the semicircular canals the hair bundles are on a protuberance called the crista and are inserted into a gelatinous sheet, the cupula, that extends across the ampulla. When the head is rotated, the endolymph in the canal lags behind the epithelium of the canal itself. The lagging fluid presses on the cupula and bends it, thereby also bending the hair bundle. In the cochlea the hair cells are on the basilar membrane; the hair bundles are inserted into the tectorial membrane. When the stapes is pulled, both membranes move upward. The membranes are hinged on different axes, so that the movement of the stapes causes a shearing motion that bends the hair bundles.

Therefore when the recording electrode is inserted into a hair cell from the saccule of the bullfrog a measurement of −60 millivolts is obtained. When the hair bundle is pushed with the probe along the axis of bilateral symmetry in the direction of the kinocilium, however, the potential difference decreases to −40 millivolts. The reduction in the potential drop across the membrane is called a depolarization. If the hair bundle is pushed along the same axis in the opposite direction, a quite different change in the membrane potential is observed: it increases from −60 millivolts to −65 millivolts. Such an increase is called a hyperpolarization. If the hair bundle is pushed in either direction along the axis perpendicular to that of bilateral symmetry, there is no change in the membrane potential.

Thus the hair cell is quite responsive to displacements along the axis of bilateral symmetry and is completely unresponsive to displacements along the perpendicular axis. Pushing the hair bundle back and forth along intermediate axes shows that the cell is highly directional in its sensitivity (confirming an inference drawn from extracellular recordings by Otto Lowenstein of the University of Birmingham and Jan Wersäll and Åke Flock of the Karolinska Hospital in Stockholm). If the hair bundle is displaced along intermediate axes, the closer the axis of displacement is to the axis of bilateral symmetry, the larger the cell's response is. Somehow the hair cell is capable of resolving any stimulus into two components: one along the axis of bilateral symmetry and the other along the perpendicular. To the first component the cell responds with a depolarization or a hyperpolarization, depending on the direction of the displacement. To the second component there is no response.

The directional sensitivity of the hair cell is of considerable sensory importance. It implies that whereas a single receptor can provide information about the magnitude of only one component of a stimulus, a group of hair cells with the appropriate sensitivities can provide accurate measurements of an acceleration in any direction. Although the usefulness of the directionality is clear, its mechanism is not known.

Indeed, the phenomenon is a puzzling one in view of the microanatomy of the hair bundle. Individual stereocilia are symmetrical about their long axis; it would appear that bending a cilium in any direction should yield the same result. The kinocilium has a plane of bilateral symmetry, which is the same as the plane of symmetry of the hair bundle. The kinocilium, however, is not responsible for the cell's directional sensitivity. The kinocilium disappears spontaneously from the hair cells of the mamma-



GLASS MICROPROBE is employed to displace the hair bundle in the experimental technique used by the author. The upper micrograph shows a single unperturbed hair cell from the saccule of the American bullfrog (*Rana catesbeiana*). In the lower micrograph the probe has been utilized to displace the hair bundle. The displacement of the hair bundle causes changes in the electrical properties of the hair cell, which in the author's work were measured by means of a small electrode inserted through the upper surface of the cell; the electrode is not shown. These light micrographs were made by differential-interference-contrast microscopy, which exploits differences in refractive index in the parts of the cell to yield bright and dark regions.

lian cochlea in the course of embryonic development. Furthermore, the kinocilium can be removed from the hair cells of the bullfrog's saccule without altering the directionality of the cells' response.

There are three hypotheses that could account for the directional response of the hair cell. The stereocilium could have an asymmetry in its cross section that is not apparent with current techniques of electron microscopy. As an alternative the complex geometric arrangement of the cilia in the hair bundle could somehow confer directionality on an ensemble of structures that do not have such directionality individually. It is also possible that the directional response is a property of the apical surface of the cell and not of the hair bundle.

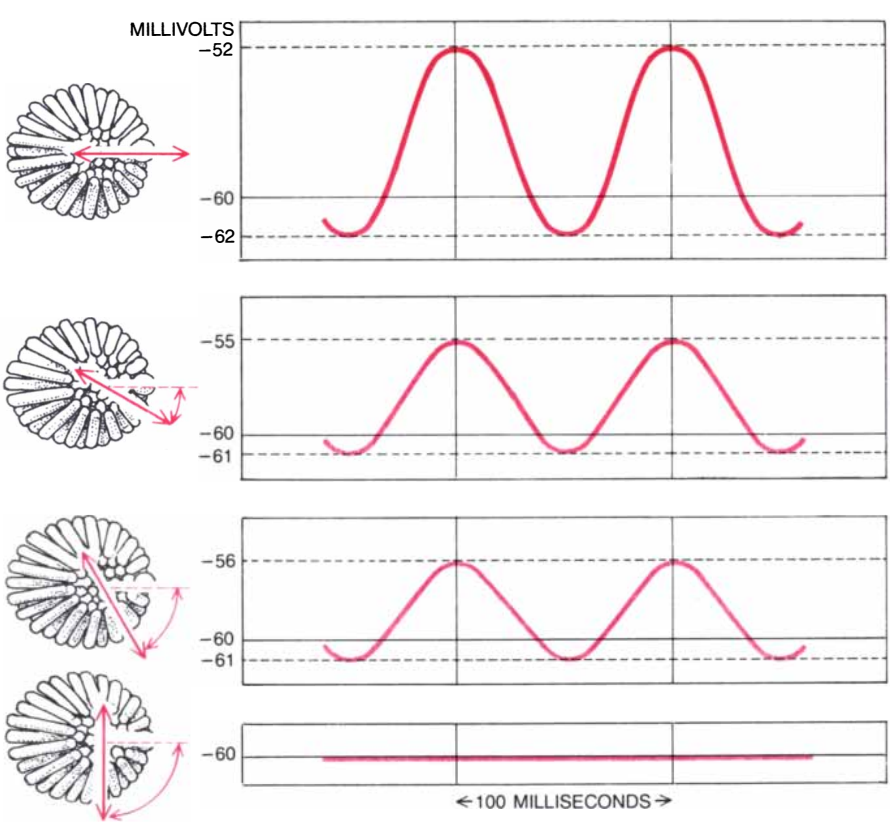
That the anatomical basis of the directionality is not known has not prevented attempts to understand how the complex set of electrical changes is induced in the individual hair cell. As I have noted, the resting membrane potential of −60 millivolts is determined by the ratio of the permeability in the apical area of the membrane to the permeability in the basal area. This suggests that the membrane potential could readily be changed by altering the permeability of the upper or lower area of the membrane. Indeed, both hyperpolarization and depolarization could be accomplished by altering the permeability of one area. Increasing the permeability of that area of the membrane could raise the ratio and decreasing it could lower the ratio, with no change in the permeability of the other area.

My results and those of other workers suggest that in principle this is how the membrane potential of the hair cell is changed. As we have seen, some channels for ions are always open, and more channels are open in the basal area of the membrane than are open in the apical area. When the tip of the hair bundle is moved on the axis of bilateral symmetry toward the kinocilium, additional channels in the apical area open. Positively charged ions, attracted by the negative potential in the cell, rush across the membrane. The increased permeability in the apical area increases its contribution to the overall membrane potential and the potential falls to −40 millivolts. If the hair bundle is pushed on the same axis away from the kinocilium, channels in the apical area of the membrane that are open in the resting state quickly close. The contribution of the apical area to the membrane potential decreases and the potential difference increases to −65 millivolts.

The "transduction channels" in the apical area of the membrane that are responsible for the change in potential do not appear to be highly selective. When the channels are open, they allow most positively charged substances .6 nanometer (billionth of a meter) or less across to enter the cell. The main positively charged ion in the endolymph, however, is potassium, and it seems that potassium is the main carrier of electric charge across the membrane. Experiments in which the potassium in the endolymph was replaced with other positively charged ions show that although other substances can carry the current, potassium probably carries most of it in the intact hair cell.

Since the molecules that enable current to flow across the membrane have not been conclusively identified, it is difficult to specify the location of the transduction channels with precision. There is, however, some indirect evidence for their location. When the transduction channels open, small electric currents flow though the fluid around the cell as well as across the membrane. By means of a recording electrode the current at various points around the perimeter of the hair bundle can be measured as the cilia are being moved. The current in the surrounding fluid is greatest at the top of the hair bundle, near the tips of the stereocilia. This suggests that when the hair bundle is deflected, channels open not at the base of the cilia where they pivot but near their tips, perhaps where adjacent cilia touch each other. Such evidence is far from conclusive and additional work is needed before the hypothesis can be accepted.

Electrical recordings from hair cells show that the initial response of the receptor is extremely fast. The fact that the auditory system of vertebrates such as bats and whales can detect tones with frequencies as high as 100 kilohertz implies the transduction apparatus can also respond very fast. Recordings from isolated inner ears confirm this implication. The hair cells of the bullfrog's ear normally respond to sounds with frequencies no higher than three kilohertz, but these cells begin to respond within a few tens of microseconds (millionths of a second) after the stimulus is applied. Hence the opening of the transduction channels, which initiates the cell's response, takes roughly the same time as a molecular event such as the action of an



DIRECTIONAL SENSITIVITY of the hair cell is shown in traces representing recordings made with an electrode inserted into a single receptor from the saccule of the bullfrog while its hair bundle was being moved. The figures are based on experimental records made by Sandra L. Shotwell in the author's laboratory. When the hair cell is in its resting state, its interior has an electric potential 60 millivolts lower than that of the surrounding fluid. If the hair bundle is displaced toward the kinocilium along the axis of bilateral symmetry, the potential difference decreases to −52 millivolts (*top*). If the bundle is displaced away from the kinocilium along the same axis, the difference increases to −62 millivolts. As the direction of displacement diverges from the axis of symmetry the size of the response decreases (*middle panels*). The cell does not respond to displacements along the axis perpendicular to that of symmetry (*bottom*).

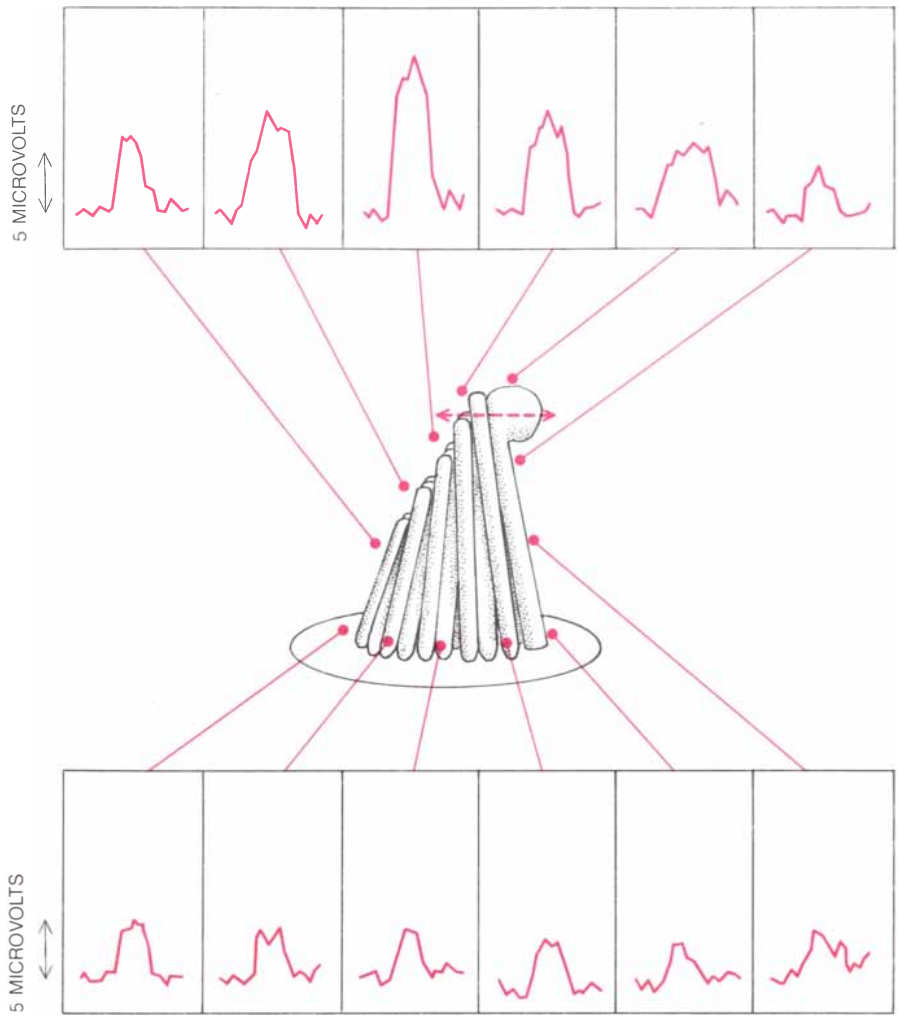
enzyme. It is unlikely that many processes intervene between the application of the force to the hair bundle and the initiation of the cell's response; it is as if the mechanical force were acting directly on the transduction channel.

A very small displacement of the hair bundle is sufficient to cause the hair cell to respond. The hair cells of lizards and turtles begin to respond when the bundle is moved a distance estimated to be as little as 340 picometers. In the mammalian cochlea the estimated threshold is even lower: about 100 picometers. Recordings of hair cells from the saccule of the bullfrog show that the entire range of the cell's response comes over a deflection of about .5 micrometer. Since the hair bundle in the bullfrog's saccule is about 8.2 micrometers long, the entire response comes over an angular deflection of only three degrees.

It seems likely that a large continuous stimulus (such as one that would result from a rapid acceleration) could deflect the stereocilia much more than three degrees. The cell might then have a high constant output. Subsequent small changes in the stimulus from the high initial level would not change the cell's response. Such a pattern would be disadvantageous to the organism because the small changes are of potentially greater significance than the large constant stimulus is.

In the saccule of the bullfrog, however, such sensory "saturation" does not occur. When an isolated hair cell from the saccule is tested in the unperturbed state, it is found to be most sensitive to displacements of a few tenths of a micrometer in either direction. If the hair bundle is moved to a new rest position that is, say, one micrometer from the first position, the result is not simply a large constant response. Instead after a few tenths of a second the response to the static one-micrometer displacement decreases and the cell becomes sensitive to small movements away from the new rest position. The basis of this adaptation is not known, but its significance is clear: it enables the organism to disregard large continuous stimuli while maintaining great sensitivity to small, rapid sensory inputs that are of greater importance.

The responses of the isolated hair cell induce wonder at a molecular mechanism capable of reacting to stimuli with an amplitude of picometers that come as often as 100,000 times per second. The means by which the transduction channels are opened and closed is currently the subject of intense scrutiny in my laboratory at the California Institute of Technology and in other laboratories. One plausible hypothesis is that the transduction channels are continually being opened and closed by the random



PATTERN OF ELECTRIC CURRENTS in the fluid around the hair cell suggests that the channels mediating the cell's electrical response are near the top of the hair bundle. The changes in electric potential in the hair cell are caused by flows of positive ions into the cell from the surrounding endolymph; the ions are attracted by the negative potential in the interior of the cell. The main positive ion is potassium. The flow of potassium ions (and hence the electric potential of the receptor) is regulated by the opening and closing of "transduction channels" in the cell membrane. When the transduction channels open and potassium ions flow into the cell, there is also a small current in the endolymph around the channels. Measurements of the voltage due to this current at various points around the hair bundle are shown. The largest of the currents has an amplitude of about 13 microvolts (millionths of a volt). The fact that the current is greatest near the top of the hair bundle suggests the transduction channels are there.

thermal motions of nearby molecules. If that is the case, the fraction of channels open at any one moment will depend on the difference in energy level between a closed channel and an open one. If the difference is greater than the kinetic energy of the average collision with a nearby molecule, the opening of a channel will be a rare event and at any moment most channels will be closed. If the energy levels of closed and open channels are similar, however, about half the channels will be open at any given time.

Suppose that initially the energy level of an open channel in the apical area of the hair-cell membrane is relatively large. Most of the channels will then be closed. If the displacement of the hair bundle somehow reduces the energy of an open channel, the fraction of channels that are open will increase rapidly under the influence of the motion of the nearby molecules. David Corey of the Yale University School of Medicine, who was then a graduate student in my laboratory, made recordings from hair cells that are consistent with this hypothesis. The central question of how the movement of the hair bundle changes the energy of the channel, however, remains unanswered.

After considering what is known about the transduction mechanism of the hair cell, the question naturally arises of how the change in potential leads to the conveying of a message to the brain. Since hair cells do not have axons, they cannot relay information directly. Instead the message is carried by a neuron

with an ending near the basal surface of the receptor. When the transduction channels of the hair cell open, the potential drop across the membrane changes very fast in all parts of the cell. In the basal area of the cell membrane there are many channels that selectively admit calcium ions. The action of the calcium channels is voltage-dependent: when depolarization occurs, the channels open.

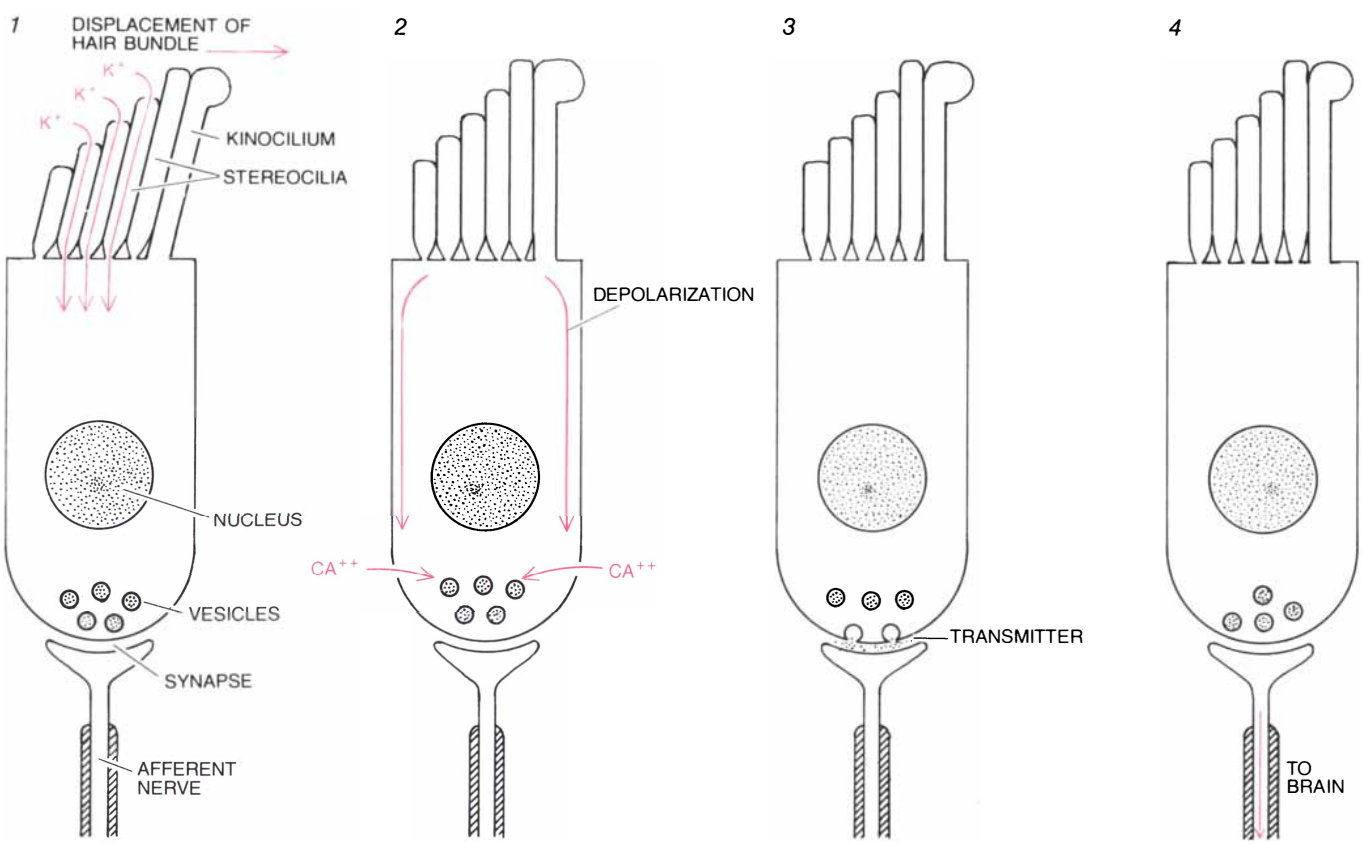
Once calcium ions are admitted to the cell, they serve to trigger another significant event. Near the base of the hair cell are many vesicles containing a chemical substance that acts as a neurotransmitter. The calcium ions cause the vesicles to fuse with the outer membrane and release their contents into the synaptic gap that separates the hair cell from the terminal of the neuron.

The exact chemical composition of the transmitter substance is not known. It is known, however, that when the transmitter molecule is released from the vesicle, it diffuses across the synaptic gap: a distance of about 20 nanometers. On combining with receptors on the membrane of the nerve terminal the transmitter causes a depolarization of the neuron much like the depolarization of the hair cell. In the neuron the depolarization elicits one or more "action potentials," electrical signals that travel along the axon.

A hyperpolarization of the hair-cell membrane has the opposite effect. Even in the resting state some of the calcium channels at the base of the cell are open; hence a small amount of the transmitter is always diffusing across the synapse and a small signal is carried to the brain. A hyperpolarization closes down the open calcium channels and reduces the amount of transmitter that is released. The frequency of occurrence of the action potentials in the neuron is thus reduced, and in the logic of the nervous system such a reduction can also serve as an information-carrying signal.

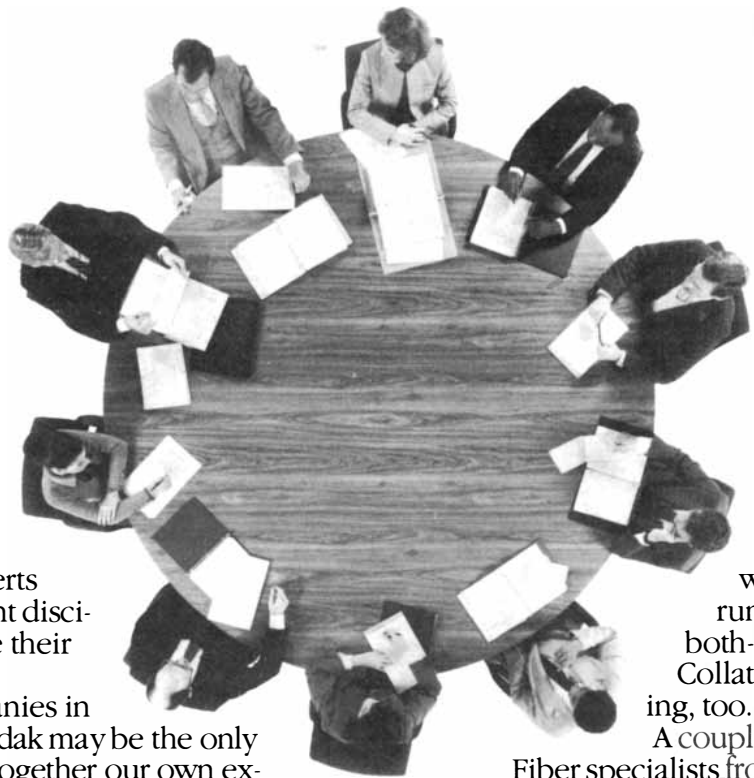
The information from all the hair cells in the inner ear is carried to the brain by axons in the eighth cranial nerve. Each of the many fibers in the nerve carries signals from hair cells of only one organ. Moreover, each fiber carries very specific information. For example, a particular fiber from one semicircular canal is excited only by angular accelerations about a particular axis in the head, and only by the accelerations in the clockwise direction.

Messages travel not only from the hair cell to the brain but also from the brain to the hair cell. Near the base of the hair cell are two sets of nerve fibers. The fibers called afferent carry information to the brain; those called efferent carry information back to the hair cell. When the efferent fibers to the cochlea are activated, the hair cells appear to become less sensitive to sound. In addition they seem to be less sharply tuned to particular frequencies. The significance of such desensitization is not clear. It is one of the many puzzles of the hair cell that remain to be solved. Work is being done on hair cells in many laboratories, and in the next few years many questions, including the central one of the transduction mechanism, will undoubtedly be answered. Further work, however, can only reinforce a sense of wonder at the sensitivity and complexity of this miniaturized piece of biological apparatus.



RESPONSE OF A HAIR CELL culminates in the transmission of an electrical signal to the brain along the afferent nerve fiber at the base of the cell. The response is shown schematically in the panels of this illustration. When the hair bundle is displaced, the transduction channels open (*1*). Potassium ions flow into the cell, and the potential difference between the cell and the surrounding fluid falls. The reduction in potential difference, which is called a depolarization, spreads almost instantaneously through the cell (*2*). In the lower part of the cell are channels that selectively admit calcium ions. The depolarization causes these channels to open, whereupon calcium flows in. Near the base of the cell are vesicles containing a neurotransmitter: a substance that transmits the nerve impulse from one cell to the next. The calcium ions cause the vesicles to fuse with the basal part of the hair cell's surface membrane (*3*). In fusing the vesicles release the neurotransmitter they contain. The transmitter, whose chemical nature is not known, diffuses across the synaptic space between the hair cell and the neuron; it then excites the neuron, which sends a message to the brain along a fiber of the eighth cranial nerve (*4*).

We discovered a new kind of brain cell.



You'll find it at places like the Princeton Advanced Center for Learning, where experts from a dozen different disciplines cross-pollinate their thoughts and ideas.

But as far as companies in America go, we at Kodak may be the only ones who can bring together our own experts in programming, optics, physics, organic and polymer chemistry, fibers, human engineering, imaging, marketing, paper technology ...we've probably skipped a few.

Call it synergism or whatever. But when these pros sit down at our table, they come up with some truly incredible results.

The Kodak Ektaprint 250 duplicator is a good case in point. It has grown out of more disciplines than any single Kodak product before it.

The nucleus is two 8085 microprocessors, with a combined yield of over 100,000 bytes, and responsibility for 134 separate functions.

There are all kinds of thoughtful touches. One is a programmable display that's as friendly as can be—with 94 different messages, not codes, to help anticipate and satisfy operator needs. And there's an internal diagnostic system that's amazingly specific.

Then there's single-pass duplexing. Lots of heads went together for this run-through-once, print-both-sides convenience. Collating and cover inserting, too.

A couple of other neat feats: Fiber specialists from our Eastman Chemicals Division in Tennessee helped create the "perfect" toner brush. In Rochester, our experts in 35 mm added sprockets to the belt drive. And a reflective "bathtub" was perfected by our processing group; it helps keep the paper path shorter and virtually trouble-free.

With all these new capabilities, you may think the 250 needs a genius to run it. Wrong. And that's the beauty. 99% of the duplicator is microprocessor-controlled. That means reliability, flexibility, productivity, easier operation. And time to think about more important matters.

For more information on how we plan to combine our talents for future applications in microelectronics, please send for our brochure ..."Resources To Innovate, Commitment To Serve..." (B3-190). Write to Eastman Kodak Company, Dept. GB-SA 6, 343 State Street, Rochester, New York 14650.

Kodak

Kodak. Where technology anticipates need.

SCIENCE AND THE CITIZEN

Orphans in the Marketplace

In a free-market economy if the demand for a commodity is not sufficient to ensure that its sale will yield a profit, the commodity will not be produced. If the commodity is a medication and the demand for it is the result of a disease, the consequences of the economic system can be life-threatening. The rule of thumb used by American pharmaceutical manufacturers is that there must be 100,000 victims of a disease to make the distribution of a drug for it profitable. There are about 2,000 diseases that afflict fewer than 100,000 people in the U.S. For the most part the pharmaceuticals industry has shown little interest in developing treatments for them. For some of the diseases medications exist but are not currently on the market.

Such orphan diseases and orphan drugs were the subject of a recent conference at the University of Michigan at Ann Arbor. It was held at a time when there is considerable interest in the subject. Bills have recently been adopted by the House of Representatives and the Senate providing tax incentives for companies that bring orphan medications to market. The U.S. Food and Drug Administration (FDA) has established an Office of Orphan Product Development to make it easier for such substances to be approved for clinical use without the support of a sponsoring corporation.

Some observers maintain that the stringent and complex FDA regulations governing premarket drug testing are responsible for the orphan status of many therapeutic substances. The regulations are said to make the cost of developing a new drug so high that it inhibits the testing of new medications for rare conditions. George J. Brewer of the University of Michigan Medical School, who was one of the organizers of the conference, disagrees with this assessment. According to Brewer, "overregulation is not the problem with orphan drugs. ... The FDA is willing to streamline the approval procedure."

Brewer argues that the main causes of the orphaning of drugs and diseases are an absence of interest on the part of the pharmaceutical companies and a lack of work on the diseases by biomedical investigators. As evidence of the manufacturers' lack of interest he has cited three substances that show promise for the cure or control of orphan diseases. One substance is L-5-hydroxytryptamine, an experimental drug for the treatment of myoclonus, a rare neuromuscular disorder. Another is cysteamine, which is employed in treating cystinosis, an inherited defect of amino acid metabolism that affects about one infant in 50,000.

The third medication is zinc acetate, a compound Brewer is testing as a means of controlling the symptoms of sickle-cell disease. Sickle-cell anemia, the result of an inherited defect in red blood cells, affects about 25,000 black Americans. Brewer's work suggests that zinc compounds can control one of the major symptoms of the disease, an abnormally high level of calcium in the red blood cells. No pharmaceutical company has shown much interest in any of the three substances, Brewer said, even though the drugs have considerable therapeutic promise.

In the case of zinc acetate the obstacles to a profitable investment are increased by the fact that the active agent, zinc, is a chemical element and therefore cannot be patented. If a company took the steps necessary to have the compound approved by the FDA for therapeutic use, the company could not claim exclusive rights to manufacture the medication. Indeed, a second company, having been spared the cost of developing the product, might be able to sell it at a lower price.

The number of drugs that offer an effective means of curing or controlling a disease but are not available commercially is a matter of controversy. Pharmaceutical companies are under no obligation to reveal how many substances they have tested but not placed on the market. At the Ann Arbor conference representatives of the Pharmaceutical Manufacturers' Association said that only about 10 orphan drugs had been found in a survey of manufacturers and biomedical investigators. According to Brewer, however, the survey was incomplete. Neither he nor other investigators he knows who have worked on orphan drugs were approached, he said. There may be considerably more than 10 orphan products, he added.

Furthermore, cures or effective therapies are known for only about 10 percent of the 2,000 orphan diseases. Workers in the biological sciences have shown little enthusiasm for identifying and testing potential medications for the remaining 90 percent. One reason is the scarcity of Federal funding for the investigation of medications and specific treatments. Most public funding in biomedicine has gone for work on disease mechanisms or on basic physiological processes. The development of drugs and treatment regimens has been left to the manufacturers.

Brewer argues that if treatments for orphan conditions are to be developed, the National Institutes of Health, the main Federal agency supporting work in the biological sciences, will have to establish a separate review process for making grants in this area. Grant applications for work on specific drugs and treatments suffer by comparison with applications for work in basic biological processes. "Our grants get thrown in with cutting-edge science, which is quite elegant," he said. Work on varying the dose of a medication for treating an orphan disease can appear "quite mundane" in such a context.

Brewer added that without increased public support there may never be effective medications for most orphan diseases. Pharmaceutical manufacturers cannot fundamentally alter their policy toward the diseases because the companies' continued existence depends on their capacity to make a profit. According to Brewer, "the drug companies can be helpful in the final stage [of drug testing and marketing], but biomedical investigators must do much of the first stage...."

Element 109 and Beyond

The reported synthesis of a new heavy element—number 109 in the periodic table of the elements—has stirred fresh interest among nuclear physicists in the prospects for finding a long-sought sequence of stable, superheavy elements with atomic numbers higher than 114. A single atom of element 109 was tentatively identified among the products of an extraordinarily "gentle" nuclear collision staged last summer at the laboratory of the Society for Heavy Ion Research in Darmstadt, West Germany. The team of experimenters was led by Peter Armbruster and Gottfried Münzenberg.

The new element was produced in the laboratory's linear heavy-ion accelerator (named Unilac) by a technique known as cold fusion. A beam of iron-58 ions was accelerated and fired at a thin foil of bismuth 209. In principle a collision between an iron nucleus with 58 nucleons (26 protons and 32 neutrons) and a bismuth nucleus with 209 nucleons (83 protons and 126 neutrons) could be expected to result in the formation of a compound nucleus with 267 nucleons (109 protons and 158 neutrons), but only under certain strictly defined conditions. The collision would have to be energetic enough for the two nuclei to fuse but not so violent that they would immediately fly apart.

Calculations showed that even under carefully controlled conditions the cold fusion of iron 58 with bismuth 209 would take place only rarely: about once in every 100 trillion nuclear collisions. Armbruster, Münzenberg and

THE BMW 733i. IT DOESN'T SCREAM WEALTH. IT JUST QUIETLY REWARDS INSIGHT.

It's no secret that there are some expensive sedans that are purchased for their sheer expensiveness.

Sedans that serve as a sort of public declaration of prosperity and status.

Such cars are the antithesis of the BMW 733i—the luxury sedan engineered in the belief that performance, not self-congratulation, is the ultimate measure of a car's true worth.

A PANACEA FOR PURISTS.

The 733i is a car whose subtleties of engineering and technology place it beyond the needs of all but a handful of purists. (Which is perhaps just as well, given its limited supply.)

For those serious drivers, however, it is as close to a necessity as a $35,000* sedan can come.

Its fuel-injected engine delivers the sort of heady response that suggests an oversized power-plant beneath the hood.

Not so. The 733i's 6-cylinder engine arrives at high performance through refinement, not brute force.

A system of microprocessors gathers data on fuel mixtures, engine speed and other factors. It then orders ignition to occur at the optimum moment—accomplishing through elegance of design what might otherwise require additional cylinders and extra girth.

Its suspension handles the landscape with precision instead of bluntness—removing almost surgically the wallowing, skittishness and related uncertainties from bumps and curves.

(One reason: a new rear suspension that incorporates one of the most significant breakthroughs in independent rear suspension design in a generation.)

The 733i is also designed for those who prefer the subtleties of road topography to the anesthetizing ride of limousine-like sedans. Its highly tactile steering places you in touch with that alternately pleasant and harsh reality called the road—filtering out just enough of its harshness, while never removing the reality.

The result is a machine that is as much system as car, one so perfectly balanced and rigorously self-policing that it even calculates when routine service is needed.

ENGINEERED FOR DRIVING, NOT JUST SITTING.

Inside, the 733i is engineered for activities of considerably more subtlety than sitting.

Its interior literally helps you drive, rather than merely accompanying you on the trip.

Its supple leather seats are designed to follow the curvature of the spine, placing the driver in postures that are anatomically correct and help reduce driver fatigue.

An Active Check/Control provides readings on 7 different measures of the car's operational readiness.

An on-board computer provides all manner of trip information, such as the distance remaining to your destination.

In fact, no detail in a BMW 733i ever suffers from inattention because it's judged 'minor.' And its warranty reflects this obsessiveness. The 733i is backed by BMW's 3-year/36,000-mile limited warranty and a 6-year limited warranty against rust perforation.†

All of which led Car and Driver to conclude that the 733i's "parts and pieces...work so well together they must have been melded in another world."

It may be seen, however, in much more convenient locations. We invite you to visit your nearby BMW dealer, who will be happy to arrange a test drive.

THE ULTIMATE DRIVING MACHINE.

*Manufacturer's suggested retail price: $34,300. Actual price depends on dealer. Price excludes state and local taxes, dealer prep, destination and handling charges. †Warranty applies to automobiles purchased from authorized U.S. BMW dealers only. See your BMW dealer for details. © 1982 BMW of North America, Inc. The BMW trademark and logo are registered. European Tourist Delivery can be arranged through your authorized U.S. BMW dealer

their colleagues planned a 10-day run on the Darmstadt accelerator in late August to search for evidence of element 109. It was not until August 29, the last day of the scheduled run, that their detector registered a likely event. Four measurements of the properties of the newly formed atom and the products of its decay confirmed that a nucleus of element 109 had evidently existed for approximately five milliseconds.

The synthesis of element 109, assuming that it can be repeated, would set a new nuclear record: it is the heaviest element detected so far. The natural elements extend through atomic number 92, and artificial ones through number 106 have been known for some time. In 1981 element 107 was identified by Armbruster and Münzenberg and their colleagues, following an earlier, disputed claim by investigators at the Joint Institute for Nuclear Research in Dubna. (Element 108 is thought to be less stable than either 107 or 109.)

The apparent success of the cold-fusion technique has stimulated efforts to extend it to the long-standing search for superheavy elements. The possibility that such elements might exist is suggested by the shell model of the nucleus, which describes nucleons as occupying successive shells much like the shells of electrons that determine the chemical properties of atoms. Each shell can hold a particular number of nucleons, and filled shells bring enhanced stability. The filling of nucleon shells in elements with atomic numbers in the range from 114 to 126 is expected to make such hypothetical nuclei unusually stable. Elements occupying such an "island of stability" would presumably be easier to detect than elements with atomic numbers ranging from 109 to 114.

The German physicists have now joined forces with a group at the Lawrence Berkeley Laboratory, headed by veteran heavy-element synthesizer Albert Ghiorso, in an attempt to make element 116 by the cold-fusion technique. The current experiments call for bombarding a target of curium 248 with an accelerated beam of calcium-48 ions in the Berkeley laboratory's SuperHILAC heavy-ion accelerator as well as in the Darmstadt machine. Ghiorso, who once bet his colleague Glenn Seaborg $100 that superheavy elements would never be found, has reportedly lowered the odds against the experiments' succeeding to no more than 100 to one.

Rosy Future

To understand the genetic control of development, molecular biologists are eager to find ways to introduce a gene into animal germ cells and monitor its expression as a protein product in the mature animal. Ideally the inserted gene would become a permanent part of the animal's genome, or total complement of genes; it would be passed on to successive generations and be expressed by the animal's offspring. Recombinant-DNA techniques make it possible to isolate a single gene, perhaps to alter it in a defined way and to clone it in quantity in bacteria. The trick then is to get the gene into animal germ cells and have it integrated into the genome without further, inadvertent alteration, so that it is correctly expressed, stably inherited and expressed in future generations. This has now been accomplished. Workers in the Department of Embryology of the Carnegie Institution of Washington report in *Science* an efficient new method of gene transfer whereby a cloned gene can be injected into embryos of the fruit fly *Drosophila melanogaster*. When the gene called *rosy* is transferred into embryos of a mutant line of flies with brown eyes, fertile adults that develop from the embryos give rise to a stably modified strain with the bright red eyes of normal *D. melanogaster* flies.

In earlier (and continuing) efforts to insert foreign DNA into animal cells investigators have "transfected" cultured cells by exposing them to naked foreign DNA, they have microinjected the DNA into the nucleus of cells or into fertilized eggs and they have exploited the ability of an animal virus to "transduce" a DNA segment into cultured cells. Gerald M. Rubin and Allan C. Spradling of the Carnegie Institution took advantage of a different natural vector: a transposon, or transposable genetic element. A transposon is a segment of DNA that moves around in the genome of a bacterium or a higher organism and is capable of inserting itself—and often other segments of DNA incorporated within it—at a large number of sites.

Rubin and Spradling worked with a family of *Drosophila* transposons called *P* elements. Rubin and several co-workers had identified a particular *P* element, much longer than the rest of them, that seems to be required for transposition. It is long enough to code for a transposase: an enzyme that catalyzes the transposition of the long element itself and of shorter "defective" *P* elements. The long element also encodes a repressor protein that blocks transposase activity. The result is that the introduction of a long *P* element into germ cells not already carrying it gives rise to transposition at a high rate; once a long element is present the repressor accumulates and prevents transposition.

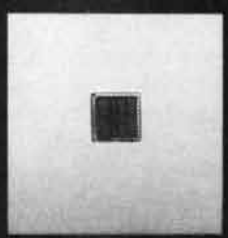
In a first series of experiments Spradling and Rubin inserted a long *P* element into a plasmid (a small circle of bacterial DNA) and injected the plasmid into a *Drosophila* embryo. They were able to show that the *P* element can catalyze its own transposition into the embryo's germ cells. Plasmids are routinely used as vectors for introducing foreign DNA into bacteria. The self-transposing ability of the long *P* element suggested that a gene of interest might be introduced into *Drosophila* germ cells as a passenger within a long *P* element. Better yet, unwanted effects of the transposase could be avoided by inserting the gene into a defective *P* element and supplying just enough transposase to catalyze the desired transposition.

Rubin and Spradling undertook to transpose the *rosy* gene, which codes for the enzyme that makes a normal fly's eyes red. They inserted *rosy* into a short *P* element, which was inserted in turn into a plasmid and was cloned in bacteria. Then they microinjected the plasmid into embryos of mutant (brown-eyed) strains. To supply the necessary transposase they tried two stratagems. One was to inject the plasmid carrying the *rosy* transposon into mutant embryos derived from a cross between males carrying the long *P* elements and females lacking them. The genome of such a hybrid undergoes transposition at a high rate; the transposition is catalyzed by the male parent's transposase, which is temporarily "derepressed" in the hybrid germ cell. The other stratagem was to coinject, into embryos lacking *P* elements, a high concentration of the *rosy* plasmid and a low concentration of the long *P* elements, which would supply the transposase. Given the excess of *rosy* transposons and the limited supply of transposase, the expectation was that many germ cells could incorporate a *rosy* transposon but not a long *P* element.

Both methods worked. Some 8 percent of the injected embryos developed into fertile adults, and 39 percent of those adults had at least some progeny with red eyes; the eye color was inherited stably in later generations. By probing the DNA of the red-eyed flies Rubin and Spradling were able to show that the *rosy* transposon was indeed integrated into chromosomal DNA and was responsible for the change in eye color. Furthermore, only the transposon was taken up by the fly genome; the bacterial DNA of the plasmid was not.

Rubin and Spradling foresee two major applications of their technique. One is the study of gene regulation in higher organisms, notably in *D. melanogaster*. The fruit fly has been the subject of intensive genetic analysis for decades. Many genes that have been identified seem to play an important role in development, but their protein product and mode of operation are not known. Now those genes and their products can be studied in detail in the developing organism. The fly's *P* element or modified versions of it may function in other animals as well; if the *P* element will not cross species lines, analogous transposons should be found in other species. Eventually, Rubin and Spradling think,

In 1981, Hewlett-Packard announced the world's densest computer chip.



Our 450,000-transistor, one-chip
32-bit CPU.

Today, it's the heart of a 32-bit co that's so affordable your top enginee can have their own mainframes.

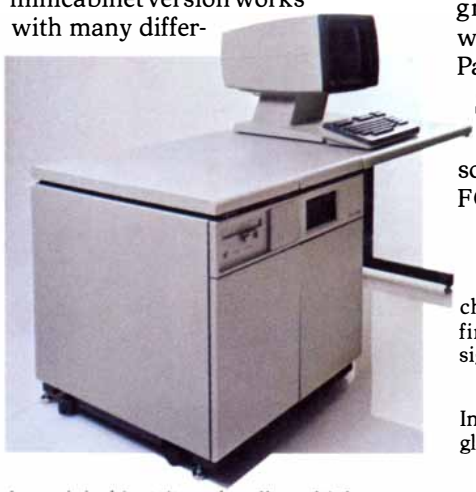
From time to time, miracles of technology come along to make previously impossible tasks not only possible, but easy. That little integrated circuit chip on the preceding page is one of those technological miracles.

Hewlett-Packard didn't develop it just to break the record for most transistors on a chip, but to put on an engineer's or scientist's desk a computer so powerful that it can do the work of mainframes costing four times as much.

32-bit computers for 32-bit applications.

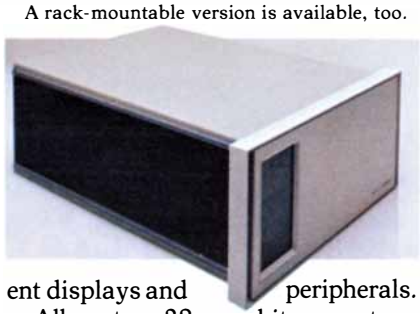
The new HP 9000 computer based on this and four other 'superchips' can handle formidable engineering and scientific problems. The scientist solving complex systems of equations, the mechanical engineer doing finite element analysis or three-dimensional modeling, the electrical engineer analyzing complex circuits or designing very large-scale integrated circuits — these are the kinds of technical people and problems the HP 9000 family is designed for.

It comes in three versions. The integrated workstation is complete with keyboard, color or monochromatic graphics display, fixed and flexible disc drives, and printer. For systems manufacturers, there's a rack-mountable box. And for a variety of single-user and multi-user applications, the minicabinet version works with many different displays and peripherals.



As a minicabinet, it can handle multiple users.

A rack-mountable version is available, too.



All are true 32-bit computers, with 32-bit CPUs, memories, and data paths. And the multi-CPU architecture lets you nearly double or triple your processing power at any time by adding one or two CPU boards. Without increasing the computer's size.

Two operating systems are better than one.

The integrated workstation is available with a choice of operating systems. One is HP's highly evolved, high-performance Enhanced BASIC, augmented with 3-D graphics and a software innovation called a run-time compiler. This substantially increases program execution speed, while retaining an interactive development environment.

The other operating system, called HP-UX, is a fully supported, extended version of the popular UNIX® HP-UX, available on all HP 9000s, adds virtual memory, graphics, data base management, data communications, and enhanced file capability to the basic UNIX 'shell.' High-level programming languages available with HP-UX are FORTRAN 77, Pascal and C.

Software, and plenty of it.

Much of the vast range of existing software written in HP BASIC, FORTRAN 77, Pascal and C is transportable to the HP 9000. HP will also be offering proprietary software packages emphasizing computer-aided design and engineering. These will tie the HP 9000 into HP's Manufacturer's Productivity Network (MPN). Third-party software suppliers will be providing many of the most widely used CAE packages for 32-bit computer systems. And both HP 9000 operating

The 32-bit CPU chip is bonded to the finstrate which doubles as a signal carrier and heat sink.

Up to three CPU boards and three Input/Output Processors can fit into a single HP 9000.

TCG-207

nputer
rs and scientists

systems offer extensive program development tools.

You also get a choice of communication tools. The HP 9000 is currently compatible with Ethernet,™ and with HP's Shared Resource Manager (SRM) which lets clusters of HP 9000 and 16-bit desktop computers share data and use common peripherals.

Links to central computers are also available. And in late 1983, HP will offer local area networks based on the IEEE-802 standard.

New technology from the silicon up.

The five superchips that make the HP 9000 possible are the 32-bit CPU, which can execute a million instructions per second; an eight-channel Input/Output processor (IOP); a random-access memory chip capable of storing 128K bits of data; a memory controller that 'heals' up to 32 bad memory locations; and an 18-megahertz clock.

Hewlett-Packard's advanced NMOS-III process makes it possible to put 450,000 transistors on a chip only 0.4 centimeters square. This tremendous density of electronic components could have required an expensive and elaborate cooling system.

Instead, HP engineers developed a new mounting structure called a finstrate, a copper-cored circuit board, which acts as both cooling fin and substrate. The finstrates containing the CPU, IOP, memory, and clock chips are housed in a lunchpail-sized module.

One user, one mainframe.

Clearly the trend in engineering and scientific computation is away from large machines shared by multiple users and towards networks of powerful personal workstations, sharing peripherals and data bases. The reason is compelling. An engineer or scientist in personal control of an HP 9000 can solve so many more problems more easily that the increased productivity alone makes the cost of individual computers easy to justify.

For complete information about this powerful breakthrough in 32-bit computing, contact the local HP sales office listed in your telephone directory. Ask a Technical Computer Specialist for a demonstration. Or write to Pete Hamilton, Dept. 56151, Hewlett-Packard, 3404 East Harmony Road, Fort Collins, CO 80525.

HEWLETT PACKARD

UNIX is a registered trademark of Bell Laboratories. Ethernet is a trademark of Xerox Corporation.

gene transfer mediated by transposons should make possible the modification of plant and animal genomes, for example to improve agricultural yield or to confer resistance to disease.

The Source at the Center

The center of the Milky Way galaxy is hidden by stars, gas and dust, but lately clues to what lies in the center have begun to emerge. One clue is that the galactic center emits gamma rays, the highest-energy form of electromagnetic radiation. Furthermore, the gamma-ray emission is marked by a sharp spectral line: all the rays have the same energy, namely .511 million electron volts. This is precisely the energy equivalent of the mass of an electron or a positron (the antiparticle of the electron), and so it is all but certain the gamma rays are generated in pairs when electrons and positrons collide and annihilate each other. The radiation from the galactic center was discovered in 1977, when a gamma-ray detector suspended from a balloon was flown over Alice Springs in Australia by a group led by Marvin Leventhal of Bell Laboratories and Crawford J. MacCallum of the Sandia National Laboratories.

Further observations now suggest that the gamma-ray emission varies with time; indeed, it seems to have stopped. A detector placed aboard the *HEAO-3* satellite by workers at the Jet Propulsion Laboratory shows that between September, 1979, and March, 1980, the emission declined by two-thirds, and the most recent measurement made by Leventhal and his colleagues (on November 21, 1981) shows no emission at all. The absence of emission is confirmed by investigators at the Goddard Space Flight Center of the National Aeronautics and Space Administration and the Saclay Nuclear Research Center in France, who launched a balloon-borne detector on November 20, 1981.

In *The Astrophysical Journal* Leventhal and his colleagues comment on the various observations. "It would appear," they write, "that the source was going into a 'low' state at the time of the 1980 March *HEAO-3* observations and has remained there through 1981 November." Moreover, "it seems safe to conclude that the time scale for change is at least as short as six months." The time scale serves to limit the size of the source. It is a "relatively compact object," considered on a galactic scale; apparently it can be no larger than one light-year, the diameter a shell of light would have if it expanded for six months from a central point. (The speed of light limits the speed at which any process spreading through space could cause a coherent change in emission.)

Is the source a black hole? Leventhal and his colleagues cite observations at infrared wavelengths showing the presence of clouds of ionized gas within 10 light-years of the center of the galaxy. The velocities of the clouds suggest they are orbiting a supermassive central object. "Apparently no problem exists" in hypothesizing that a variable flux of positrons boils off the disk of hot gas thought to ring a black hole. The positrons, passing through the surrounding clouds of ionized matter, would then have a good chance of colliding with electrons, yielding gamma rays.

Platonic Chemistry

Ever since Friedrich August Kekulé grasped the shape of the benzene molecule in a dream of snakes biting their tails, the topology and geometry of organic molecules has been one of the more charming pursuits of structural chemistry. The DNA molecule twists thousands of times before its ends can join to form a ring, and certain enzymes can tie DNA into topologically interesting knots. Only recently, however, has anyone been able to create the molecular analogue of a Möbius band, a strip joined end to end after a single half twist. Likewise, the orderly arrangement of atoms in a crystal can exhibit remarkable periodic symmetry, yet creating molecules with the form of the five Platonic solids has proved a considerable challenge. The synthesis of a molecular dodecahedron has now been reported.

The molecular Möbius band was synthesized by David M. Walba, R. Curtis Haltiwanger and Rodney M. Richards of the University of Colorado at Boulder; they made the band in much the way it is done with paper, scissors and paste. First carbon and oxygen atoms are bound together in two parallel chains, and the chains are fused by double bonds between carbon atoms lying opposite each other on the chains. The double bonds form the rungs of a ladder-shaped molecule called THYME diol ditosylate, and the network of bonds in the ladder defines a surface analogous to the surface of a long, narrow sheet of paper. The top and bottom of the ladder have a chemical affinity for each other, and so the ends can be joined, forming a ring. If the legs of the ladder cross when they join, the surface defined by the network of atomic bonds has the same topological properties as a strip of paper given a half twist and then pasted together at the ends.

One striking characteristic of the Möbius band is that it has only one continuous edge; run a finger along the edge of the paper model and the entire edge will be traversed before the starting point is regained. Moreover, slitting the band lengthwise around its entire circumference yields a single, thinner band that is twice as long and has two half twists. Walba and his colleagues carried out the chemical analogue of the slitting operation. When the molecule is immersed in ozone, the double-bonded carbon rungs of the molecular ladder are broken; the resulting molecule has indeed proved to be a ring twice the original size. The next step, according to Walba, will be to build longer molecular ladders so that more twists of the legs are possible before the ends are joined. Slitting a band with two half twists gives two separate bands linked together; slitting a band with three half twists gives a single band that winds through itself to form a trefoil knot.

The synthesis of the molecular dodecahedron is a task of even greater delicacy: the pieces of the 12-sided polyhedral cage must be attached in a strict sequence or the entire structure will collapse of its own energetic imbalance. Leo A. Paquette and his co-workers at Ohio State University are the only structural chemists to have achieved the synthesis so far, although other investigators are testing other routes to the same result. Paquette and his colleagues began with the hydrocarbon cyclopentadiene, made up of five carbon atoms arranged in a regular pentagon that eventually becomes one of the sides of the dodecahedron. In subsequent steps sides of the dodecahedron are added in such a way that the intermediate molecules remain roughly symmetrical about an axis. The final molecule, called dodecahedrane, is framed by 20 carbon atoms that make up the vertexes of the dodecahedron. Twenty hydrogen atoms, one attached to each carbon atom, lie outside the polyhedral surface defined by the carbon atoms and the carbon-carbon bonds.

The dodecahedron is the last of the five regular Platonic solids to be synthesized. Cubane, in which carbon atoms form the vertexes of a cube and a hydrogen atom is attached to each carbon atom, was made in 1964 by Philip E. Eaton and Thomas W. Cole, Jr., of the University of Chicago. The corresponding four-sided carbon skeleton would be called tetrahedrane; it has not been synthesized, but carbon atoms have been made to form a tetrahedron by investigators in Germany, who emplace butyl groups (C_4H_9) at the four vertexes. Both octahedrane, the eight-sided organic molecule, and icosahedrane, the 20-sided one, are probably too unstable to exist, but molecules in the two forms can be readily synthesized from atoms other than carbon.

Such geometric modeling may seem little more than an elegant trick, but the high degree of symmetry of the Platonic molecules makes it possible to investigate the properties of chemical bonds in great detail. Paquette also notes that a highly symmetrical molecule with an amino group attached to one vertex

readily passes through the membrane of a cell and tends to destroy virus particles inside it.

The Migrating Marsupial

When marsupial mammals began to flourish some 50 million years ago, they inhabited the New World (particularly South America), western Europe and Australia. The keen edge of Occam's razor makes it unlikely that the three groups evolved independently, but the question has remained whether New World marsupials had colonized Australia, as they evidently had Europe, or whether the order originated in Australia and radiated from there. The further question of how the animals could have migrated in either direction was essentially answered some years ago by students of continental drift, who found that when the marsupials were developing, South America was connected to Australia by way of Antarctica.

Until recently, however, the Antarctic fossil record, which is surprisingly rich in fishes, reptiles, birds and even sea mammals, gave no evidence of marsupials having passed that way either coming from or going to Australia. Such is no longer the case. Writing in *Science,* Michael O. Woodburne of the University of California at Riverside and William J. Zinsmeister of Ohio State University have announced the discovery of fossil marsupial jawbones and teeth of late Mesozoic to early Cenozoic age in a sandstone formation on Seymour Island, off the Antarctic Peninsula.

The fossils, the first evidence of land mammals found so far in the Antarctic, have been assigned to the extinct marsupial family of polydolopids. Remains of the group were unearthed previously in Patagonia, Bolivia and Brazil. Because polydolopids had not been known from anywhere except South America, their presence in the Antarctic is strong evidence that they came there from the New World rather than from Australia. On the other hand, late Mesozoic and early Cenozoic marsupial fossils from Australia are unknown, and later ones all resemble one or another of the eight marsupial families that still survive there rather than any potentially ancestral foreign marsupial. Hence the issue of which way the Mesozoic migration went is still unresolved.

Unforgettable Data

In the earliest electronic computers space for the storage of information was a precious resource; as a result programs were written in the most compact form possible. As the cost of computer memory has decreased, the need for succinctness in programs has diminished. More emphasis can now be put on other qualities, such as the speed at which a program operates or the ease with which it can be written, understood and altered. The effects of further decreases in the cost of memory have recently been discussed by George P. Copeland of the Servio Logic Corporation in Portland, Ore. Writing in *Computer,* he describes the likely direction of events by speculating on what a computer system would be like if the mass storage of information were completely free.

Mass storage, which commonly takes the form of magnetic tapes or disks, is the cheapest kind of computer memory. Large quantities of information, such as the payroll and employment records of a business, can be kept indefinitely in a mass-storage device; when a particular record is needed, it is called up into the internal memory of the computer. In the way such a "data base" is maintained today, records are periodically added, altered and deleted. The most important change that would be brought about by free mass storage, Copeland finds, is that no information would ever be deleted from the system.

The ability to delete information might seem to have certain advantages even apart from the clearing of space for new data. In particular, when an error is discovered, the usual means of correcting it is to delete the erroneous information and replace it with a revised version. Copeland points out that there are hazards inherent in this practice. For example, an error in stating an employee's salary could be corrected by the deletion-and-replacement method, but in the meantime the incorrect amount might have been used in calculating pension benefits, insurance premiums, tax withholding and so on.

Forbidding deletion eliminates such attempts to "go back into the past and change history," but a new awkwardness is introduced. Each time a record is modified a new version is created, but the old one also remains in the system; hence multiple copies of the record accumulate. Copeland proposes to keep track of the proliferating data by assigning each record two identifying numbers. One number is called a surrogate, and it remains constant for all versions of the record; the other number serves as a "time stamp" indicating when the version was created. Copeland notes that the nondeletion policy, in which errors can be adjusted but never expunged, is not a new idea in record keeping. It is an established practice in accounting, where "no journal entry or ledger posting deletes any information."

The cost of mass storage in computer systems will never decline to zero, but it may well become low enough to make the permanent retention of information feasible. Indeed, there is at least one existing low-cost storage medium that seems to require a nondeletion policy: the optical disk. On such a disk information is recorded as a pattern of microscopic pits etched into the surface by a laser beam. The inability to erase the pits has occasionally been cited as a disadvantage of optical-disk recording, but in Copeland's view it may be a virtue.

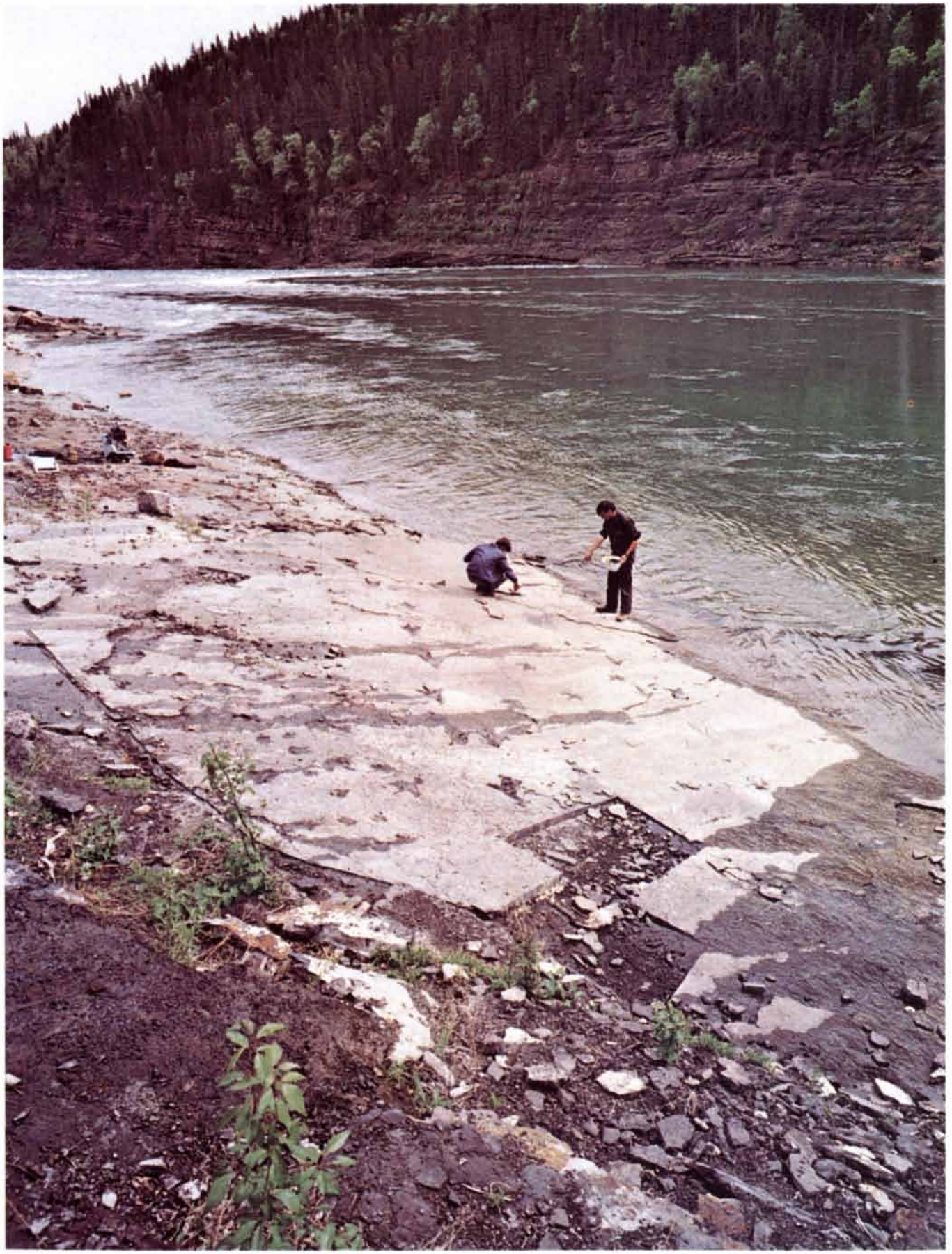
Natural Gas

The smokestacks of the world are viewed with concern in some quarters because the carbon dioxide and other gases they release into the atmosphere might inhibit the escape of heat into space. The concern is put into a new perspective by the recent report that the smokestacks are outgassed by termites, which munch huge amounts of biomass and give off various gases as by-products. Termites may contribute twice as much carbon dioxide to the atmosphere as smokestacks do, and the insects also emit significant amounts of methane (CH_4), molecular hydrogen (H_2) and other gases that are trace components of the atmosphere. The termite emissions are discussed in *Science* by P. R. Zimmerman and J. P. Greenberg of the National Center for Atmospheric Research, S. O. Wandiga of the University of Nairobi and P. J. Crutzen of the Max Planck Institute for Atmospheric Chemistry in Mainz.

The authors estimate the world's termite population at 2.4×10^{17} and calculate that the insect's efficiency of digestion is 60 percent or more. Digestion depends primarily on the decomposition of food by microorganisms in the termite's gut. On the basis of measurements made on termite colonies maintained in the laboratory the authors estimate that the gross amount of carbon dioxide produced by termites worldwide is 4.6×10^{16} grams per year, "more than twice the net global input from fossil-fuel combustion." The corresponding amount for methane is 1.5×10^{14} grams and for molecular hydrogen 2×10^{14} grams.

Zimmerman, Greenberg, Wandiga and Crutzen point out that the ecological impact of termites, as indicated by the ratio of what they eat to the net primary productivity of biomass, is greatest for wet savanna, Temperate grasslands, cultivated land in underdeveloped countries and areas that have been cleared and burned.

The authors caution that their estimates are not definitive. The laboratory measurements have an uncertainty of 50 percent and additional uncertainty arises from the possibility that not all termite species yield gases at the rates measured in the laboratory. Moreover, in a natural setting microorganisms in the soil may consume or modify some of the emissions by termites. "We estimate that all of these uncertainties may cause the actual gas productions to vary from those reported here by a factor of two."



FOSSIL-FOOTPRINT BONANZA, uncovered during five years of salvage work in advance of flooding at the Peace River Canyon dam site in British Columbia, included this array of dinosaur trackways preserved in early Cretaceous sediments. In the foreground one rectangular block of footprints has been removed. In the background investigators from the Natural History Museum of Alberta are preparing casts of other prints. Over four miles salvage workers have found 1,700 footprints representing at least 10 different dinosaur species.

The Footprints of Extinct Animals

Vertebrate animals have left their tracks in sediments ever since they first appeared on dry land 370 million years ago. Indeed, most of the known extinct species are known only from their footprints

by David J. Mossman and William A. S. Sarjeant

The skeletons of extinct land animals in museums suggest that the main fossil evidence for such animals is bones. This is not the case. The bones of the animals are very much rarer than their tracks. Indeed, many extinct land animals are known only from their tracks; their bodily remains appear never to have been preserved. In England, for example, the middle Triassic sediments of Worcestershire and the late Triassic sediments of Nottinghamshire, both laid down in the Mesozoic "age of reptiles," contain no fossils of vertebrate animals. Footprints in these same formations, however, indicate that at least eight different kinds of reptiles traversed the sediments of Worcestershire and six other kinds did the same in Nottinghamshire.

Tracks similarly testify to the earliest known invasion of dry land by vertebrates. The markings are found in the Old Red Sandstone formation of the Orkney Islands, of middle Devonian age (370 to 360 million years ago). They record the passage of a primitive fish. Like today's lungfishes, it lived at the mercy of seasonal dry spells. It left its track when it crawled on its belly overland from one pool of water to another, propelled by its stubby lobate fins.

From the perspective of the present the Devonian period may seem remote. Within the absolute-time framework of earth history, however, it is not so long ago. The point is readily appreciated if one views the 4.5 billion years that have passed since the earth formed as being a single year, with each day lasting for 12.3 million years. On such a time scale the earth's first forms of life—primitive plants resembling modern single-celled algae—appeared in the seas in early May. Many-celled forms of life, however, did not arise until early November.

By about November 20 primitive fishes were swimming in the planet's waters. Toward the end of the month their descendants ventured onto the land. By December 7 reptiles had become the dominant terrestrial animals, and by mid-December the first mammals had appeared. At about 5:00 P.M. on the last day of the year two early hominids left their footprints in a fresh fall of volcanic ash on the Laetoli Plain of Kenya. Our own genus, *Homo,* did not appear until about an hour before midnight—some 500,000 years ago. Thus the entire span of vertebrate life on land occupies less than six full weeks of an earth-history "year." Even though that important interval spans some 360 million years, it is a very small part of the history of the planet.

The Terrestrial Record

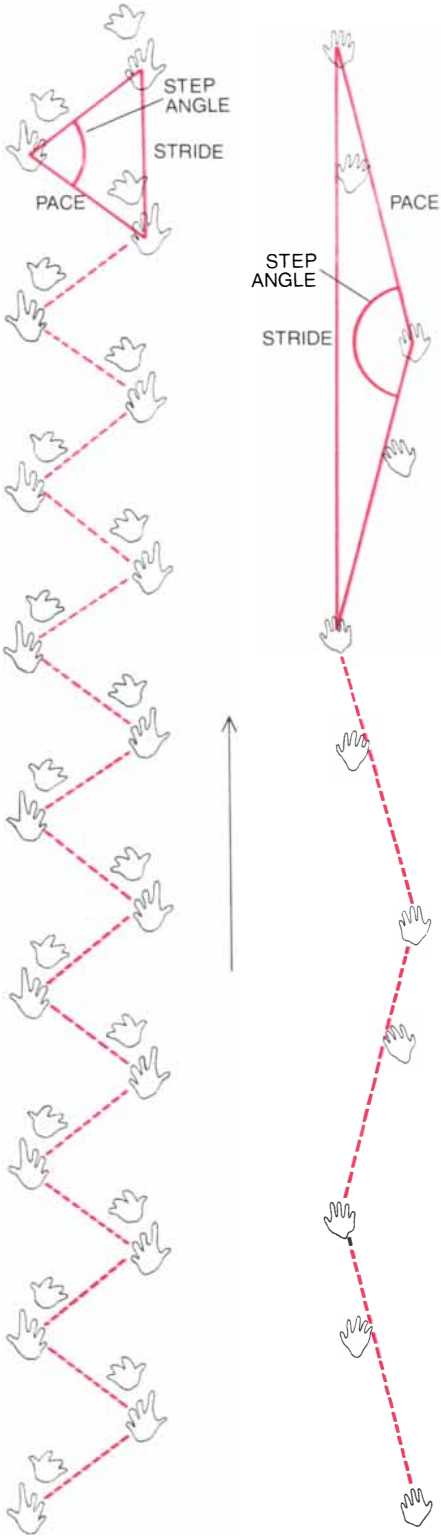
Throughout this interval two kinds of evidence were preserved that help paleontologists to interpret the animal life of the past. The remains of the animals themselves are evidence of one kind: fossilized bones, articulated or separated, with or without associated structures such as teeth, scales or feathers and, in rare instances, with fossilized impressions of skin or other tissues. From these remains one can reconstruct the animal's appearance with fair accuracy and deduce how it moved about and fed. Yet an animal has only one body and its remains, if they are preserved at all, can be found in only one place. The other kind of evidence is the animal's tracks and trails, which are a much more extended and dynamic testament. Studying these clues to an animal's behavior, the paleontologist is able, as it were, to see the animal in action.

The lobe-finned fishes that first ventured onto the land in Devonian times gave rise to four-legged amphibians. In evolutionary terms the transition was quick: the earliest amphibian tracks are found in formations of the late Devonian. Amphibian trackways are characteristically broad and show the short stride that is typical of rather inefficient quadruped locomotion. The individual prints show a full impression of the palms (indicative of a slow walking pace) and clawless digits that spread widely and are directed outward rather than forward. Amphibian footprints are abundant in the sediments of the Carboniferous period, which followed the Devonian. They are less abundant in the Permian (280 to 225 million years ago) and are rare in the fossil record thereafter, largely because the surviving amphibians preferred, as they still do, wet conditions that are unsuitable to the preservation of tracks.

The rise of the reptiles to dominance is clearly seen in the later Carboniferous and in the Permian as the number of their footprints steadily increases. Both the diversification of these animals and their progressively efficient locomotion are made evident by changes in the form of the digits, including the development of a short, thumblike fifth digit and the appearance of claws. In some reptilian lineages the trackway narrows and the stride lengthens. The animals were traveling faster but their gait remained quadrupedal. Other tracks show that by late Permian times some small reptiles maintained a quadrupedal gait when they moved slowly but adopted a bipedal one when they moved quickly. Their body pivoted at the hips, a long tail counterbalancing their forelegs and nearly erect trunk.

This was the reptile lineage that gave rise to the dinosaurs at the start of the Mesozoic era, 225 million years ago. Many Triassic dinosaurs were largely or entirely bipedal. Even in the groups that later resumed quadrupedal locomotion there was a marked disparity in limb size. The body continued to pivot on the pelvis and the hind feet were usually much larger than the forefeet. Of all fossil footprints dinosaur tracks are the most spectacular; they are found in abundance in terrestrial sediments of Mesozoic age in most parts of the world.

In time two groups of small dinosaurs evolved wings. In one group, the pterosaurs, membranes that enabled the animals to glide stretched outward from the flank of the body, supported by an elongated fourth digit of the forelimb. W. L. Stokes of the University of Utah has found fossil trails of these flying reptiles in sediments of the Morrison for-



TRACKWAY MEASUREMENTS appear in this diagram. At the left is the broad trackway characteristic of an amphibian; it shows the successive front and rear footprints of one of the amphibians that became extinct near the end of the Paleozoic era. As the nearly equilateral colored triangle indicates, both the pace and the stride of the animal were short. At the right is the narrower trackway characteristic of a reptile; it shows the successive front and rear footprints of a dinosaur. Both the pace and stride distances are proportionately much longer than those of the amphibian, an indication of more efficient locomotion.

mation, a late Jurassic deposit in Arizona. The trackways make it clear that the animals not only were capable of gliding but also could walk readily on all fours, with the wing membrane trailing on the ground behind the forefeet.

In the second group a modification of the epidermal scales led to the evolution of feathers and a more efficient means of flight. The evolution of birds from dinosaurs produced no immediate modification of the hind feet, but footprints distinctly avian in style were not long in appearing. The transition is evident in the beautifully preserved footprints of birds found in the middle Cretaceous sediments of eastern British Columbia by Philip J. Currie of the Tyrrell Museum of Paleontology in Drumheller, Alberta.

Not all the Mesozoic reptiles became bipedal. The synapsids were one persistently quadrupedal group, a reptilian lineage that was separated from the dinosaurs throughout their history. Here too, however, the stride of the animals lengthened, the trackway narrowed and the digits were directed forward, all changes that are evidence of increasingly efficient and fast locomotion. The structure of the synapsid foot became increasingly like the structure of the foot of a mammal. Indeed, in the sediments of late Triassic and early Jurassic age in the province of São Paulo the Brazilian paleontologist Giuseppe Leonardi has found numerous tracks so exactly intermediate in form between reptile and mammal footprints that they could be attributed to either group.

With the end of the Mesozoic dinosaur tracks vanish from geologic history. From the beginning of the Cenozoic era, 65 million years ago, the variety of mammal tracks corresponds to the increasing diversification of these advanced vertebrates. That diversification is one of the main evolutionary events of the Tertiary period, the first and longest part of the Cenozoic. It seems only fitting that with the succeeding Quaternary period, as geologic history comes up to the present, we find among the mammal tracks at Laetoli evidence that some of man's upright precursors have walked their way into the fossil record.

The Trail of Discovery

What is the earliest documented discovery of fossil vertebrate footprints? One might think it was made in Europe by one of the pioneers of geology, but it was not. The locale was the valley of the Connecticut River near South Hadley, Mass., where red sandstone beds of the late Triassic to the earliest Jurassic age lie exposed. The discoverer was Pliny Moody, a local farm boy, the year was 1802 and the tracks were those of small dinosaurs. At that time, however, dinosaurs were unknown and so the footprints were taken to be those of large birds. Some even went so far as to suggest that they were the tracks of Noah's raven.

A quarter century passed before a scholarly account of fossil footprints appeared. The footprints were in a sedimentary formation of Scotland, the New Red Sandstone. The age of the formation was then uncertain, but it is now known that the sediments were deposited in the final interval of the Paleozoic era. The discoverer was a local clergyman and amateur naturalist, the Reverend Henry Duncan. His description of the tracks, published in 1828, astonished the scientific world of the time.

The prints, which show the movement of an animal with a short stride and a broad trackway, are now recognized as those of a caseasaur, one of a group of large, ponderous plant-eating reptiles. Like the dinosaurs of the Connecticut River valley, the caseasaurs were still unknown, and so the tracks were attributed, quite reasonably, to some kind of turtle. Having learned of Duncan's discovery, a distinguished English geologist, William Buckland, undertook to test the turtle-track hypothesis. John Murray III, a well-known publisher, described Buckland's investigation:

"I went on Saturday last to a party at Mr. Murchison's house, assembled to behold tortoises in the act of walking upon dough. Prof. Buckland acted as master of the ceremonies. There were present many other geologists and savants, among them Dr. Wollaston. At first the beasts took it into their heads to be refractory and to stand still. Hereupon the ingenuity of the professor was called forth in order to make them move. This he endeavored to do by applying sundry flips with his fingers upon their tails; deil [devil] a bit however would they stir; and no wonder, for on endeavouring to take them up it was found that they had stuck so fast to the piecrust as only to be removed with half a pound of dough sticking to each foot. [It] was found necessary to employ a rolling pin, and to knead the paste afresh; nor did geological fingers disdain the culinary offices. It was really a glorious scene to behold all the philosophers, flour-besmeared, working away with tucked-up sleeves. Their exertions, I am happy to say, were at length crowned with success; a proper consistency of paste was attained, and the animals walked over the course in a very satisfactory manner; insomuch that many who came to scoff returned rather better disposed toward believing."

After Duncan's discovery numerous other footprints began to be reported but not necessarily accepted. For example, in 1841 on the shores below Horton Bluff in Nova Scotia, William Logan, a visiting Scotsman, found a slab of sedimentary rock bearing well-pre-

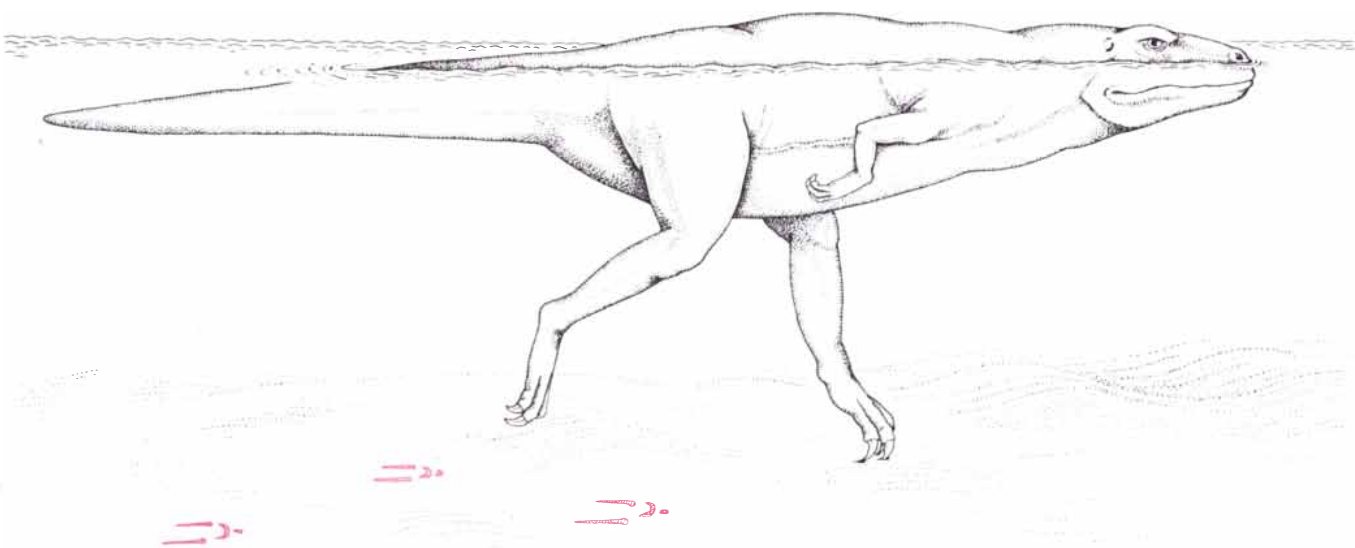
served vertebrate footprints. The formation was Carboniferous in age, a time, in the prevailing opinion of Logan's day, too early in geologic history for any vertebrates other than fishes to have existed. When, on returning to Britain, Logan exhibited his find to the Fellows of the Geological Society, the evidence was disbelieved.

Logan was not vindicated until 1872. In that year John W. Dawson, working with a founding father of geology, Sir Charles Lyell, was investigating Carboniferous strata on the east coast of the Bay of Fundy. The strata included the coalfields at Joggins in Nova Scotia, where a number of upright tree stumps had been preserved as fossils. Dawson had predicted that the fossil remains of small amphibians and reptiles might be discovered in the hollow interior of the stumps. His prediction was correct; reptile remains were indeed found inside the stumps. Perhaps the animals had been trapped in the stump hollows while searching for water or insects. In any event it was now clear that the Carboniferous fauna included terrestrial vertebrates as well as fishes.

Logan and Dawson were destined to be major figures in Canadian science. Logan was the first director of the Geological Survey of Canada and Dawson became principal of McGill University. Both men were eventually knighted.

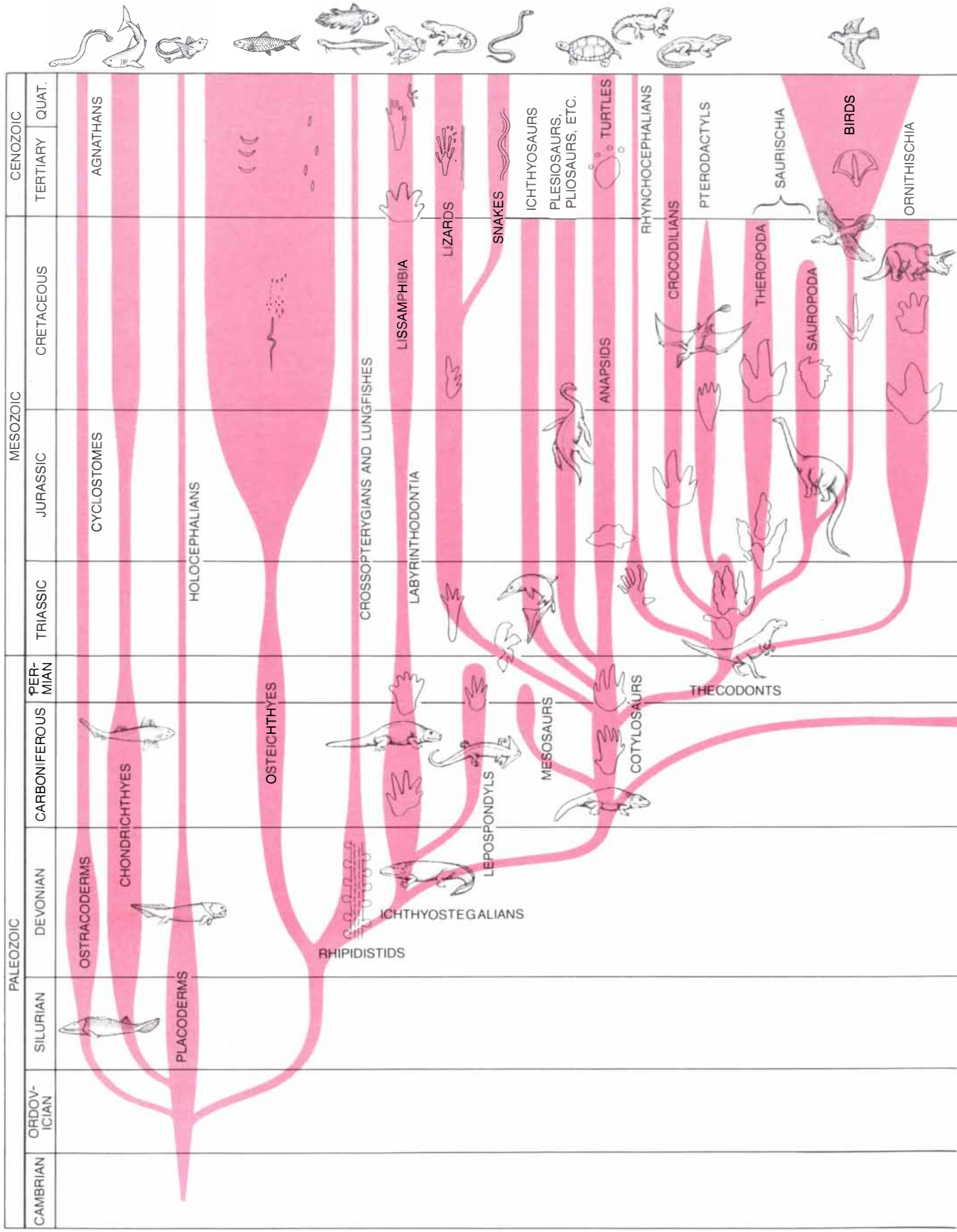
A Waning Interest

Although many descriptions of fossil vertebrate tracks were published in the second half of the 19th century, such papers appeared in diminishing numbers in the first half of the 20th. This branch of paleontology, known formally as ichnology (from the Greek *ikhnos,* track), fell progressively into disfavor among geologists. The reasons are not hard to find. Between 1930 and 1960 the

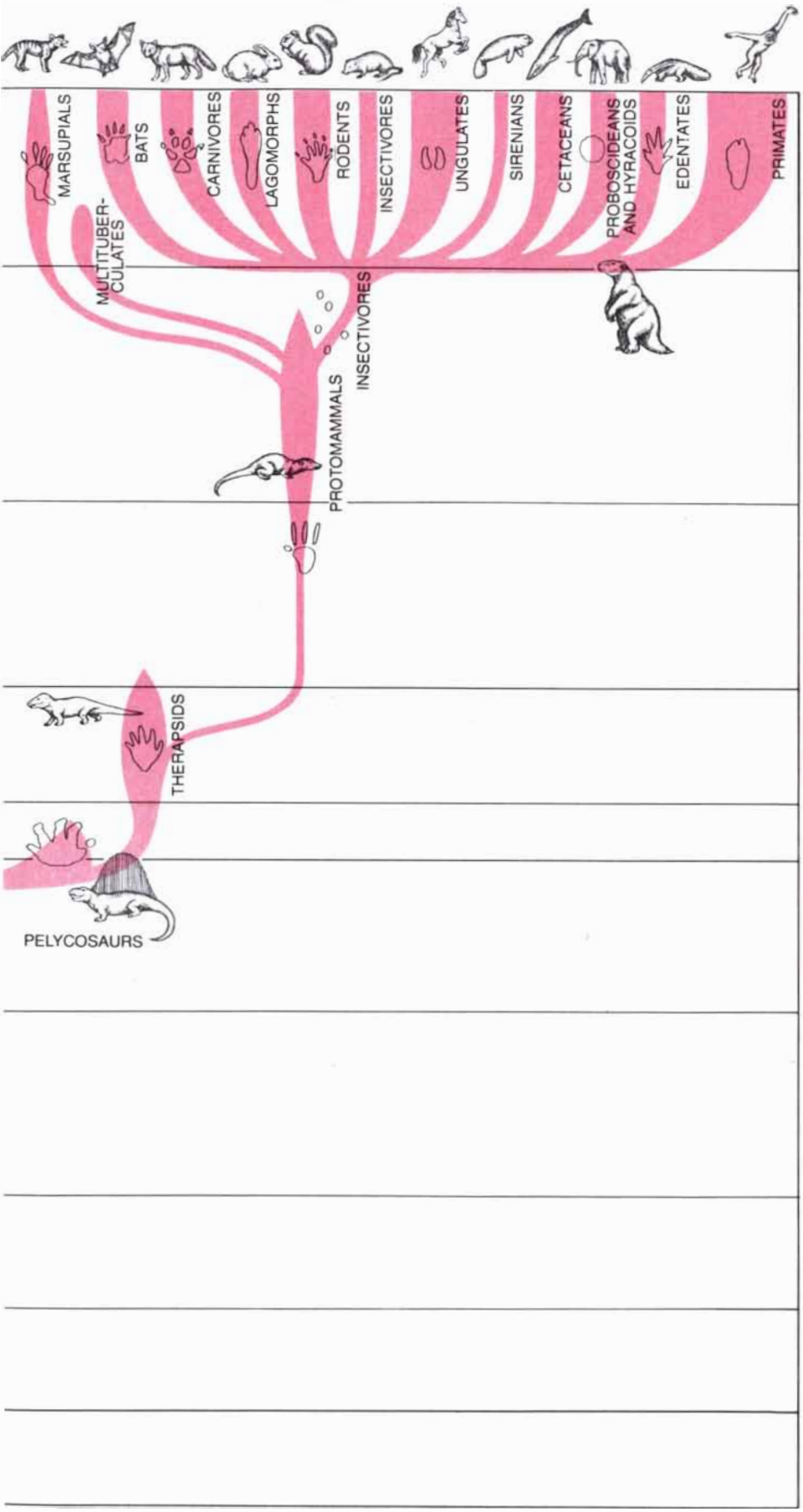


SWIMMING TRACKS of a carnivorous dinosaur found recently in sediments of Jurassic age consist of clear toe and claw marks that give no indication of foot pressure on the bottom. The depiction of the swimming carnivore is based on a restoration of *Megalosaurus* by Matthew Hyman. The tracks indicate that herbivorous dinosaurs, the carnivores' natural prey, could not escape by going into the water.

DINOSAUR TRACKS in Connecticut River valley red sandstones of early Jurassic age indicate that the animal was essentially bipedal, as at the left, but that sometimes it walked on all fours, as at the right. This is one of the many extinct reptiles known only from their tracks; from these Richard Lull of Yale University reconstructed the animal as it is seen here. It was one of the early ornithopods, a herbivore.



VERTEBRATE EVOLUTION has contributed numerous tracks to the fossil record, including the earliest terrestrial trackway: that of a crossopterygian fish left in the Old Red Sandstone of the Orkney Islands in middle Devonian times (*fifth band from left*). Thereafter, as



the representative footprints indicate, a succession of amphibians, reptiles, birds and mammals have left their traces. The last print, at the far right, was made by a hominid in Kenya.

few papers in the field were written primarily by amateurs rather than professionals, appeared in relatively obscure journals and were poor in quality. The authors' descriptions of the tracks they had discovered were woefully inadequate and their findings were seldom related in any systematic way to tracks that had been recorded previously. Photographs of tracks, if they were included at all, generally were made with overhead illumination rather than with the side lighting needed to bring out shallow impressions or the low relief of casts. More often the illustrations were not even poor photographs but were inadequate line drawings. The authors almost never stated where the specimen could be examined by others, chiefly because most of the specimens were never placed in study collections. Here the fault was not necessarily the authors'. Since footprint slabs are large and awkward to handle, they tend to be unpopular with museum curators. After the slabs are received they are likely to end up outside the museum, either flat on the ground, where they are worn away by the feet of visitors, or propped up, where exposure to the elements eventually destroys them.

In the 1950's the investigations of the German paleontologist Walter H. Häntzschel and the missionary endeavors of his disciple Adolf Seilacher resulted in a surge of interest in the fossil tracks left by marine organisms. Even then there was no parallel revival of interest in the tracks of vertebrates. Today only a handful of investigators are concerned with them, most notably Leonardi and Rodolfo Casaniquela in South America, Donald Baird of Princeton University, Justin Delair in Britain, Georges Demathieu and Albert F. de Lapparent in France, Hartmut Haubold in East Germany and O. S. Vialov in the U.S.S.R.

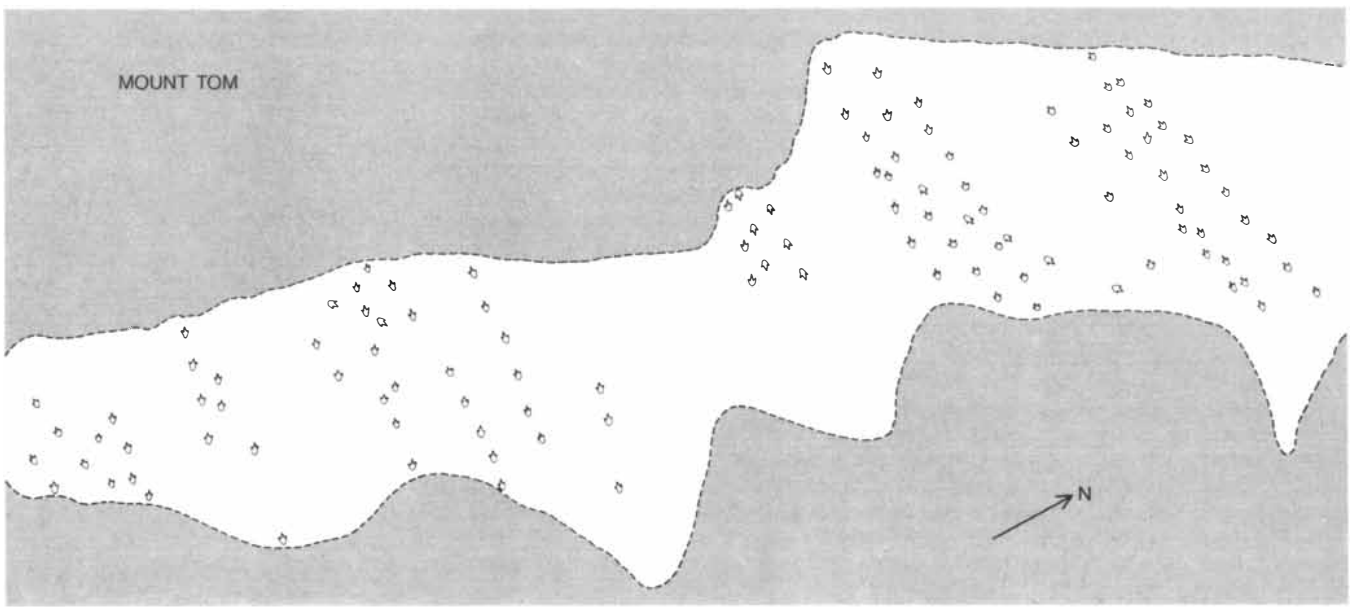
In addition to the problems of an inadequate scholarly literature the student of fossil footprints is confronted by other difficulties. Foremost among them is the initial character of the tracks themselves. Ideally for clear foot impressions to be formed there first must be a moist, fine-grained and cohesive bed of sediment for the animal to traverse. This it should do slowly, leaving detailed impressions of its forefeet and hind feet (assuming it is quadrupedal). Under such ideal circumstances even the exact outlines of claws or nails, the shape of pads and the pattern of scales may be preserved.

Actually few such high-quality impressions are found. If the sediment is too coarse, it will not retain details. If it is too wet, its deeper hollows will fill with water, distorting the shape of the print. If it is too dry, it will not be cohesive enough to preserve the impression. A strong wind may obliterate the print,

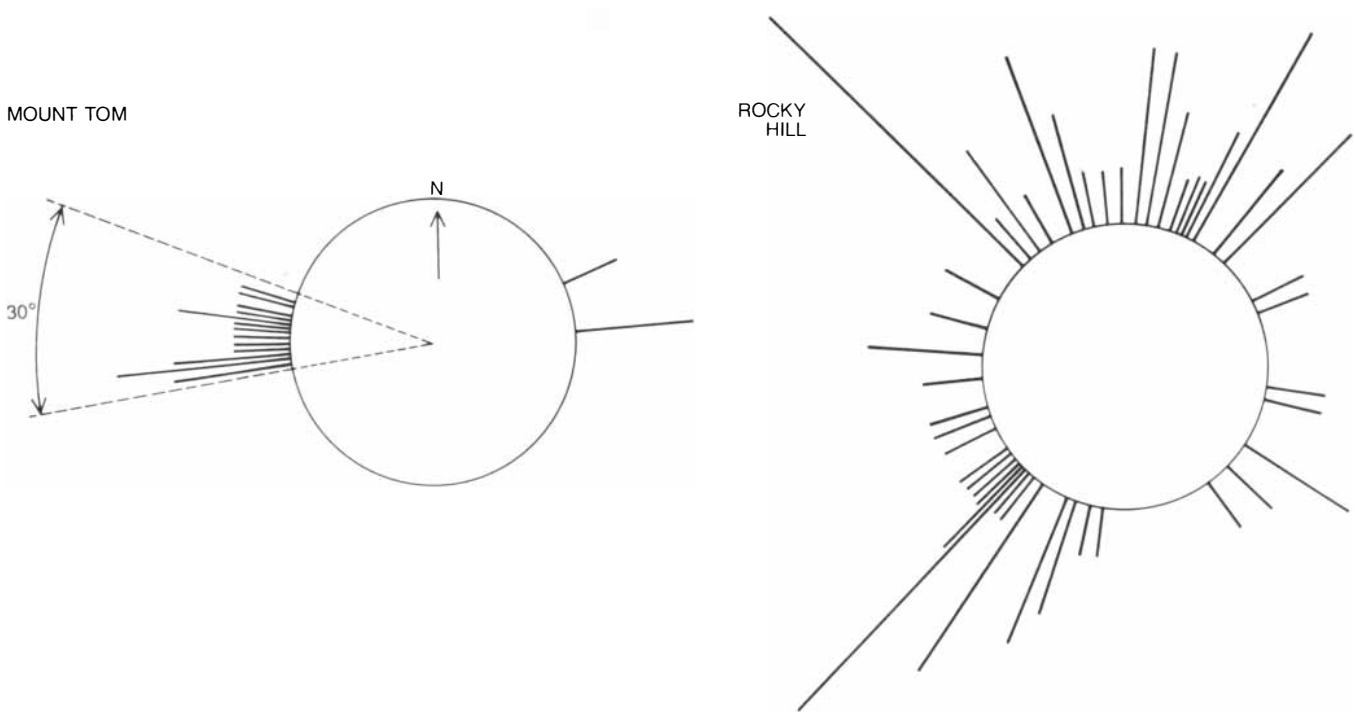
and if it was made near the seashore, a rising tide may do the same. Even the deposition of sediment on top of the tracks, which is clearly crucial to their preservation, may mean their obliteration. If the new sediment is too much like the imprinted one, the two strata will tend not to separate and so the tracks may never be discovered.

The most favorable conditions for track preservation exist after high waters have receded, leaving fine-grained sediment freshly exposed. At the seashore this means after the highest level of a spring tide. Inland it means after a rainy season has raised the stream or pool to its highest level. Imprints in such a moist surface will solidify as the sediment dries and will be buried by the next deposition of sediment. Even then it may be not the imprint that is preserved but rather its cast, particularly if the newer sediment is coarser and more cohesive than the older one. Such casts must be sought on the underside of sedimentary layers and may escape detection unless the layer has been turned over by the collapse of a cliff, say, or in the course of a commercial quarrying operation.

Lastly, the probability that a terrestrial animal track will be preserved and later recognized is directly proportional



SOCIAL BEHAVIOR among dinosaurs is suggested by the orientation of these fossil trackways of Triassic age in the Connecticut River valley near Holyoke, Mass. Analysis of the tracks by John H. Ostrom of Yale reveals that at least 19 carnivorous dinosaurs of the genus *Eubrontes* had crossed damp sand traveling from east to west. The common direction suggests that the animals were hunting in a pack.



ROSE DIAGRAMS contrast the nonrandom orientation of the *Eubrontes* tracks, at the left, with the far more random orientation of more than 1,000 similar dinosaur footprints at another Connecticut River valley site, Rocky Hill, at the right. Not all the recorded tracks at either site are shown. At the left, however, 70 percent of all the trackways lie within a 30-degree arc. The randomness evident at the right suggests that the Rocky Hill trackways were imprinted over a longer period of time, but this does not rule out gregarious activity.

to the animal's size. Large animals need to range widely for food and to visit water holes frequently. Moreover, their footprints are more deeply impressed and therefore less readily destroyed. Although small animals are much more numerous than large ones, they range less widely, require less food and water and leave shallower footprints. That is why a dinosaur's tracks are more likely to be preserved (and far more likely to be recognized) than a lizard's.

The Analysis of Tracks

Ideally a trackway suitable for analysis should consist of a sequence of at least three footprints or casts. Among quadrupeds locomotion is initiated with the hind foot on one side, followed by the forefoot on the same side and then by the hind foot and the forefoot on the other side. In fast locomotion two or three feet touch the ground simultaneously; in slower locomotion three feet or all four do so. In jumping locomotion all four footprints are set close together; they do not overlap. Jumping tracks are extremely rare in the fossil record; some were recently found by Leonardi and one of us (Sarjeant) in a Brazilian formation of Mesozoic age. The tracks of bipeds normally show the left and right feet alternating; the footprints are rarely side by side.

The analyst's first job is to make four basic measurements of a trackway. The first is a measurement of stride, or forward movement; it is made from a fixed point on one footprint to the same point on the next print of the same foot. The second is a measurement of pace; it is the distance between the right forefoot and the left forefoot and between the right hind foot and the left hind foot. The length of the stride is usually identical for the forefeet and the hind feet, but the length of the pace may differ greatly. The third is a measurement of pace angulation, or step angle; it is the angle formed by joining the midpoint of three successive hind-foot prints or three successive forefoot prints. The fourth is a measurement of trackway breadth. As we have noted, the mark of an inefficient walker is a broad trackway and a short stride. Conversely, a narrow trackway and a long stride are evidence of an efficient walker moving fast. A trackway moderate in both width and stride, however, can be evidence of either a not too efficient walker or an efficient walker moving at a relatively low speed.

For quadruped tracks a further important measurement is one from the midpoint between two consecutive hind-foot impressions to the midpoint between two similar forefoot impressions. This distance is actually between the hipbone socket and the shoulder socket of the living animal. It corresponds approximately to the length of the animal's trunk (head, neck and tail omitted).

Because the margins of both the footprints and the casts may grade smoothly into the surrounding matrix exact boundaries are often difficult to determine. As a result inaccurate measurements of footprints are all too common and the field investigator must guard against them. A further hazard arises from the fact that in some instances the weight of the animal making the track was impressed not only on the surface layer of the sediment but also on layers farther down. Care must therefore be taken to distinguish between an actual footprint and a "ghost print": a subsurface impression. Ideally the investigator should examine not only the track but also a cast of it. If what has been found is itself a cast, he should examine a mold made from it (which, of course, re-creates the missing track). If it is a true footprint (a mold), then a cast should be made of it. Such casts are made by applying plaster of Paris, latex or fiberglass to the footprints in the field. They provide a valuable record for subsequent study.

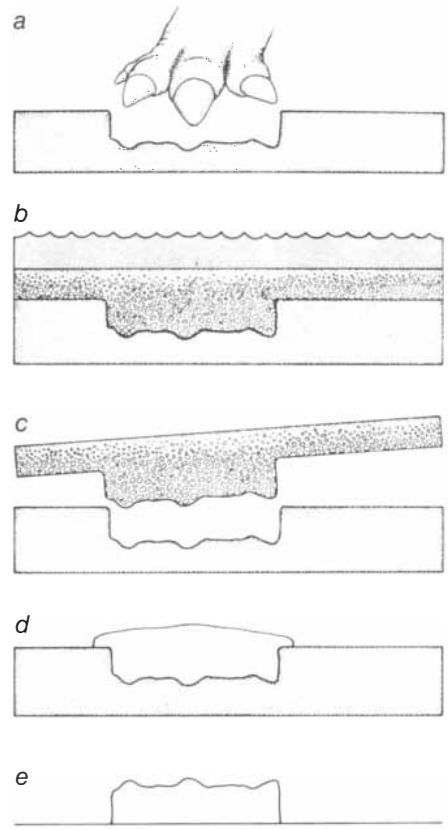
Another aid to track analysis is the measurement of the degree of overlap between the forefoot and hind-foot impressions. With the exception of certain flexible-bodied animals such as salamanders, the extent and consistency of overlap yields a second measurement of the distance between the quadruped's hip and shoulder. Still another useful measurement determines the position of the footprints with respect to the midline of the trackway; it can yield information on a quadruped's gait. For example, if the gait is a sprawling one, the midline of the trackway will not correspond exactly to the midpoints between paired footprints.

Reconstructing Unknown Animals

The hazards of attempting the reconstruction of an animal known only from its fossil footprints are well demonstrated by a 19th-century example. The animal is *Chirotherium,* from the Greek roots for "hand animal." Its tracks were first discovered in 1834, imprinting a red sandstone in the central German state of Thuringia dating to the Triassic (from 225 to 195 million years ago). The hind-foot prints are about the size of a man's hand; four of the digits are directed forward but the fifth is set at an angle to the other four, much as the thumb of the human hand is. Some hind-foot prints were associated with smaller but very similar impressions that were taken to be forefoot prints.

Even though the dominant Triassic land animals were reptiles, the Thuringian fossil footprints were at first attributed to some kind of ape or bear. What blinded those who arrived at this conclusion was a misinterpretation: they identified the angled digit as a thumb even though it was on the outside of the print. Indeed, some early reconstructions pictured *Chirotherium* as a toadlike amphibian that walked cross-legged; at least this served to put the "thumb" in a normal position. Addressing the British Association for the Advancement of Science in 1874, the zoologist L. C. Miall demonstrated that the cross-legged reconstructions were false and proposed instead that the tracks were those of dinosaurs. That the tracks were at least reptilian became evident when it was recognized that the *Chirotherium* trackways were narrow. Amphibian trackways are invariably broad.

The most convincing evidence for the tracks having been made by a reptile was the fact that the best-preserved *Chirotherium* prints showed the impressions of scales. Moreover, they had strong claws, suggesting the reptile was carnivorous. Track analysis suggested the body of the animal was about a meter long and, since the gait was predominantly bipedal, was probably balanced in an upright position by a tail of equal length. Making allowance for the ani-



FOOTPRINT PRESERVATION requires, once a print has been made (*a*), that the indented sediment be filled with sediment of contrasting texture (*b*) and that both sediments harden. The two sediments may later split apart. If only the filling sediment is preserved (*c*), it is a natural cast. If only the indented sediment is preserved (*d*), it is a natural print and can be used to make casts (*e*).

BIRD FOOTPRINTS appear scattered across an exposure of Mesozoic sediments in British Columbia. This is the famous Peace River Canyon site, where large numbers of dinosaur tracks were discovered in 1922. The shore birds that left these tracks are otherwise unknown in the fossil record. Their footprints are the only traces of bird life found in early Cretaceous strata.

GROUND-SLOTH TRACK of elephant-size footprints crosses this exposure of the Río Negro formation in Argentina. The track was made in late Pliocene times, 3.5 million years ago.

mal's head and neck, this gave *Chirotherium* a length from the tip of its snout to the tip of its tail of about two and a half meters.

It is now known the animal that made the tracks was one of the pseudosuchian reptiles, an important early group ancestral to both the dinosaurs and the crocodiles. Although nearly two dozen species of *Chirotherium* are now recognized on the basis of footprint form, only a few fossil skeletons of corresponding types are known. One of them, *Ticinosuchus,* discovered less than 20 years ago in a formation of middle Triassic age in Switzerland, is a prime candidate for being the animal responsible for the tracks that puzzled the 19th-century paleontologists.

Just as the pattern of scales in certain *Chirotherium* footprints helped to clinch the animal's identity as a reptile, so the prints of other animals can reveal details of soft parts that are only rarely fossilized. An example is hadrosaur tracks found in late Cretaceous formations in the Canadian province of Alberta. Hadrosaurs were large, bipedal herbivorous dinosaurs. The Alberta footprints show that, like elephants of today, these animals had evolved pads on the soles and digits of their hind feet to cushion their great weight.

Tracks and Behavior

Fossil footprints have proved to be both informative and corrective on the individual and group behavior of the animals that made them. For example, the dinosaurs of the sauropod group, first described in 1842 by the British zoologist Richard Owen, include the largest land animals that ever lived. For more than a century it was assumed these massive quadrupedal herbivores, with their long neck and long tail, must have been habitually aquatic, depending on their buoyancy in water to support their enormous weight. In fact, it was seriously questioned whether they could move about on land at all. Then Roland T. Bird of the American Museum of Natural History discovered sauropod tracks in two Texas formations of Cretaceous age: at Paluxy Creek in 1944 and at West Verde Creek in 1954. His finds decisively settled the issue. A brontosaur trackway from Paluxy Creek, now displayed in the American Museum, is particularly dramatic. It is a narrow trackway, with huge footprints deeply impressed in what could only have been a mud flat either partially or completely exposed to the air. The animal's steps, one fully two meters from the next, are unmistakably those of an efficient terrestrial walker. So much for the assumed habitual aquatic habitat. The drama is supplied by something else. Superposed on the brontosaur tracks are the footprints of a large carnivorous

bipedal dinosaur—possibly *Allosaurus*—moving in the same direction. It seems likely that the carnivore was pursuing the herbivore.

Bird's other site, West Verde Creek, shows in addition to sauropod footprints a continuous deep groove. It evidently was made by the dinosaur's dragging tail, an indication that the mud flat was completely exposed and additional confirmation that sauropods could walk entirely unsupported by water. This does not mean, however, that these giant dinosaurs never went swimming. Another of Bird's findings is informative on the point. This particular sauropod track consists solely of the impressions of forefeet except for a single print of a left hind foot. Apparently the animal's hindquarters and tail were afloat in water shallow enough to enable it to walk along the bottom with its forefeet, just as the hippopotamus does today. Then, when turning left, the sauropod made the extra track as it kicked the bottom with one hind foot.

In depictions of these giant herbivores it has been traditional to show them afloat, contentedly browsing on aquatic plants while their natural enemies, the carnivorous dinosaurs, stand in frustration on the shore, unable to pursue their prey. Tracks in the State Dinosaur Park at Rocky Hill, Conn., suggest a different scene. There hundreds of footprints, most of them of the predatory genus *Eubrontes,* are preserved in lake sediments of early Jurassic age. Some prints are perfectly clear, others are faint; some are deeply imprinted, others are shallow. All show the three digits characteristic of a bipedal carnivorous dinosaur. One trackway in particular, discovered by Walter P. Coombs, Jr., of Western New England College, sheds light on the swimming ability of carnivorous dinosaurs. It consists of eight successive footprints averaging just over one meter between steps. The step with the right foot forward is consistently about 20 centimeters longer than the step with the left foot, which suggests to Coombs that the swimming movements had a kind of galloping rhythm. Further, some of the prints show exceptionally clear claw marks, precluding the possibility that they are ghost prints, and yet give no indication that the animal was putting any weight on its feet. According to Coombs's interpretation, the animal was swimming in shallow water, kicking the bottom with the tips of its toes. Moreover, the track sequence begins and ends abruptly, indicating that the swimming motion intermittently lifted the animal clear of the bottom. Coombs concludes that the traditional image of the frustrated carnivore standing at the water's edge is in need of revision.

Fossil evidence that dinosaurs may have lived in social groups has until recently been quite limited. One example is an accumulation of more than 20 dinosaur skeletons in an early Cretaceous formation at Bernissart in Belgium. All the animals were large bipedal herbivores of the genus *Iguanodon.* It has long been suggested they had died when a whole herd tumbled into a limestone ravine, mute evidence that the iguanodons traveled in herds. Recently David Norman of the University of Oxford has demonstrated that the animals were not all found at the same level in the formation and that the formation is actually a series of sediments. His findings cast serious doubt on the herd interpretation.

Other Social Dinosaurs

A second example is an accumulation of dinosaur skeletons found in a Triassic formation at Ghost Ranch, N.M., in 1947. All the animals belonged to the same small carnivorous genus *Coelophysis.* Again, some see this as evidence that these carnivores hunted in packs and others view it as an unexplainable freak aggregation. Indeed, by their very nature fossilized skeletal remains cannot be expected to yield direct evidence on an animal's behavior.

One source is the nests made by dinosaurs. The discoveries in Mongolia in the 1920's of dinosaur nests complete with unhatched eggs, symmetrically arranged like bird eggs, was the paleontological sensation of its day. One nest, that of a small quadrupedal dinosaur of the genus *Protoceratops,* provided evidence that, unlike such modern reptiles as the turtle, these Cretaceous animals may have guarded their unhatched eggs. Beside the nest was the fossilized skeleton of a predatory coelurosaur—a small bipedal carnivorous dinosaur—that had apparently met a violent end. Its discoverer assigned the fossil to a new genus, appropriately named *Oviraptor.*

A second nest discovery, made only two years ago in Montana, continues the story of dinosaur sociality. In the Two Medicine formation, a deposit of Cretaceous age, John R. Horner of Montana State University found the skeletons of 18 infant hadrosaurs. The young animals were judged to be one or two weeks old. Their continued association with a nest strongly suggests sibling social behavior and also implies the possibility of parental attendance on the young after their hatching.

What social behavior can be deduced from footprints? For one thing, the first clear evidence for dinosaur herd behavior has come from the same two Texas sites studied by Bird. Some 20 of the sauropod tracks he uncovered were parallel to one another, indicating movement in a common direction, and few of the footprints overlapped. Moreover, the prints differed in size, showing that the herd included both young animals and mature ones.

A few years after Bird's discovery John H. Ostrom of Yale University uncovered 25 sets of dinosaur footprints in an early Cretaceous formation at Lake Eanes in Texas. All were made by bipedal herbivores, probably of the genus *Camptosaurus.* All were moving in the same direction: to the northwest. Ostrom has also found evidence suggestive of hunting-pack behavior among carnivorous dinosaurs preserved in the Triassic sediments of Mount Tom in Massachusetts. There on a bedding-plane surface are at least 19 separate tracks of the dinosaur genus *Eubrontes.* Again, all the animals were moving in the same direction. Evidently they were walking abreast, because none of the trackways overlap.

Thus the evidence of the footprints makes it clear that at least some herbivorous and carnivorous dinosaurs were gregarious. This in itself is an advanced mode of social behavior without close parallel among the reptiles of today. On the basis of Bird's Texas tracks Robert T. Bakker of Johns Hopkins University has even suggested these sauropods traveled in "structured" herds, with the juveniles at the center and the adults forming a protective perimeter. Supplementary evidence indicating that the structured arrangement was deliberate rather than a matter of chance would be most welcome.

A Canadian Bonanza

The inventory of fossil footprints has been greatly enriched in the past few years by urgent work on the middle Cretaceous sediments of the Peace River Canyon in British Columbia. The 17-mile canyon has been known to be rich in tracks ever since a pioneer paleontologist, Charles M. Sternberg of the Royal Ontario Museum, uncovered 400 dinosaur footprints there in 1922. He published his findings in 1930, naming eight new types of footprint, all made by dinosaurs and all but one showing a bipedal gait.

In 1975 construction began on a large dam downstream from the site of Sternberg's discoveries. This spurred efforts by the Natural History Museum of Alberta to rescue at least some of the other fossil footprints presumed to exist in unsurveyed parts of the canyon before its inevitable flooding. Five seasons of work under the leadership of Philip Currie have brought spectacular results.

Although only the lower four miles of the canyon could be examined, more than 100 trackways were found. They included a total of 1,700 dinosaur footprints representative of at least 10 different species, a series of bird footprints and a single turtle trackway. More than 1,000 of the footprints have now been mapped and almost as many have been measured. Casts have been made of

some 200 of the prints, and 90 prints have been removed for further study and eventual public display.

Among the splendid dinosaur prints is an entire growth series—from juvenile to adult—of an early hadrosaur. The series shows how the shape of the animals' feet changed as they matured and their weight increased, a record unprecedented in ichnology. A herd of these herbivores appears to have traveled across the Cretaceous sediments, its members spread out along a broad front and sometimes walking side by side. Judging by the frequent overprinting of juvenile footprints on adult ones the younger animals followed along after the adults. One series of tracks shows that 11 animals proceeded southward for a time and then suddenly switched to travel eastward. (The preservation of a single hadrosaur trackway moving in the opposite direction, from south to north, shows that the left turn of the others was not dictated by some natural obstruction.) A close study of four of the 11 trackways has even revealed that when one of the hadrosaurs lurched sideways, three of the others veered away to avoid a collision.

The Peace River prints have also added to the evidence that carnivorous dinosaurs, small to medium in size, hunted in packs. For example, at one location six trackways of medium-size carnivores are all headed in the same general direction, although two of the tracks cross each other twice. In contrast the trackways of the larger carnivores—probably tyrannosaurs—found in the

TWO TEXAS TRACKWAYS record a dramatic Mesozoic episode. Discovered by Ronald T. Bird at Paluxy Creek in 1944, the set of larger footprints are those of a quadrupedal herbivorous dinosaur, probably one of the brontosaurs. The smaller three-toed footprints visible to the left of the main track were made by a bipedal carnivorous dinosaur, possibly of the genus *Allosaurus*. The carnivore was moving in the same direction as the herbivore, probably in pursuit.

canyon are those of individuals or at the most pairs. Thus there is still no evidence from any footprint locality in the world to suggest that the larger carnivorous dinosaurs hunted in packs.

Footprints can also yield evidence on how fast the animal that made them was traveling, although care in interpretation is called for. A case in point is provided by dinosaur tracks discovered early in the 1960's at an English quarry. There, near Herston in Dorset, the Purbeck Beds, a late Cretaceous formation, lie exposed. On the quarry floor were found two parallel lines of footprints of a bipedal herbivore. The initial interpretation was that these prints had been made by the opposed feet of a single animal. The modest length of the stride suggested the dinosaur must have moved quite ponderously; one analyst even proposed that since the animal was clearly putting its weight on the ball of its foot, it had been ascending what was then a steep incline at a slow walk. This picture fitted the then current image of dinosaurs as sluggish, slow-moving animals. When the quarrying was extended, however, it was discovered that the two tracks diverged; they had been made not by one dinosaur climbing slowly uphill but by two much faster-moving dinosaurs that were possibly running side by side.

Evidence from the Peace River tracks makes it clear that some dinosaurs were able to move quite rapidly. With a formula developed by R. McNeill Alexander of the University of Leeds the speed of the animal's movement can be calculated from the observed pattern and spacing of its footprints. The highest speeds recorded by the Peace River tracks are those of medium-size carnivores: 16.5 kilometers per hour (10.3 miles per hour). This is close to the highest running speed that can be attained by a human being. The other carnivores traveled at lower speeds: between 6 and 8.5 k.p.h. (3.8 and 5.3 m.p.h.). The herbivores moved more slowly. Their top speed was the same 6 k.p.h. that was the slowest pace of the carnivores.

Overall, then, recent footprint studies demonstrate that the long-prevailing picture of dinosaurs as essentially solitary, slow-moving and even clumsy and slow-witted animals is far from true. It appears likely that when equally careful analyses are done of the fossil tracks of other terrestrial vertebrates—for example amphibians, other kinds of reptiles and the mammals of the early Cenozoic—new information about their behavior will be forthcoming and will prove to be equally at variance with past conclusions. Are we at the beginning of a renaissance in the study of fossil footprints? Certainly it seems safe to predict that over the next few decades this area of investigation will be one of the most exciting in paleontology.

NMR Spectroscopy of Living Cells

The chemical reactions of metabolism have traditionally been studied one by one in the test tube. With nuclear-magnetic-resonance (NMR) spectroscopy reactions can be followed as they occur in living tissue

by R. G. Shulman

In the middle of the 19th century Louis Pasteur showed that living cells conduct specific chemical reactions: he demonstrated that yeast, which ferments carbohydrates into substances such as carbon dioxide and ethyl alcohol, consists of living cells. Then at the end of the century the brothers Eduard and Hans Buchner ground up yeast cells and showed that the carbohydrates they added to the resulting solution were turned into the same end products. In this way they established that substances inside yeast cells catalyze chemical reactions. The discovery set a strategy for the advance of biochemistry. Throughout this century biochemists have tried to understand how substances extracted from cells and purified could account for the chemical transformations of which a cell is capable. Different investigators have traced different reactions, clearing paths through the welter of events that make up intracellular metabolism. Although some aspects of metabolism remain unexplored, the cleared reaction pathways, crossing, joining and circling back on themselves, are beginning to match the complexity of the living cell itself.

To the extent that objections persist about the validity of modern biochemistry, they continue to be about reducing the processes of life to sequences of chemical reactions. "The reactions may take place in the test tube," one hears, "but do things really happen that way inside the living cell? And what happens in multicellular organisms?" Here I shall describe how one technique is beginning to answer these questions by detecting chemical reactions as they occur inside cells, tissues and organisms including man. The technique is nuclear-magnetic-resonance (NMR) spectroscopy. It relies on the fact that atomic nuclei with an odd number of nucleons (protons and neutrons) have an intrinsic magnetism that makes each such nucleus a magnetic dipole: in essence a bar magnet. Such nuclei include the proton (H-1), which is the nucleus of 99.98 percent of all hydrogen atoms occurring in nature, the carbon-13 nucleus (C-13), which is the nucleus of 1.1 percent of all carbon atoms, and the phosphorus-31 nucleus (P-31), which is the nucleus of all phosphorus atoms.

In NMR spectroscopy two fields are applied to cells, to tissue or to parts of a living organism. The first field is a strong magnetic field. It causes the nuclear dipoles (that is, the hydrogen-1, carbon-13 and phosphorus-31 nuclei in the sample) to orient themselves so that the dipole of each nucleus is aligned either with the field or against it. Alignment with the field is a state in which the nucleus stores less energy than it does when it is aligned against the field.

Then the second field is applied. It consists of electromagnetic radiation in the radio-frequency part of the spectrum. For any particular strength of the magnetic field in which the sample is placed there is a particular frequency of the electromagnetic radiation for which each photon, or quantum of the radiation, will carry precisely the right amount of energy to allow the alignment of a certain type of nucleus to "flip." Hence if the magnetic field strength is held constant and the radio frequency is varied (or conversely the radio frequency is held constant and the magnetic field strength is varied), there comes a time when those nuclei resonate: they absorb the radio photons. For example, in a magnetic field with a strength of 84,000 gauss the hydrogen-1 nucleus resonates at a frequency of about 360 megahertz (360 million cycles per second), the phosphorus-31 nucleus at about 146 megahertz and the carbon-13 nucleus at about 90 megahertz.

In cells, tissues and organisms the precise frequency of the resonance depends on a further circumstance: the chemical environment of the nuclei. The frequency has a "chemical shift," that is, it differs from what it would be if the nuclei were isolated, because in cells, tissues and organisms the nuclei are a constituent of atoms and the atoms are constituents of molecules. Consider an NMR spectrum showing the resonances (in the form of peaks, or absorption lines) in a sample of the amino acid glutamate. Although carbon 13 has a natural abundance of only 1.1 percent, the peaks from C-13 in the five carbon positions in the glutamate molecule are clearly apparent, and if the spectrum is obtained by varying the magnetic field strength, they extend over a range of about 200 parts per million (.02 percent of the field strength). Two of the carbon atoms in the glutamate molecule are each bound to two oxygen atoms (COO^-); their peaks are close together. These peaks are separated quite widely, however, from the peak of a carbon atom that is bound to a hydrogen atom and from the peaks of two carbon atoms that are each bound to two hydrogen atoms.

A given peak in an NMR spectrum can be assigned to a carbon atom at a particular position in a given molecule by a variety of techniques, the most direct of which is to measure the spectrum after a particular position has been labeled with carbon 13 so that the number of C-13 atoms at that position in a sample of the substance is far greater than one would predict from the natural abundance of C-13. The absorption peak from carbon at that position will be correspondingly enhanced. The technique shows one of the advantages of employing carbon 13 to study the chemical reactions of metabolism, namely that it is possible to follow the C-13 introduced as a label as it passes through various positions in various molecules.

The phosphorus-31 nucleus has different advantages. Since the P-31 nucleus is 100 percent of all phosphorus nuclei, no labeling technique is required. Moreover, the P-31 nucleus gives strong NMR signals from a small number of important compounds. When one applies NMR spectroscopy to a suspension of the bacterium *Escherichia coli,* one typically detects peaks from the three phosphorus atoms in adenosine triphosphate (ATP), a molecule in

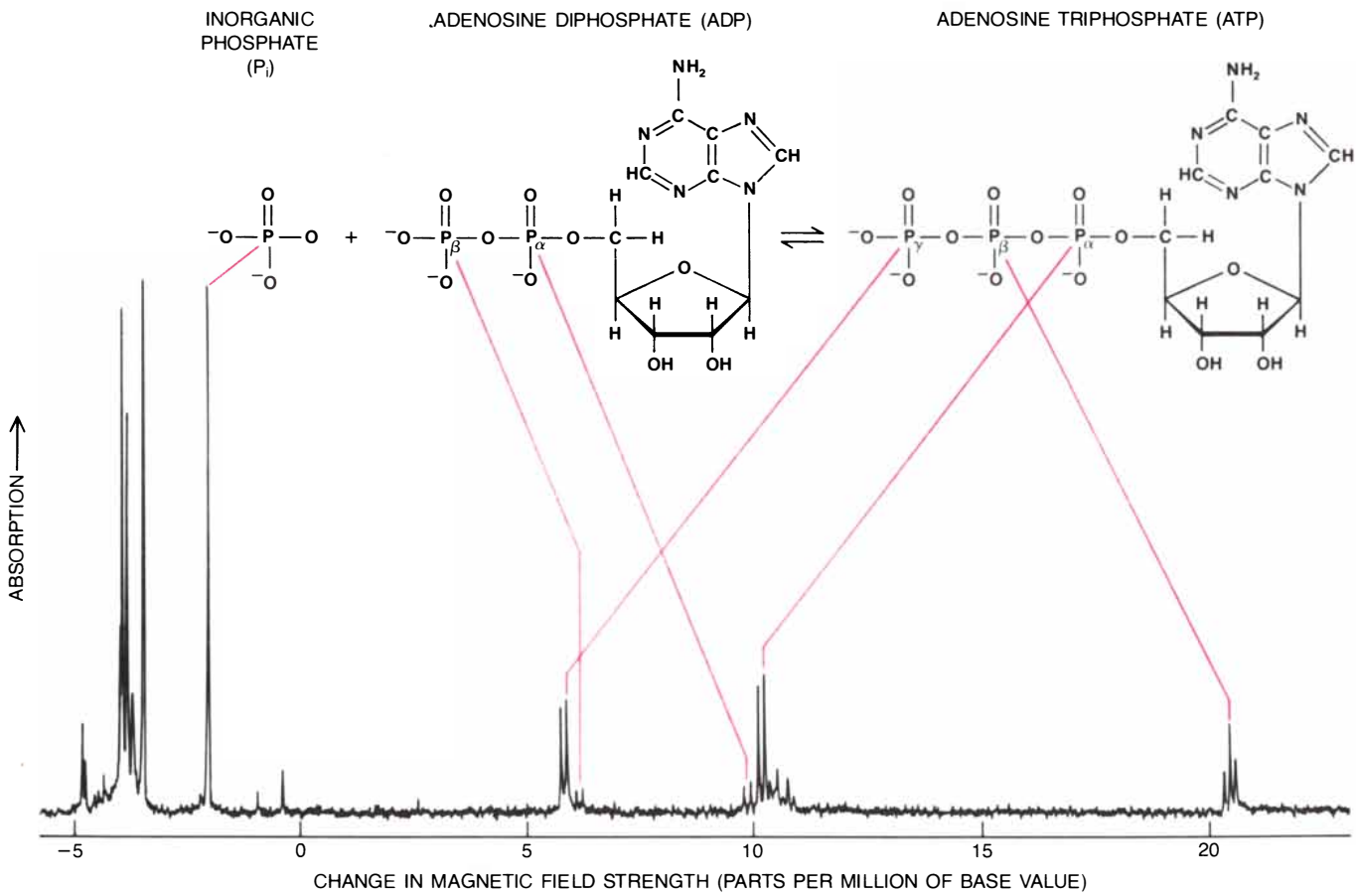
which the cell stores chemical energy. One detects a peak from the phosphorus atom in inorganic phosphate (P_i), which is broken off ATP when its energy is tapped. One detects peaks from the two phosphorus atoms in nicotinamide adenine dinucleotide, a molecule that takes part in intracellular oxidation. Sometimes one detects peaks from adenosine diphosphate (ADP), which has one less phosphorus atom than ATP; from phosphoenol pyruvate, which the cell can tap for P_i to convert ADP into ATP, and from polyphosphate chains.

When the bacteria are living off a supply of the sugar glucose, they utilize the sequence of metabolic reactions known as glycolysis or the glycolytic pathway. Molecules that arise along the sequence, in particular fructose-1,6-biphosphate (FBP), can then be detected too. Such molecules give sharp NMR peaks because each molecule is rather small (a typical one has less than 1,000 times the mass of a hydrogen atom) and small molecules in solution tumble rapidly (on the order of 10^9 times per second). The rapid tumbling means the chemical bonds in the molecule sweep through all possible angles with respect to the environment of the molecule. In effect the tumbling averages out the angles of the bonds with respect to the environment. In contrast, a large molecule tumbles slowly and a molecule locked in place in a solid does not tumble at all. The absorption peaks from a quantity of such molecules are broadened by contributions from nuclei in molecules oriented at various angles with respect to their environment. Even a great concentration of phosphorus-31 nuclei in such a substance may therefore go undetected in NMR spectra because the absorption peaks are simply too broad.

Among the observable phosphorus-31 peaks the one due to inorganic phosphate has been the simplest to understand and among the most informative. Inorganic phosphate is found inside cells in combination with one or two protons. At an intracellular *p*H of 6.7 the populations of the singly protonated form and the doubly protonated form are equal. At higher *p*H's (that is, in a more alkaline environment) the singly protonated form dominates, and at lower *p*H's (a more acidic environment) the doubly protonated form dominates. Each form has a distinctive chemical shift, but the inorganic-phosphate peak in an NMR spectrum of cells will be found in some intermediate position depending on the average degree of protonation of all the P_i in the cells.

Thus the position of the P_i peak yields an accurate measure of the degree of protonation of the P_i, which in turn yields an accurate measure of the intracellular *p*H. Working with me at Bell Laboratories a few years ago, Gil Navon and Kamil Ugurbil showed by these means that the *p*H inside *E. coli* is kept quite constant near 7.4 when the bacteria have a source of energy such as glucose. In the same NMR spectra the conversion of glucose into FBP can be followed, and so can the subsequent increase in the concentration of ATP as the cell stores the energy it has acquired.

The British biochemist Peter Mitchell has proposed that the interconversion of ATP with ADP and inorganic phosphate in the cell is coupled to the passage of protons through the membrane of the cell. Specifically, the reaction that



PEAKS IN AN NMR SPECTRUM represent the absorption of radio waves by atomic nuclei at specific positions in molecules that have been placed in a magnetic field. The top part of the illustration shows the chemical reaction that joins inorganic phosphate (P_i) and adenosine diphosphate (ADP) into adenosine triphosphate (ATP), a molecule in which the living cell stores chemical energy. The bottom part of the illustration shows a phosphorus-31 NMR spectrum of the contents of mouse tumor cells. Almost all the absorption peaks in the spectrum are due to the nuclei of phosphorus-31 (P-31) atoms in the molecules shown at the top. The peak due to the gamma-P-31 in ATP and the peak due to the beta-P-31 in ADP are close together because these phosphorus atoms are in a similar chemical environment. The peak from the beta-P-31 in ATP is more isolated because its environment is unique: it alone is flanked by phosphate groups.

liberates chemical energy by splitting ATP into ADP and P_i serves to expel protons from the cell, thereby creating a proton gradient that stores both electrical and chemical energy. Conversely, the reentry of protons into the cell provides the energy required for the reaction that combines P_i and ADP into ATP. A change in the concentration of protons inside the cell constitutes a change in the intracellular *p*H; hence the chemical shift of the P_i peak in NMR spectra provides a test of Mitchell's hypothesis.

One begins the test by giving glucose to *E. coli* bacteria that have exhausted their food supply. The *p*H in such cells will have dropped to a value considerably less than 7.0. As the cells start to metabolize their fresh supply of glucose, the peak due to inorganic phosphate is observed to split into two peaks. One of them, corresponding to a *p*H of 7.4, arises from P_i inside the cells. The other, corresponding to a lower *p*H (which signifies a greater concentration of protons), arises from P_i in the medium in which the cells are suspended. The peaks confirm that a proton gradient is being established. At the same time the peaks due to ATP are rising.

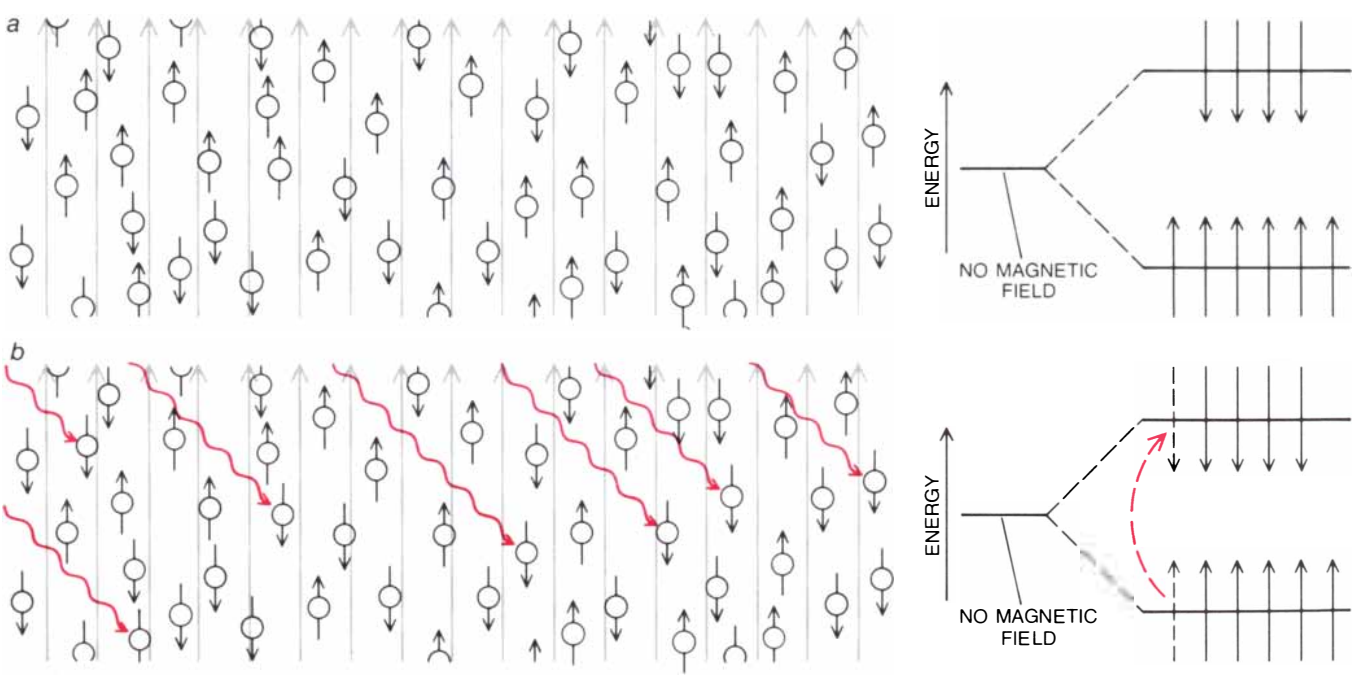
The reaction that splits ATP into ADP and inorganic phosphate is catalyzed by ATPase, an enzyme that couples the reaction to the transport of protons across the cell membrane. If a substance that inhibits the action of ATPase is added to a suspension of *E. coli* along with a fresh supply of glucose, one finds again that the glucose is consumed. Soon FBP and then ATP appear. Yet even when appreciable amounts of ATP have formed, observations of the P_i peak reveal that a proton gradient does not arise. Moreover, the glucose supply lasts much longer than it did before. This shows that an appreciable fraction of the consumption of ATP in the cell goes to maintaining the proton gradient.

The importance of NMR spectra in the study of metabolism is particularly well illustrated by a series of experiments done on muscle by George K. Radda, David G. Gadian and their colleagues at the University of Oxford. In 1974 these workers showed that NMR spectra could be employed to observe metabolism in living muscle tissue from the frog. Their spectra revealed phosphorus-31 peaks from inorganic phosphate, from ATP and from phosphocreatine (PCr), a molecule that muscle cells draw on to convert ADP into ATP. When the muscle was deprived of an external source of energy, the spectra showed a loss of PCr and a constant level of ATP, maintained by the reservoir of PCr. Only when the PCr was exhausted did the level of ATP fall, leaving ADP and P_i. The positions of the ATP peaks revealed that essentially all the ATP in the cells was complexed with positively charged metal ions (thought to be magnesium).

These preliminary studies (supplemented by similar experiments done by Michael Barany, C. Tyler Burt and Thomas Glonek of the University of Illinois Medical Center) succeeded, then, in making measurements that earlier investigators had made by the chemical analysis of numerous muscle samples taken at various times after the tissue had been nourished. Subsequent studies went further. Since the function of muscle is to do work, Gadian, with Douglas Wilkie and Joan Dawson of the University College London School of Medicine, undertook to make the phosphorus-31 peaks in NMR spectra reveal correlations between metabolites and the fatigue of the muscle: the decline of muscle force after prolonged exertion. They suspended a frog muscle in a vessel much like a test tube, which they placed in the NMR spectrometer.

In some experiments the solution surrounding the muscle was oxygenated; in others it was deoxygenated. An electrode stimulated the muscle, causing it to contract. A strain gauge measured the force the muscle developed. Gadian and his colleagues concluded that a decline in the force is well correlated with an increase in the concentration of ADP. Furthermore, it is proportional to the rate at which ATP is consumed. On the other hand, it does not depend on the concentration of ATP. The experimenters could follow the rate of consumption of ATP by measuring the decreasing level of phosphocreatine, the source of ATP in muscle, and simultaneously measuring the level of inorganic phosphate, which arises from the consumption of ATP.

In the past two years advances in NMR spectroscopy and the building of new spectrometers have allowed the



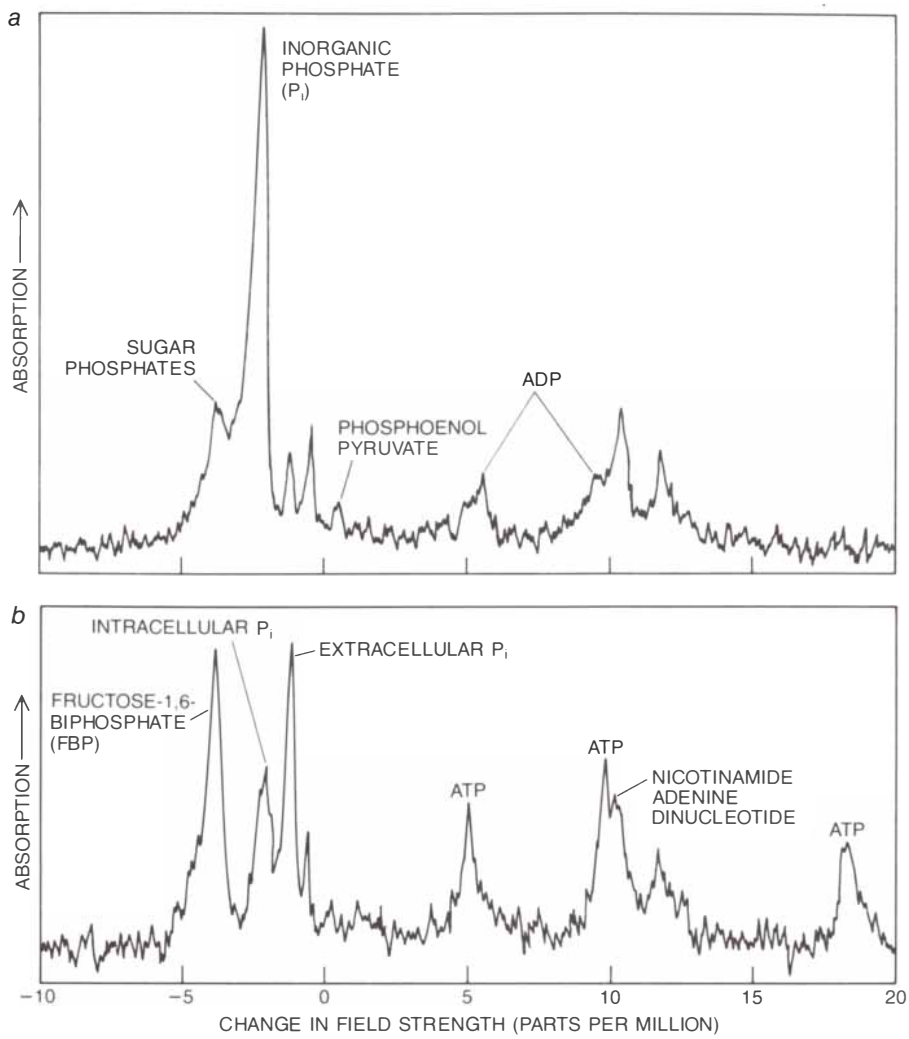
INTRINSIC MAGNETIZATION of certain nuclei (including the nuclei of hydrogen 1, carbon 13 and phosphorus 31) makes NMR spectroscopy possible because each such nucleus is the equivalent of a bar magnet. In a strong magnetic field the nuclei become aligned either with the field or against it (*a*). If the nuclei are then bathed in radio waves, they will absorb the quanta of radiation that give them the energy to "flip" from alignment to antialignment (*b*). The types of nuclei and the chemical environments they are in determine the combinations of radio frequency and magnetic field strength at which absorptions appear as peaks (or "lines") in the NMR spectrum.

study of metabolism in muscle to be extended to the muscles in the limbs of living people. The earlier spectrometers, built for chemical studies, had been designed to accommodate samples in a cylindrical tube. Inside the spectrometer the tube with its contents was surrounded by a coil that emitted radio-frequency radiation, and the tube and the coil were immersed in a static, homogeneous magnetic field. Clearly the examination of intact human limbs would require a different apparatus.

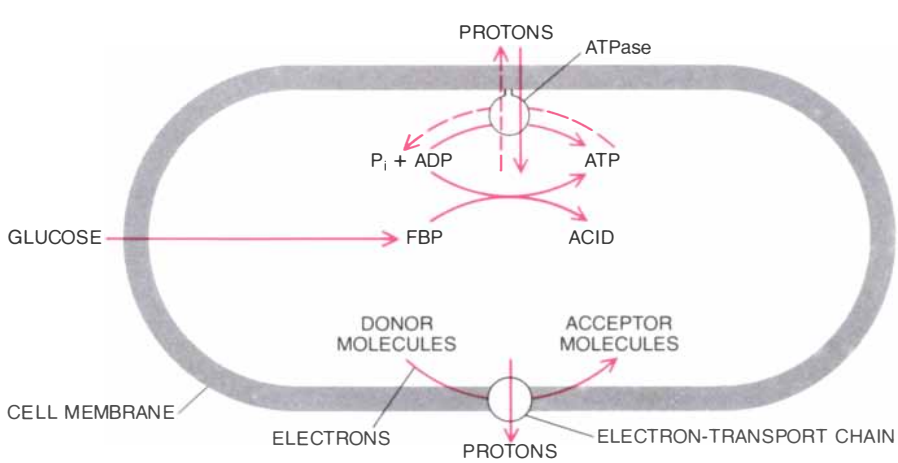
In the new spectrometers a flat coil is laid on the surface of the sample. The flat coil does not excite the entire sample with radio waves; it irradiates only the immediate vicinity of the coil. This turns out to be an advantage: the experimenter can follow chemical reactions in a specific region of tissue. To circumscribe the region still further the magnetic field in the new spectrometers can be adjusted so that it is homogeneous only in a volume of one or two cubic centimeters. Sharp NMR peaks are detected from this volume. In the rest of the volume under the influence of the magnet the magnetic field is quite inhomogeneous, so that if radio waves are absorbed there, the absorption is spread over a wide range of frequencies or magnetic field strengths and merges into the background of the spectrum.

By combining a flat surface coil with a focused magnetic field, Radda and Gadian, in collaboration with workers from Oxford Research Systems, the company that builds the new spectrometers, have been able to measure phosphorus-31 absorption peaks from the liver of living rats to the exclusion of peaks from the surrounding muscle tissue. They knew their spectra represented the liver alone because the spectra showed no peaks from phosphocreatine, which is absent in the liver but is present in muscle in high concentration. More recently Radda, working with a group of clinical investigators led by Brian D. Ross of Oxford, has reported the application of NMR spectroscopy to the diagnosis of human disease. Specifically, the investigators diagnosed McArdle's syndrome in a 51-year-old patient. McArdle's syndrome is a genetic disorder in which the enzyme phosphorylase cannot function. Phosphorylase participates in converting the carbohydrate glycogen into glucose on the demand of muscle cells. Someone who suffers from the syndrome becomes tired after the slightest exertion.

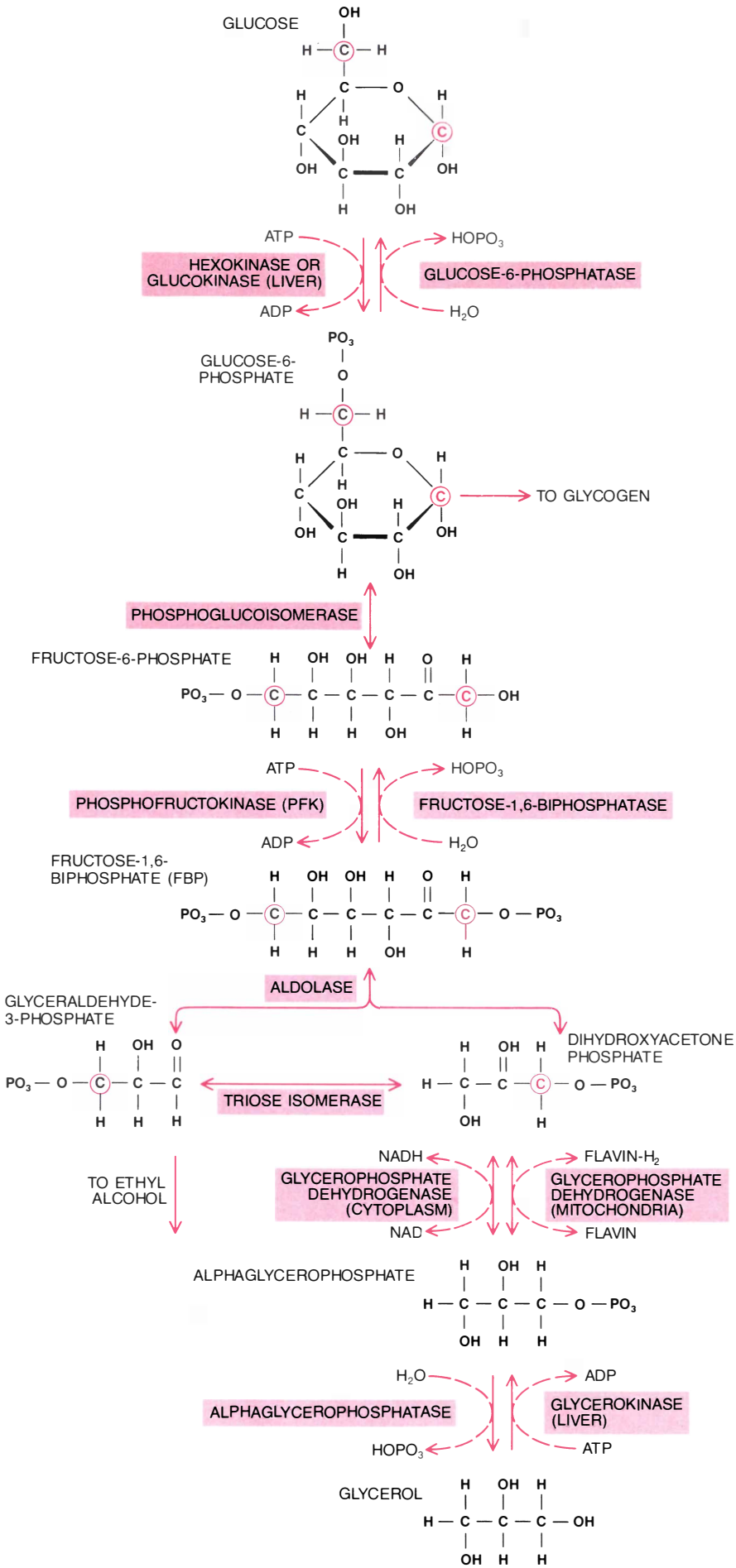
To make the diagnosis the patient was asked to put his arm in the NMR spectrometer, and phosphorus-31 spectra of his forearm were measured as he exercised. The investigators had already shown that in the muscles of normal people the *p*H drops during exercise. In one instance it dropped from 7.0 to 6.4,



EFFECTS OF GLUCOSE CONSUMPTION on the compounds containing phosphorus in the bacterium *Escherichia coli* are traced in these phosphorus-31 NMR spectra, which were made before the cells were given glucose (*a*) and then from four to six minutes after they were given glucose (*b*). In the second spectrum the inorganic-phosphate peak is found to have split. From the positions of the two resulting peaks it can be deduced that the interior of the cells has become more alkaline than the medium in which the cells are suspended. The remaining peaks are due to intracellular substances such as ADP, ATP and fructose-1,6-biphosphate (FBP).



INTRACELLULAR MECHANISMS thought to underlie the changes in the NMR spectra shown in the illustration at the top of the page are diagrammed. The metabolism of glucose gives rise to FBP and ultimately to substances such as lactic acid. In the later steps of the process ADP is converted into ATP. The energy in ATP can be tapped by the membrane-bound enzyme ATPase, which employs the energy to expel protons (hydrogen ions) from the cell. As a result a proton gradient arises and the cell becomes alkaline with respect to its environment. In the reverse reactions, also catalyzed by ATPase, protons are readmitted and the regained energy is utilized to make ATP. The mechanism shown at the bottom expels protons independently.



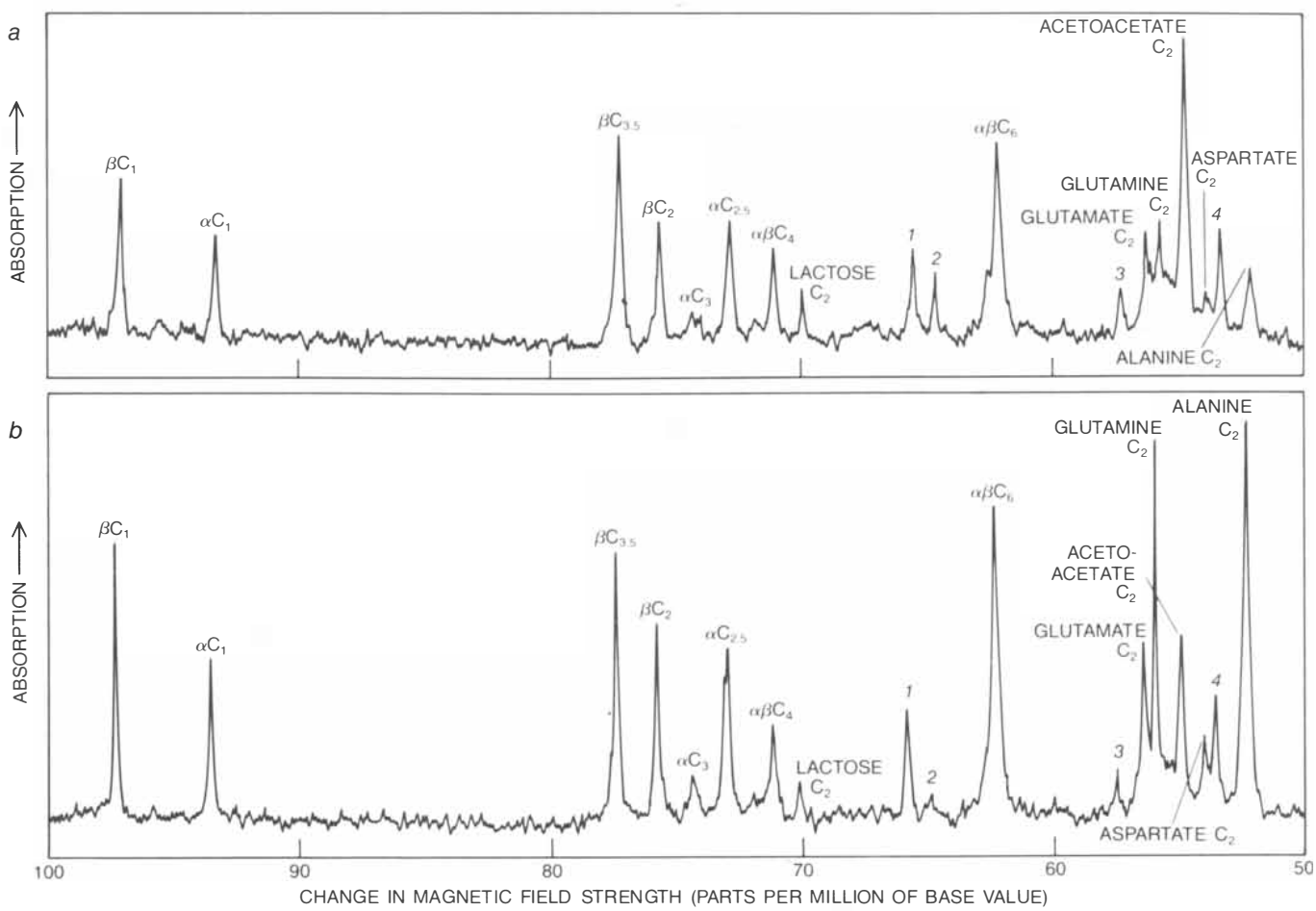
then returned to about 7.0 when the exercising stopped. In contrast, the *p*H of the patient's muscle began at 7.1 and increased to 7.2. It was also observed that the peak in the patient's spectra due to phosphocreatine decreased markedly as the patient exercised.

The explanation of these findings is that in the absence of the action of phosphorylase the patient's muscle cells had little glucose available to be metabolized by glycolysis. The glycolysis would have made lactic acid, whose temporary accumulation lowers the *p*H of normal muscle cells. Glycolysis would also have increased the supply of ATP, which would then have been available to provide the energy needed for muscular exertion. In the patient's muscle cells the required ATP was being made from phosphocreatine. The diagnosis that phosphorylase was inactive in the patient was later confirmed by a chemical test of a muscle biopsy specimen. Ross and his colleagues note that the noninvasive, harmless assays NMR makes possible could be helpful not only in diagnosing diseases such as McArdle's syndrome but also in evaluating the success of treatment.

The details of the glycolytic reactions by which glucose is metabolized can be examined inside the living cell by means of NMR. For example, in the process of glycolysis each glucose molecule, which has six carbon atoms, is converted into two lactic acid molecules, which have three carbon atoms each. As I have noted, the corresponding acidification shows up in phosphorus-31 NMR spectra of healthy muscle tissue as a shift of the inorganic-phosphate peak in accord with the lowered *p*H.

The chemical transformation of glucose can be followed more directly, however, by NMR spectroscopy of cells that are fed glucose in which most of the molecules are labeled with carbon 13 at the carbon position designated C_1. This strategy was pursued in my laboratory at Yale University by J. A. den Hol-

GLYCOLYTIC PATHWAY is the sequence of chemical reactions by which cells metabolize glucose. The sequence can be followed in NMR spectra by feeding cells glucose labeled with carbon 13. Here glucose has been labeled at the carbon position designated C_1 (*colored ball*). Then in the course of glycolysis the label has moved to the C_1 position of FBP and from there into dihydroxyacetone phosphate, a three-carbon compound. If the reverse reactions are occurring, a substantial part of the label will move to glyceraldehyde-3-phosphate (*black and colored ball*) and from there to the C_6 position in FBP. Ultimately it moves to the C_6 position in glucose. The "scrambling" of the label in FBP, as measured by the intensity of the carbon-13 peaks in NMR spectra, therefore constitutes a measure of the reversal of glycolysis in the living cell.



EFFECT OF THYROID HORMONE on cells in the liver of the rat is to increase the rate at which they synthesize glucose by a process that is essentially the reverse of glycolysis. One NMR spectrum (*a*) shows peaks due to carbon 13 in substances that rat liver cells have synthesized after the rat was fed alanine (an amino acid) labeled with carbon 13 at the C_3 position. For the other spectrum (*b*) a rat was treated with thyroid hormone and then fed labeled alanine. The peaks due to carbon 13 at various positions in two forms of glucose (α *and* β) are notably higher. Other peaks also show differences. The peaks marked *1, 2, 3* and *4* in each spectrum have not yet been identified.

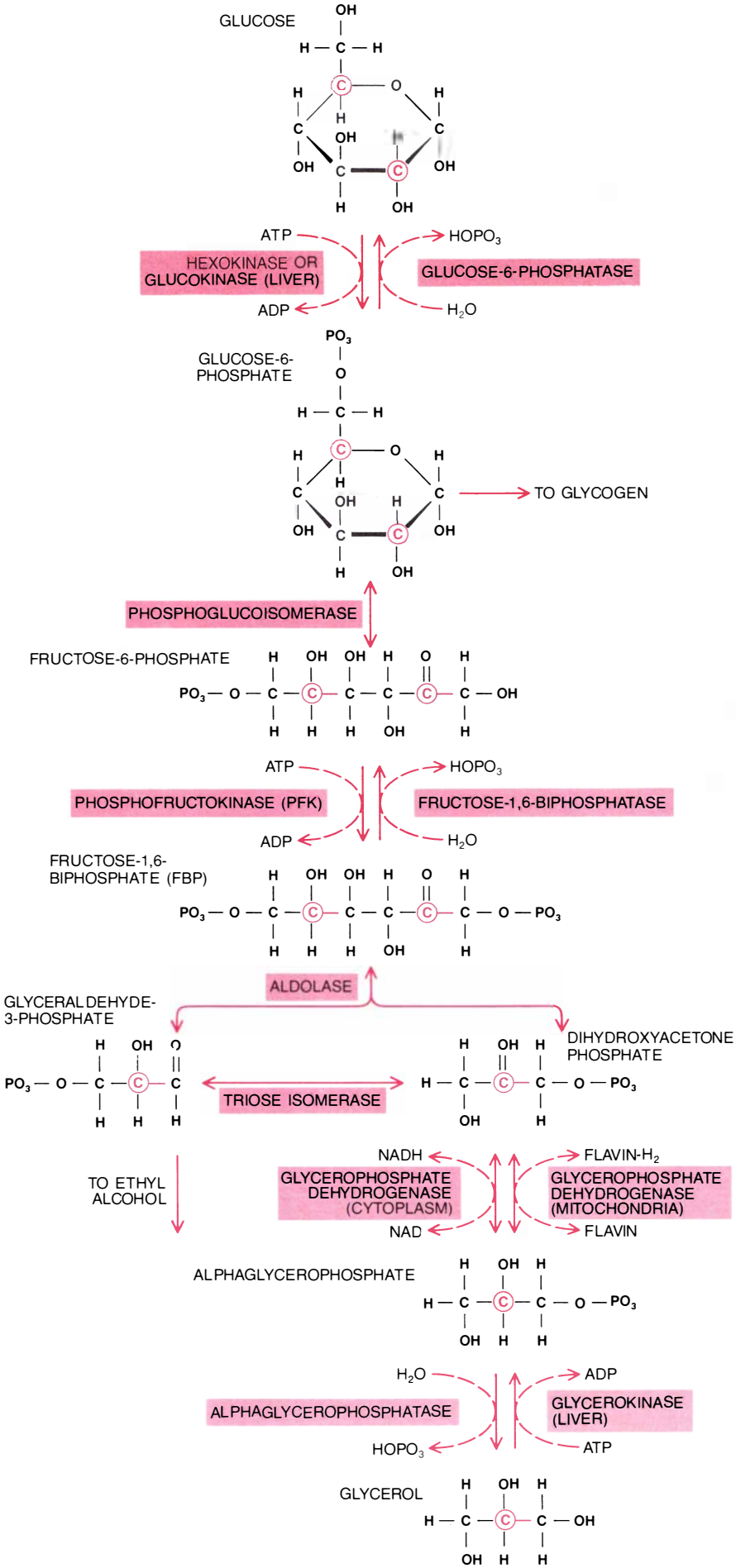
lander, following earlier work done in collaboration with Ugurbil and T. R. Brown. The aim was to measure the rates at which intermediate substances along the glycolytic pathway are transformed from one into the next inside yeast cells, and then to correlate these rates with the properties of the enzymes that catalyze each transformation.

For some time it has been known that the rate at which yeast cells consume glucose is higher in the absence of oxygen than it is if oxygen is present. The increase is called the Pasteur effect. Den Hollander, Ugurbil and Brown began, therefore, by measuring the peaks due to glucose in carbon-13 spectra of a suspension of yeast cells to which they had fed the labeled molecules. The results showed quite clearly that glucose was absorbed and consumed twice as fast in the absence of oxygen.

The next question was whether the Pasteur effect is caused by changes in the rates at which certain enzymes act. One possibility in particular seemed worth examining. Several steps along the glycolytic pathway there is a reaction that transforms fructose-6-phosphate into fructose-1,6-biphosphate, or FBP, under the governance of the enzyme phosphofructokinase (PFK). The enzyme is known to be allosteric: the rate at which it catalyzes the reaction depends not only on the concentration of its substrate (the fructose-6-phosphate on which it acts) but also, at least in the test tube, on the concentration of about 10 different small molecules, which are called for that reason effectors. Partly because of this wealth of possible controlling mechanisms it had long been presumed that PFK is controlled in living cells. No experiment had shown, however, which effectors in a cell really control the enzyme.

Den Hollander, Ugurbil and Brown employed phosphorus-31 spectra to investigate the precise changes caused in yeast cells by the presence of oxygen. Their measurements showed a significant change in the P_i peak. It decreased about fourfold in intensity, and it shifted in position in a way that showed the intracellular *p*H had increased from 7.04 to 7.45. At the same time the peak due to fructose-6-phosphate increased severalfold, and the peaks due to ATP and ADP changed slightly. Test-tube measurements of the activity of PFK made recently by den Hollander and David Reibstein show that under these conditions the enzyme becomes less active.

Remarkably, however, the intracellular concentration of FBP (the product of the reaction catalyzed by PFK) stayed fairly constant in den Hollander's spectra, even though the rate at which FBP formed had markedly decreased. The rate at which FBP was being consumed must also have decreased, and by the same amount. The flux of FBP (the flow of molecular matter through that stage of glycolysis) can be inferred from carbon-13 NMR spectra by a method that capitalizes on the nature of the glycolytic reactions. In the first steps of glycolysis the carbon-13 label den Hollander introduced at the C_1 position of glucose makes its way directly to the C_1 position of FBP. Then comes a reaction in which FBP splits into two three-carbon compounds. Each such compound can undergo reactions that take it further down the pathway. Alternatively, each three-carbon compound can be transformed into the other three-carbon compound and the two can then combine to remake the



FBP. These reverse reactions "scramble" the carbon 13 so that in the remade FBP it appears at the C_6 position.

Overall, therefore, the competition between the downward flow along the glycolytic pathway and the reverse direction of flow, in which FBP reappears, determines the distribution of labeled atoms between the C_1 and the C_6 positions of FBP. If there is little flow down the pathway, there will be ample time for FBP to be remade. In that event the intensity of the NMR peak from carbon 13 at the C_6 position of FBP should equal the intensity of the peak from the C_1 position. If, on the other hand, the downward flow is rapid, there will be little time for the label to be scrambled and the C_6 peak will be nil. Den Hollander's spectra show that the oxygenation of yeast cells increases the scrambling. He was able to calculate that the downward flux through FBP had been reduced severalfold. Hence the flux was being controlled at some point further along the pathway.

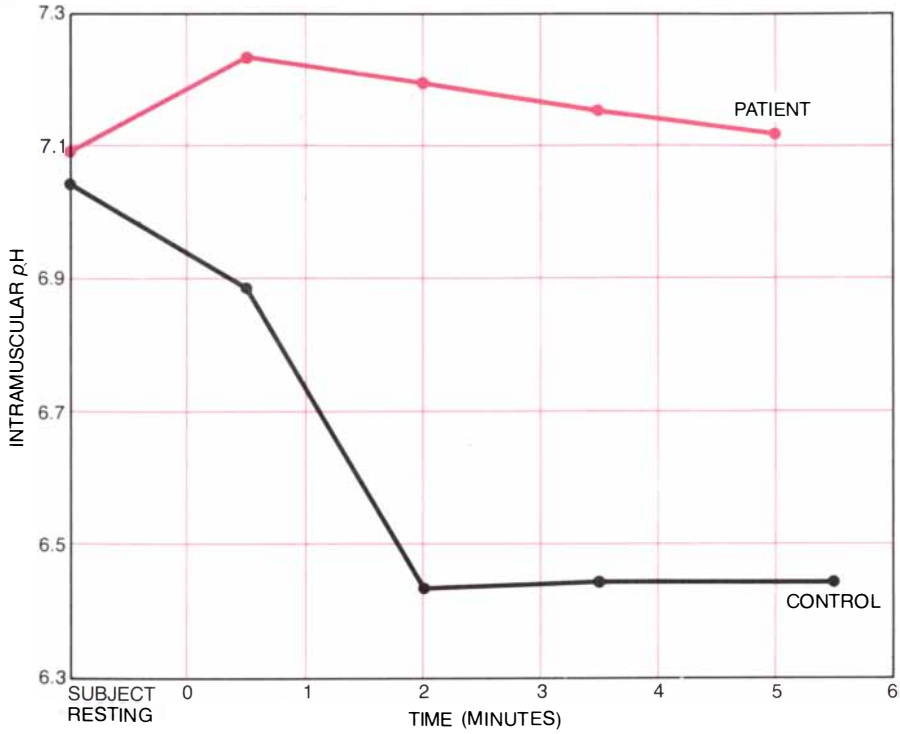
The reversal of flow along the glycolytic pathway is well worth studying because many cells reverse the flow to synthesize glucose if they lack it. FBP is an intermediate in both directions of flow. The enzyme that makes FBP along the pathway leading away from glucose is PFK; the enzyme that catalyzes the opposite reaction, which brings FBP one step closer to glucose, is fructose-1,6-biphosphatase. It is notable that the making of FBP by PFK requires inorganic phosphate to be transferred from ATP to fructose-6-phosphate. In the reverse reaction, however, the phosphate is simply cast off. If both flows occur, the net result is the waste of an ATP molecule in what is called a futile cycle. Many such cycles are known.

What controls a futile cycle? One hypothesis about the cycle involving FBP was stimulated by the knowledge that FBP, like many other sugars, exists in two forms (designated alpha and beta) that differ only in the interchange of two carbon bonds. A mixture of the forms takes only about one second to reach an equilibrium distribution in which 15 percent of the molecules are of the alpha

LABELING OF GLYCEROL (an end product of glucose metabolism) makes it possible to follow the gluconeogenic pathway: the series of reactions that reverses glycolysis. When the middle carbon position of glycerol is labeled with carbon 13 (*color*), the label ends up distributed between the C_2 and the C_5 positions of glucose. A small amount of the label (not shown) ends up at the C_1 position of glucose because it passes through an alternate series of reactions called the pentose-phosphate pathway. NMR studies of gluconeogenesis show that thyroid hormone increases the production of glucose in hyperthyroid rats by increasing the activity of the mitochondrial enzyme glycerophosphate dehydrogenase.

form and 85 percent are the beta. Test-tube studies show that fructose-1,6-biphosphatase, which takes FBP toward glucose, acts preferentially on the alpha form. Aldolase, the enzyme that takes FBP a step further from glucose, acts preferentially on the beta form.

It had accordingly been suggested that the living cell controls the direction of flow through the futile cycle by controlling the distribution of the forms of FBP. The suggestion could not be tested by extracting FBP from the cell and analyzing the distribution. The extraction would take longer than the one second needed for equilibrium to become established. The NMR spectroscopy of intact *E. coli* bacteria makes a different test quite feasible. Den Hollander, Ugurbil and Brown measured carbon-13 peaks from the alpha and beta forms of FBP inside the cells while the cells were metabolizing glucose. The results showed that the FBP was in equilibrium. The direction of flow was not, therefore, being controlled by disequilibrium.



DIAGNOSIS OF DISEASE by means of NMR spectroscopy has been reported by a group of workers led by George K. Radda and Brian D. Ross of the University of Oxford. The disease, McArdle's syndrome, is characterized by fatigue following slight exertion and results from the incapacity of the enzyme phosphorylase, which takes part in making glucose in muscle tissue. The metabolites of glucose decrease the *p*H of cells; in NMR spectra, therefore, the disease is revealed by a lack of change in *p*H, as is shown by the unchanging position of the inorganic-phosphate peak in spectra of the patient's arm made as the patient exercised. The subjects of the measurement wore a cuff to keep fresh glucose from arriving in arterial blood.

When glucose is consumed in the cells of a mammal, the end products of the glycolysis (such as lactic acid) are carried by the blood to the liver along with the end products of the metabolism of fat (such as glycerol). The liver takes these wastes and converts them back into glucose. The liver has thus been made the subject of extensive investigation into the reversal of glycolysis: the gluconeogenic pathway. Paradoxically, the results are too abundant: the study of isolated livers, liver slices and liver cells reveals a number of possible ways the pathway might be controlled but no insight into how it really is controlled in the liver of a living animal. I shall summarize the possible types of control to give a background for discussing the progress being made by NMR spectroscopy in disentangling the actual controls from the merely plausible ones.

First, the rate of production of glucose in the liver can be influenced by the concentrations of the substrates of the gluconeogenic path. An increased level of lactic acid gives rise to a higher rate of glucose production. Second, the rate of production of glucose can be influenced by the supply of food to the organism. Animals deprived of glucose have increased levels of the enzymes that synthesize glucose. Third, the activity of the enzymes that synthesize glucose can be influenced by the extent to which phosphate groups are attached to the enzymes. This phosphorylation, which usually is controlled by rapidly acting hormones such as glucagon, is known to activate a number of enzymes in cells. Fourth, several enzymes that synthesize glucose are controlled by effectors. Fifth, the concentration of various enzymes can be influenced by slow-acting hormones such as the ones synthesized by the thyroid gland.

Sheila Cohen, who was then at Bell Laboratories, employed the NMR spectroscopy of compounds labeled with carbon 13 to follow the gluconeogenic pathway in the intact liver of the rat. One of her studies followed the synthesis of glucose from glycerol, a pathway consisting of six reactions. Several differences emerged between normal rats and rats that had been rendered the equivalent of hyperthyroid by treatment with high doses of a thyroid hormone. In the hyperthyroid rats the rate of production of glucose showed a twofold increase and the concentration of alpha-glycerol phosphate, the compound immediately after glycerol in the pathway, showed a two-and-a-half-fold decrease.

Two enzymes compete for alpha-glycerol phosphate as a substrate. Both are called glycerophosphate dehydrogenase, but one of them is found in the intracellular organelles called mitochondria and the other is in the cytoplasm of the cell. Cohen calculated that under the influence of the hormone the mitochondrial enzyme got a fourfold increase in substrate at the expense of the cytoplasmic enzyme. The finding is in agreement with the levels and activities of the enzymes measured in extracts from liver cells.

There were also some unexpected results. For example, Cohen's NMR spectra showed that the thyroid hormone somehow gave rise to a severalfold increase in the flux through a futile cycle in the gluconeogenic pathway. The basis of such changes is probably the altered concentrations of enzymes brought about over time by the hormone.

A recent experiment done by investigators at Oxford Research Systems and Yale at the instigation of Jeffrey R. Alger in my laboratory suggests the future of NMR spectroscopy. The experimenters fed glucose labeled with carbon 13 to a rat. Then in NMR spectra they watched the glucose disappear from the living animal's stomach. Over a period of hours the level of glycogen in the animal's liver increased; the glucose was being stored as glycogen there. Finally the glycogen peak disappeared as the labeled glycogen was consumed. The NMR spectra had enabled the investigators to follow metabolism throughout the animal's body. Clearly NMR promises the ability to watch precisely how hormones, diet and illness affect the chemical processes of life. Considering the short time in which NMR spectra have served the study of metabolism it is reasonable to expect that within a few years the technique will revolutionize our understanding of how biochemistry proceeds in living animals.

The Physics of Organ Pipes

The majestic sound of a pipe organ is created by the carefully phased interaction of a jet of air blowing across the mouth of each pipe and the column of air resonating inside the pipe

by Neville H. Fletcher and Suszanne Thwaites

No other musical instrument can compare with the pipe organ in power, timbre, dynamic range, complexity of tone and sheer majesty of sound. Like many other musical instruments the organ was brought to a high state of perfection by generations of craftsmen drawing on a slow accumulation of empirical knowledge. By the end of the 17th century the organ had attained essentially its modern form. Two of the 19th century's most illustrious physicists, Hermann von Helmholtz and Lord Rayleigh, reached opposite conclusions about the fundamental mechanism of how organ pipes generate sound but lacked the technical means for resolving the issue. With the advent of oscilloscopes and other modern devices it has become possible to develop a detailed understanding of the mechanism. It turns out that the analyses of Helmholtz and Rayleigh are each valid over certain ranges of the pressure at which air is blown into the organ pipe. The present understanding, which will be described here, is at variance with many of the explanations of organ-pipe mechanisms still found in textbooks.

Pipes cut out of reeds or other hollow-stemmed plants were probably the first musical instruments. They could be made to emit sound by blowing across the end of the pipe, by blowing into the end in such a way that the lips vibrated or by pinching the end in such a way that blowing through it caused the sides of the pipe itself to vibrate. Modern versions of these three primitive wind instruments are flutes, trumpets and clarinets, all of which have been developed so that the performer can play many notes spread over a substantial range of acoustic frequencies.

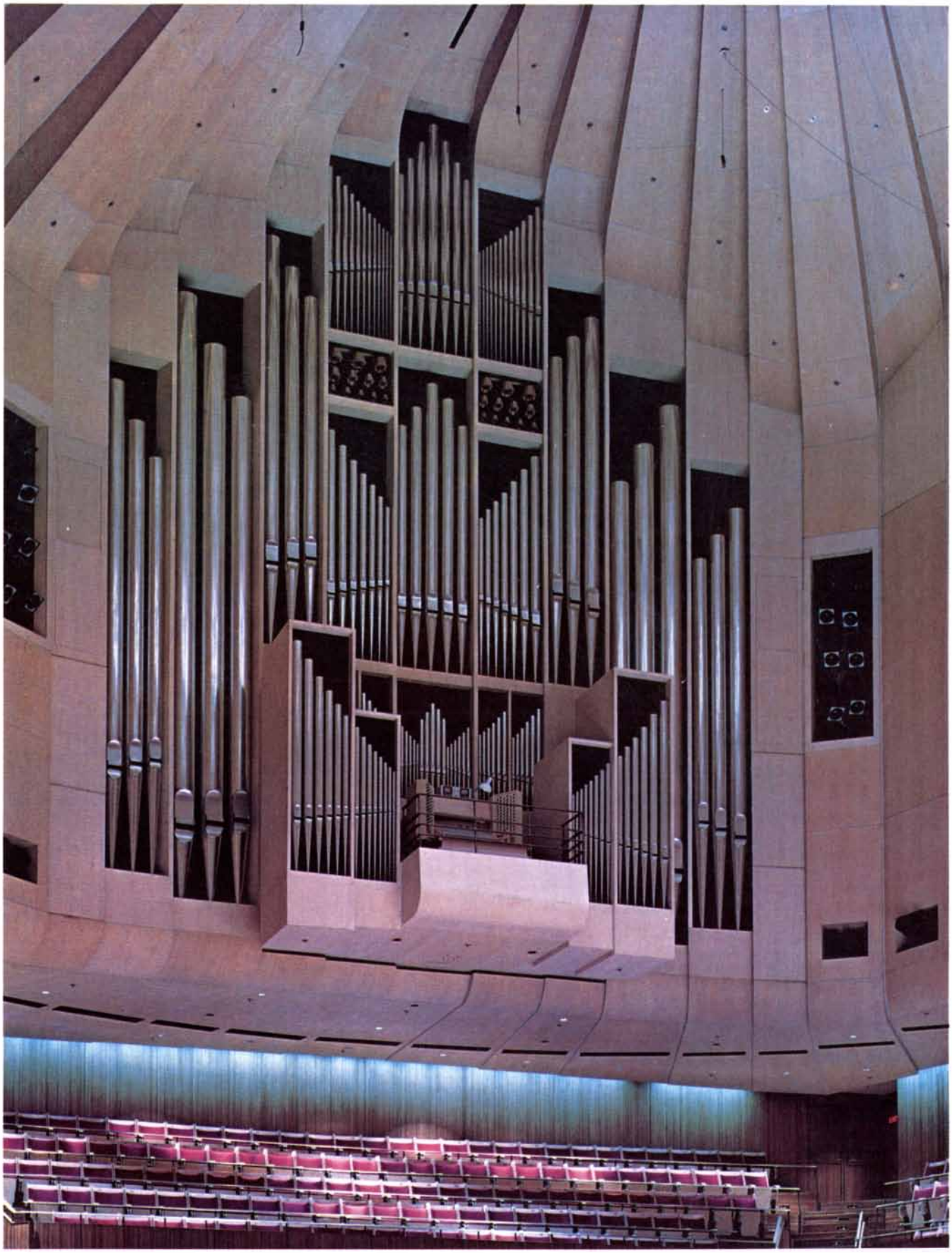
There has been a parallel development in pipe instruments in which a different pipe is dedicated to each note. The simplest such instrument is the panpipe (from "pipes of Pan"), which commonly has about 20 pipes of different lengths, all closed at one end and excited by being blown across the open end. At the other extreme in size and complexity is the pipe organ, an instrument with as many as 10,000 pipes that the player controls through complex mechanical linkages. The pipe organ has a long history. Pottery figures of Alexandrian musicians playing an arrangement of pipes excited by bellows date back to the second century B.C. By the 10th century the organ began to be adopted in Christian churches, and treatises on organ building, written by monks, were in circulation in Europe. The great organ built in the 10th century for Winchester cathedral in England was reputed to have had 400 metal pipes, 26 bellows and two keyboards with a total of 40 keys, each key controlling 10 pipes. In the succeeding centuries there were substantial mechanical and musical advances in organ building, and as early as 1429 an organ with 2,500 pipes was built in Amiens cathedral. By the late 17th century in Germany organs had evolved to their modern form.

The terminology for describing organ pipes reflects their origin in pipes blown by mouth. Organ pipes are open at the top and tapered at the bottom, with a "mouth" (a slot) running across a flattened section above the taper. Inside the pipe is a "languid," or tongue (a horizontal plate), with a "flue" (a narrow slit) between it and the lower "lip" of the mouth. The air that excites the pipe, supplied by a large bellows, reaches the tapered foot of the pipe at a typical pressure of 500 to 1,000 pascals (five to 10 centimeters measured on a water gauge). When the air is admitted to the pipe by the actuation of the appropriate drawstop knob and key, it flows upward and forms a sheetlike jet as it emerges from the flue slit. The jet flows across the mouth and strikes the upper lip, where it interacts both with the upper lip and with the column of air in the body of the pipe to maintain the steady oscillation that generates the pipe's "speech." The problem of how the pipe makes the abrupt transition from silence to steady speech is a difficult one and fascinating in its own right but will not be part of our story. We shall be concerned primarily with the processes that control the steady speech of organ pipes and endow them with their characteristic tone qualities.

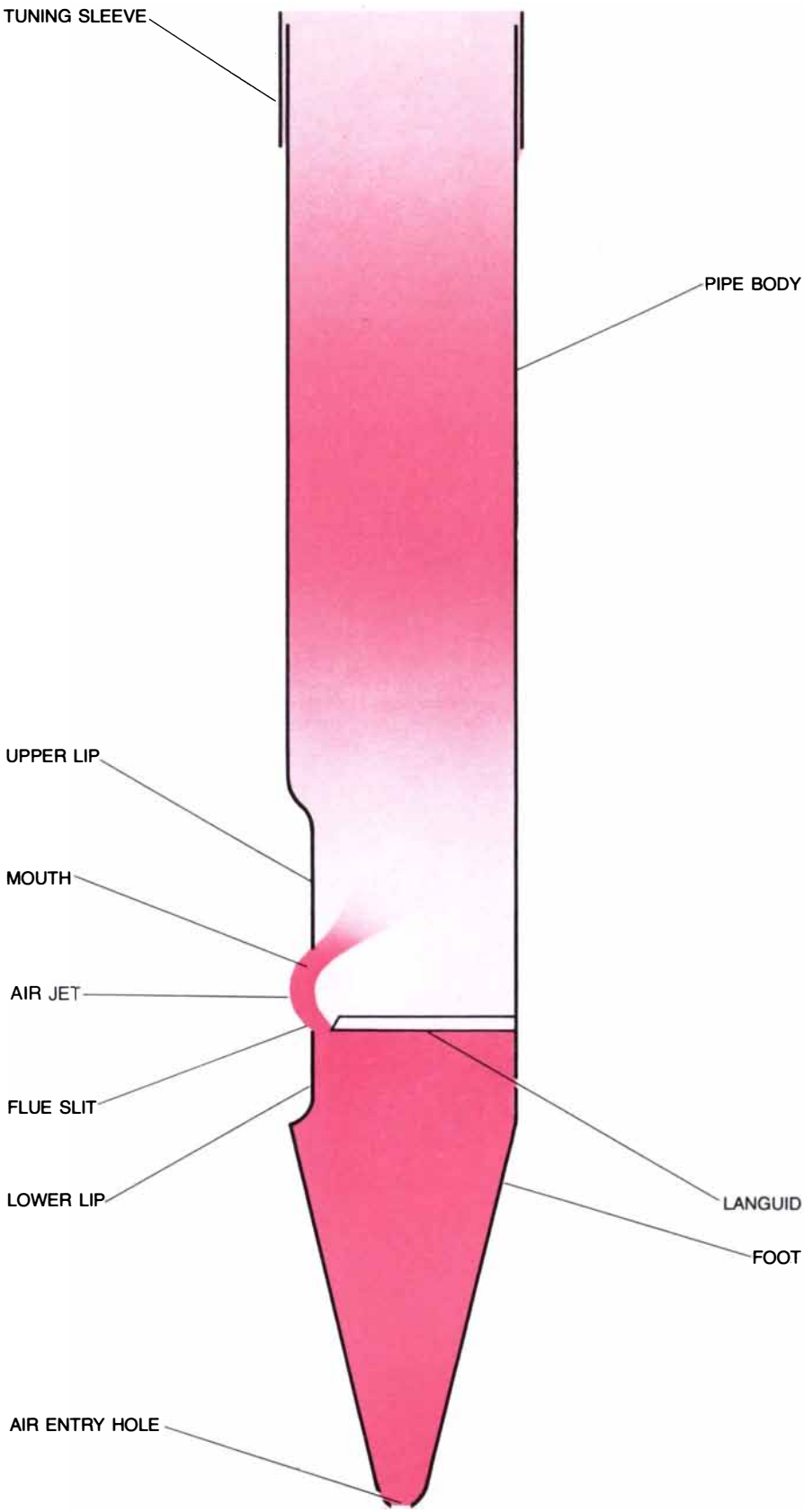
The behavior of the jet that initiates and sustains the organ's speech would seem to present a straightforward problem in fluid flow. It turns out, however, that even a steady jet flowing in a smooth laminar manner is quite difficult to understand theoretically, and the fully turbulent jet that flows in a real organ pipe is impossibly complex. Fortunately the complexity introduced by turbulence actually simplifies the behavior of the jet. If the flow were laminar, the jet would interact with its surroundings through viscosity. In the actual jet viscosity is replaced as the mechanism of interaction by turbulence, and on a scale directly related to the width of the jet. Organ builders go to some lengths to ensure that the jets in their pipes are fully turbulent by cutting fine nicks along the edge of the languid. Paradoxically a turbulent jet is stable and reproducible and a laminar jet is not.

A fully turbulent jet mixes gradually with the surrounding air, thereby broadening and slowing down in a simple manner. When the jet's velocity is plotted against the distance from its central plane, the resulting curve has a bell-like shape, with the maximum velocity in the center. The width of the jet increases linearly with the distance from the flue slit. The momentum of the flow must be conserved, which means that the velocity decreases with the square root of the distance from the slit. This description can be supported theoretically and agrees with experiment (when a small transition region near the flue slit is included).

In an organ pipe that is already excited and emitting sound the jet emerges from the flue slit into the intense sound field in the mouth of the pipe. The air motion associated with the sound is directed into and out of the mouth and therefore at right angles to the plane of the jet. Half a century ago Burniston



CONCERT HALL ORGAN at the Sydney Opera House in Australia, completed in 1979, is one of the largest and most advanced in the world. Designed and built by Ronald Sharp, it has some 10,500 pipes controlled by the mechanical action of five keyboards and a pedal board. The mechanical action, which regulates the flow of air into the pipes, is duplicated by an electric action that is under microprocessor control. The organ can therefore be operated by a magnetic tape on which an original performance has been recorded digitally.



ORGAN PIPE is excited by air blown in from the bottom. The air is formed into a jet by the flue slit between the lower lip and the languid: a plate running across the pipe. When the air blows across the mouth of the pipe, it interacts with the column of air in the pipe at the upper lip, blowing alternately into and out of the pipe. Waves propagating along the turbulent jet maintain a steady oscillation in the air column, causing the pipe to "speak." The standing-wave pattern of pressure in the pipe is represented by the colored shading. The upper end of the pipe is surrounded by an adjustable sleeve, as is shown here, or is closed by an adjustable stopper. Either makes it possible to change the length of the air column slightly for tuning.

Brown of University College London made beautiful photographs of smoke-laden laminar jets emerging into a sound field and found they developed sinuous waves that grew as they traveled along the jet until the jet broke up into a double row of vortex rings rotating in opposite directions. It was the somewhat naive application of these observations and similar ones that led to much of the confused discussion of organ-pipe physics still found in many textbooks.

A more fruitful way to study the behavior of real jets in a sound field is to remove the organ pipe and generate the sound field by means of a loudspeaker. Such studies, conducted by John W. Coltman of the Westinghouse Electric Corporation and by our own group at the University of New England in Australia, have provided much of the current understanding of the physics of organ pipes. Actually Rayleigh had developed a careful and nearly complete mathematical description of the behavior of laminar jets of nonviscous fluids. Because of the fortunate circumstance that turbulence has been found to simplify jet behavior rather than complicate it, Rayleigh's treatment can be carried over with little modification to describe the real jets produced and analyzed experimentally by Coltman and by us.

If one were to forget about the flue slit in the organ pipe, one would expect that the jet, being just a sheet of moving air, would simply be moved back and forth by the acoustic vibrations along with all the other air in the mouth of the pipe. As the jet leaves the flue, however, it is effectively held still by the flue slit itself. This has the same effect as superimposing on the general back-and-forth motion in the sound field an exactly balancing displacement localized at the flue. The localized displacement, which matches the sound field in frequency and amplitude in order to maintain zero displacement at the flue, is carried along with the moving air in the jet and imposes a sinuous wave motion on it.

As Rayleigh showed for a limiting case in his jets, and as we have verified in detail for diverging turbulent jets, the wave motion propagates along the jet at a speed that is a little less than half the speed of the air at the central plane of the jet. In addition, however, the wave grows in amplitude nearly exponentially as it propagates along the jet. In a typical case the amplitude of the wave doubles in traveling about one millimeter along the jet, so that the effects of the wave rapidly dominate the simple back-and-forth lateral motion imposed by the acoustic vibrations.

We find that the growth rate of the wave is greatest when its wavelength along the jet is about six times the width

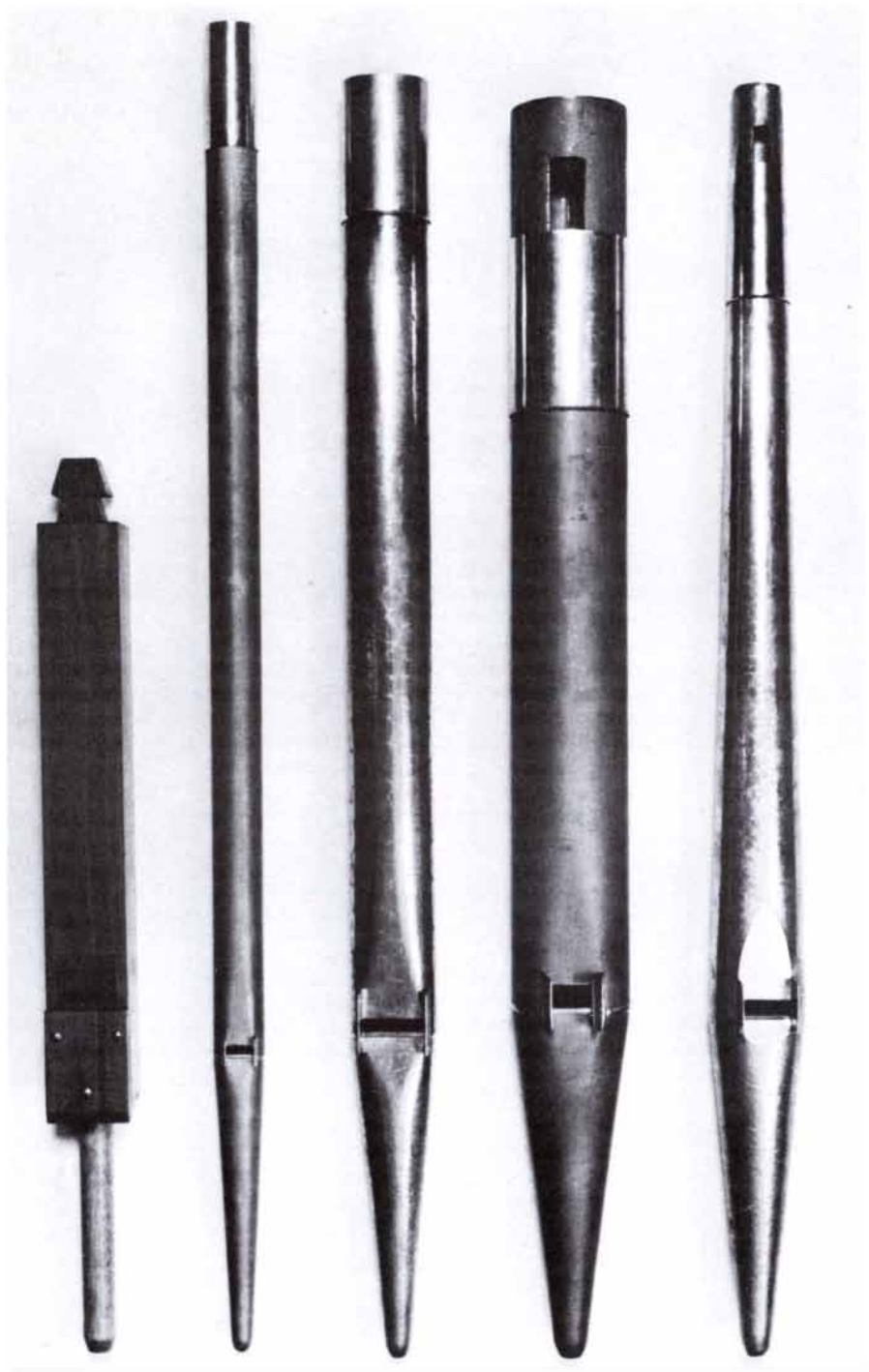
of the jet at that point. On the other hand, when the length of the wave is less than the width of the jet, the wave does not grow at all and may even die out. Since the jet slows and broadens as it moves away from the flue slit, only the long waves, that is, those of low frequency, can propagate with large amplitude along long jets. This fact will be an important one when we come to consider the harmonic content of organ-pipe sounds.

Let us now consider the effects of the organ pipe's sound field on the jet. As one can readily imagine, the large acoustic waves associated with the sound field in the pipe mouth will cause the tip of the jet to flick back and forth across the upper lip of the mouth so that the jet will flow alternately into the pipe and outside it. The situation resembles what happens when one pushes a swing that is already swinging. The air column in the pipe is already oscillating, and if the puffs of air enter the pipe in step with the oscillations, the puffs will maintain the oscillations against sundry energy losses, such as the radiation of sound from the pipe and the drag at the pipe walls. If, on the other hand, the puffs are out of step with the oscillations of the air column, the jet will tend to damp the oscillations and the sound will die away.

The relations between the acoustic motions of the air in the pipe mouth and the time the air pulse arrives inside the upper lip are determined simply by the time required for a wave on the jet to travel from the flue slit to the upper lip. The organ builder calls this distance the cut-up. If the cut-up is large or the blowing pressure is low so that the jet velocity is low, the travel time will be long. Conversely, if the cut-up is small or the blowing pressure is high, the travel time will be short.

In order to discover just what the correct phase relation between the oscillation of the air column in the pipe and the arrival of jet pulses inside the upper lip is, it is necessary to know more about how the pulses act on the air column. Helmholtz thought that the dominant factor in the relation was the volume of flow contributed by the jet. Then if the pulses in the jet flow were to impart as much energy as possible to the oscillation of the air column, they should enter the pipe at the times when the acoustic pressure inside the upper lip was at a maximum.

Rayleigh took a different view. He argued that since the mouth is actually not very far from the open end of the pipe, little acoustic pressure can build up just inside the mouth for the jet flow to work against. He considered that the jet flow was essentially stopped as soon as it entered the pipe, thus quickly building up pressure that could act on the flow in the pipe. Therefore according to Rayleigh the maximum amount of energy would be transferred from the jet if the jet flow entered the pipe when the acoustic flow, not pressure, was at a maximum. The difference between these two maximums amounts to a quarter of an oscillation period in the oscillation frequency of the air column in the pipe. In terms of the swing analogy the difference is between giving the swing a push when it is at the top of its arc and has acquired its maximum potential energy (Helmholtz) and giving the swing a push when



FIVE PIPES OF DIFFERENT DESIGN yield a note of the same pitch but different tone qualities. The second pipe from the left is the dulciana, which has a sweet, thin tone resembling that of a stringed instrument. The third pipe is the open diapason, which has the bright, ringing sound most characteristic of an organ. The fourth has the sound of a very dull flute. The fifth is the *waldflöte* (forest flute), which has a soft sound. The wood pipe at the left is closed with a stopper. It has the same fundamental frequency as the other pipes but only odd harmonics: overtones with frequencies that are odd-numbered multiples of the fundamental. The other pipes differ slightly in length because "end corrections" are needed to give them the same pitch.

it is at its lowest point and is moving fastest (Rayleigh).

The problem remained unresolved, and indeed virtually uninvestigated, for nearly 80 years. It has now been fairly satisfactorily decided as a result of work by Lothar Cremer and Hartmut Ising of the Heinrich Hertz Institute for Vibration Research in Berlin, by Samuel A. Elder of the United States Naval Academy, by Coltman and by us. Briefly, both Helmholtz and Rayleigh were partly right. The balance between the two driving mechanisms depends on the blowing pressure and the frequency of the sound, with the Helmholtz mechanism dominating for low blowing pressures and high frequencies and the Rayleigh mechanism dominating for high blowing pressures and low frequencies. For an ordinary organ pipe the Helmholtz mechanism is usually more important.

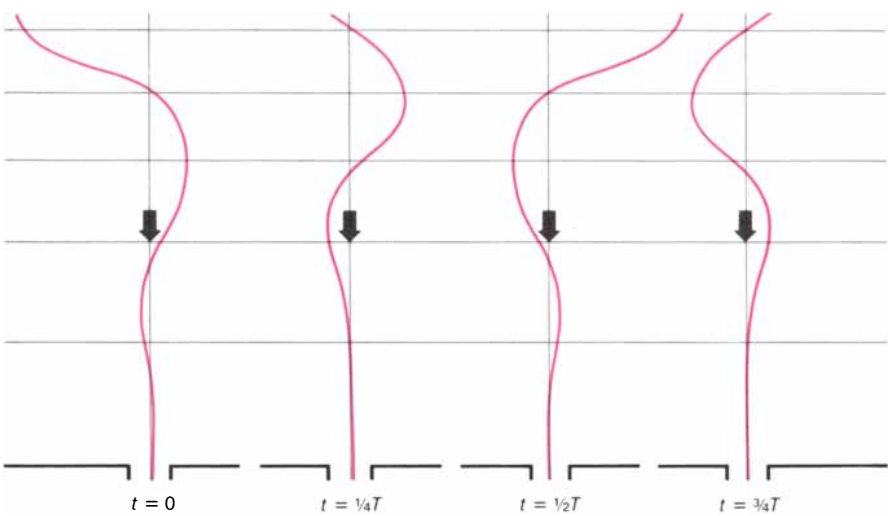
A simple and effective method for studying the properties of the jet was devised by Coltman and has been modified and extended in our laboratory. Essentially one examines the behavior of a jet at the mouth of an organ pipe that has been fitted with wedges of absorbent felt or foam at the far end so that the pipe is prevented from speaking. A sound wave is then directed down the pipe from a loudspeaker outside the far end. The wave is reflected from the mouth end, with and without the jet blowing. In both cases the incident and reflected sound waves interact inside the pipe to create a pattern of standing waves. By measuring with a small probe microphone the changes in the pattern when the jet is turned on, we can determine whether the jet adds energy to the reflected wave or subtracts it.

What our experiments actually determine is the acoustic "admittance" of the jet, which is defined as the ratio of the acoustic flow out of the mouth, produced by the presence of the jet, to the acoustic pressure just inside the mouth. The acoustic admittance has both a magnitude and a phase angle, which can be plotted as a function of frequency or of blowing pressure. When either the frequency or the blowing pressure is varied separately and the admittance is appropriately plotted, it takes the form of a spiral [*see top illustration on opposite page*]. The distance from the origin of the spiral represents the magnitude of the admittance and the angular position along the spiral represents the delay in the phase of the sinuous wave that is imposed on the jet by the acoustic oscillations in the pipe. A delay of one wavelength corresponds to 360 degrees around the spiral. Because of the special properties of turbulent jets it turns out that if the admittance is multiplied by the square root of the blowing pressure, all the measurements for a particular organ pipe fall in the same spiral.

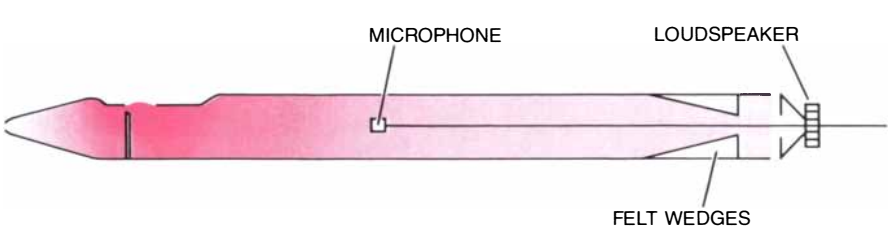
If we keep the blowing pressure constant and increase the frequency of the input sound wave, the point representing the admittance spirals inward toward the center in a clockwise direction. If we hold the frequency constant and increase the blowing pressure, the point spirals outward in the opposite direction.

When the point representing the admittance lies in the right half of the spiral, the jet is taking energy from the pipe flow and energy is therefore being dissipated. When the point lies in the left half, the jet is imparting energy to the pipe flow, which means that the jet is acting as an acoustic generator. When the point is in the upper half of the spiral, the jet lowers the natural resonance frequency of the pipe; when the point is in the lower half, the jet raises the natural resonance frequency. The reference angle from which the phase delay is measured depends on whether the Helmholtz or the Rayleigh excitation mechanism is dominant, and this, as we have remarked, depends on the blowing pressure and the frequency. It is never far, however, from the zero, or three-o'clock, position on the right arm of the horizontal axis.

Since 360 degrees around the spiral represents a phase delay of one wavelength in the sinuous wave propagated along the jet, it follows that for phase delays from rather less than a quarter of a wavelength to nearly three-quarters of a wavelength a representative point on the spiral will lie to the left of the center line, or in the region where the jet acts as an acoustic generator. We have



SHAPE OF THE AIR JET is shown in consecutive "snapshots" as it emerges from the flue slit into the acoustic velocity field generated in the mouth of the pipe by the resonating air column inside the pipe. The periodic displacement of the air in the mouth creates a sinuous wave that propagates with a speed equal to half of the speed of the air at the central plane of the jet and grows exponentially until its amplitude exceeds the width of the jet. The horizontal divisions show distances traveled by the jet wave in successive quarter periods of the oscillation period *T*. The lines get closer together as the jet loses speed. In an organ pipe the upper lip is positioned at about the location of the arrow, so that the air jet blows alternately into and out of the pipe.



ACOUSTIC GENERATING PROPERTIES OF THE JET can be measured by inserting felt or foam wedges in the open end of the pipe to keep it from sounding and introducing a sound wave of small amplitude from a loudspeaker. The sound wave is reflected from the opposite end of the pipe and interacts at the mouth with the air jet. The interaction of the jet with the standing-wave pattern in the pipe is examined with a movable microphone. In this way one can establish whether the jet is adding energy to or subtracting energy from the wave reflected at the base of the pipe. In order for the pipe to sound, the jet must add energy. The measurements are in terms of the acoustic "admittance": the ratio of the acoustic flow out of the mouth to the acoustic pressure just inside the mouth. A plot of the admittance for various combinations of blowing pressure and frequency yields the spiral curve at the top of the opposite page.

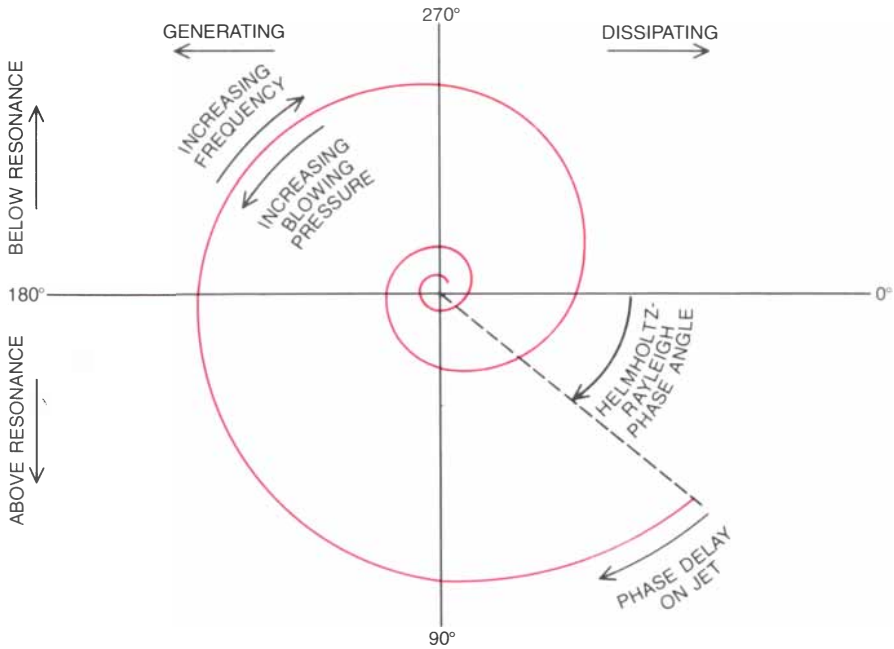
also seen that at a constant frequency the phase delay is a function of blowing pressure, which controls both the speed of the jet and the speed with which the sinuous wave propagates on the jet. Since the wave speed is half the jet speed and the jet speed is proportional to the square root of the blowing pressure, the blowing pressure has to change considerably to alter the phase of the jet wave by as much as half a wavelength. Actually there is about a ninefold range of pressure over which the pipe can emit its fundamental sound if other conditions are right. In practice, however, the pipe will overblow to a mode of higher frequency before the upper end of this pressure range is reached.

We should point out that the spiral can have more than one loop extending far enough into the left side to overcome dissipative losses in the pipe and generate a stable sound. The second such loop, which is generally the only other one that can cause the pipe to sound, corresponds to about three half wavelengths on the jet. Because the jet has only a small admittance at this point the sound generated is weak compared with the sound for a point in the outer loop of the spiral.

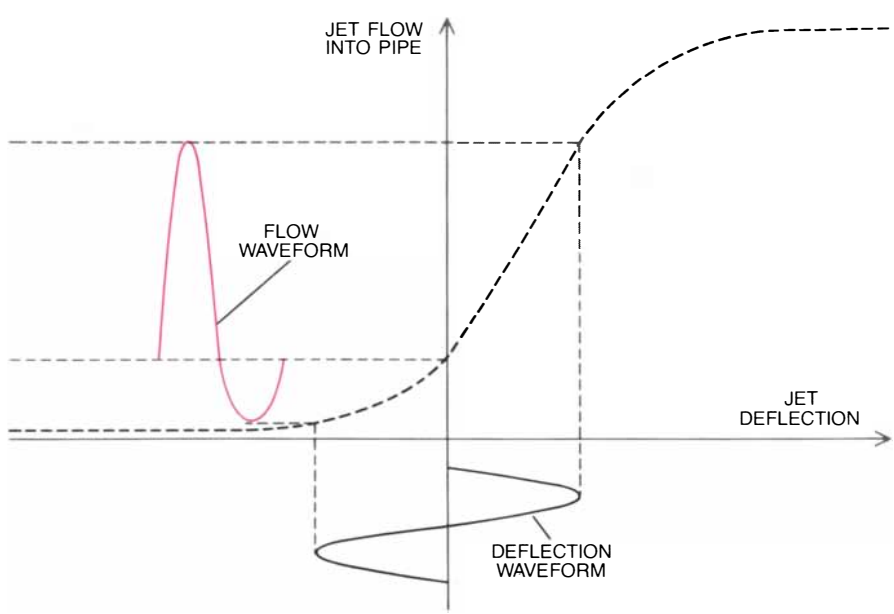
The admittance spiral we have been discussing has an additional complication if the deflection of the jet at the upper lip becomes greater than the width of the jet. The jet then blows nearly completely into and out of the pipe mouth in each cycle and the impulse it gives to the reflected wave in the pipe becomes independent of further increases in amplitude. The efficiency of the jet as a generator decreases accordingly The admittance spiral simply shrinks in size as the amplitude of the jet deflection increases.

The loss of jet efficiency with increasing deflection amplitude is accompanied by increasing energy losses in the organ pipe. The pipe oscillation rapidly settles down at an amplitude for which the energy supplied by the jet exactly balances the energy lost by the pipe. Rather surprisingly it turns out that in most instances the energy losses through turbulence and viscosity are much larger than the losses through sound radiation from the mouth and the open end of the pipe.

The sound of a real organ pipe is not, of course, limited to a single frequency; it has many components of higher frequency. One can demonstrate that these components are all exact harmonics of the fundamental frequency: integer multiples of that frequency. Under steady blowing conditions the shape of the sound wave seen on an oscilloscope remains exactly the same. Even the smallest deviation from exact integer multiples for the frequencies of the components would lead to a slow



ACOUSTIC ADMITTANCE OF THE JET as a function of frequency or of blowing pressure takes the form of a spiral in which the distance from the origin represents the magnitude of the admittance and the angular position represents the phase relation between the acoustic flow out of the mouth of the pipe and the pressure just inside the mouth. When the outflow is in phase with the pressure, the admittance lies in the right half of the spiral and the energy of the jet is being dissipated. For the jet to act as an acoustic generator the admittance must lie in the left half of the spiral, which requires that the back-and-forth displacement of the jet be offset, or delayed, in phase with respect to the pressure inside the mouth of the pipe. The wave reflected from the jet is then larger than the incident wave. The reference angle from which the phase delay is measured depends on which of two excitation mechanisms discussed in the text is dominant, the one proposed in the last century by Hermann von Helmholtz or the one proposed by Lord Rayleigh. When the admittance falls in the upper half of the spiral, the jet lowers the natural resonance frequency of the pipe; when it falls in the lower half of the spiral, the jet raises the resonance frequency. The outer parts of the spiral have been reduced in size for clarity.

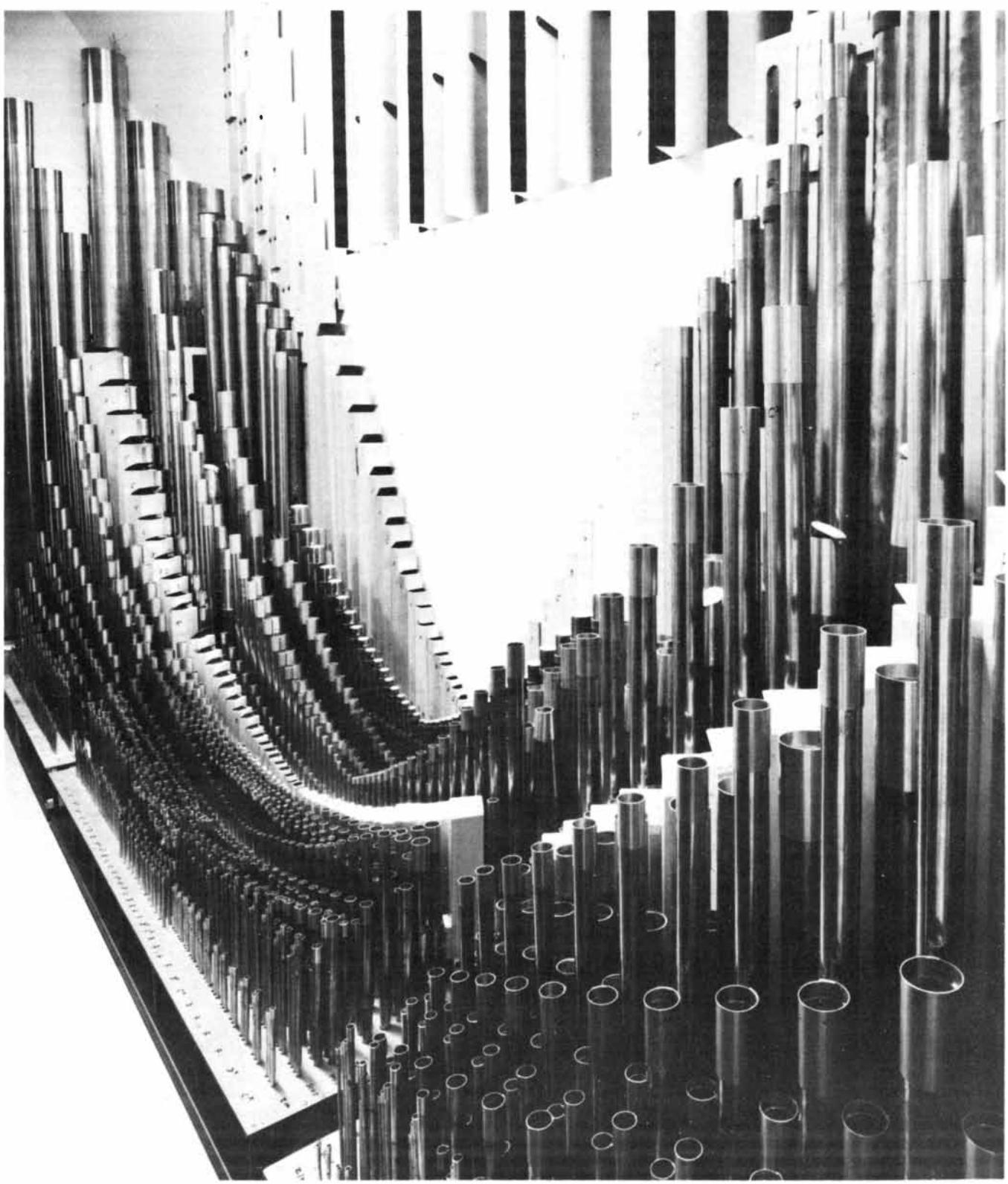


CURVE GIVING THE JET FLOW INTO A PIPE (*broken curve*) **for a given deflection is asymmetric around zero deflection because the lip of the pipe has been adjusted to cut the jet slightly away from its central plane. When the jet is deflected in a simple sinusoidal way with a large amplitude** (*solid black curve*)**, the jet flow into the pipe** (*colored curve*) **"saturates" first at one end of its range, where the jet is blowing completely outside the pipe lip. For a still larger amplitude the jet flow will also saturate at the other end of its range, where the jet is blowing completely into the pipe. The offset of the lip gives the flow waveform an asymmetric shape that has frequency components at all integer multiples of the deflection frequency. It is these harmonic components that are responsible for the rich sound of a well-adjusted organ pipe.**

but clearly discernible change in the waveform.

This behavior is interesting because the resonances of the air column in an organ pipe, or indeed in any open pipe, have frequencies that differ slightly from exact harmonic multiples. The reason is that the effective length of the pipe decreases slightly with increasing frequency because of changing acoustic flow at the open ends. As we shall see, the harmonics in the sound of an organ pipe are generated by the interactions of the jet and the pipe lip, so that the pipe acts largely as a passive resonator as far as the upper harmonics are concerned.

The resonances of a pipe develop when there is maximum movement of the air at the open ends of the pipe, which is the same as saying that the admittance of an organ pipe should be at a



INTERIOR VIEW OF ORGAN at the Sydney Opera House shows some of its 26 ranks of pipes, most of which are of metal but some of which are of wood. The length of the speaking part of each pipe doubles at every 12th pipe; the pipe diameter doubles at about every 16th pipe. Through long experience master organ builders arrived at the proportions necessary for achieving balanced tone quality.

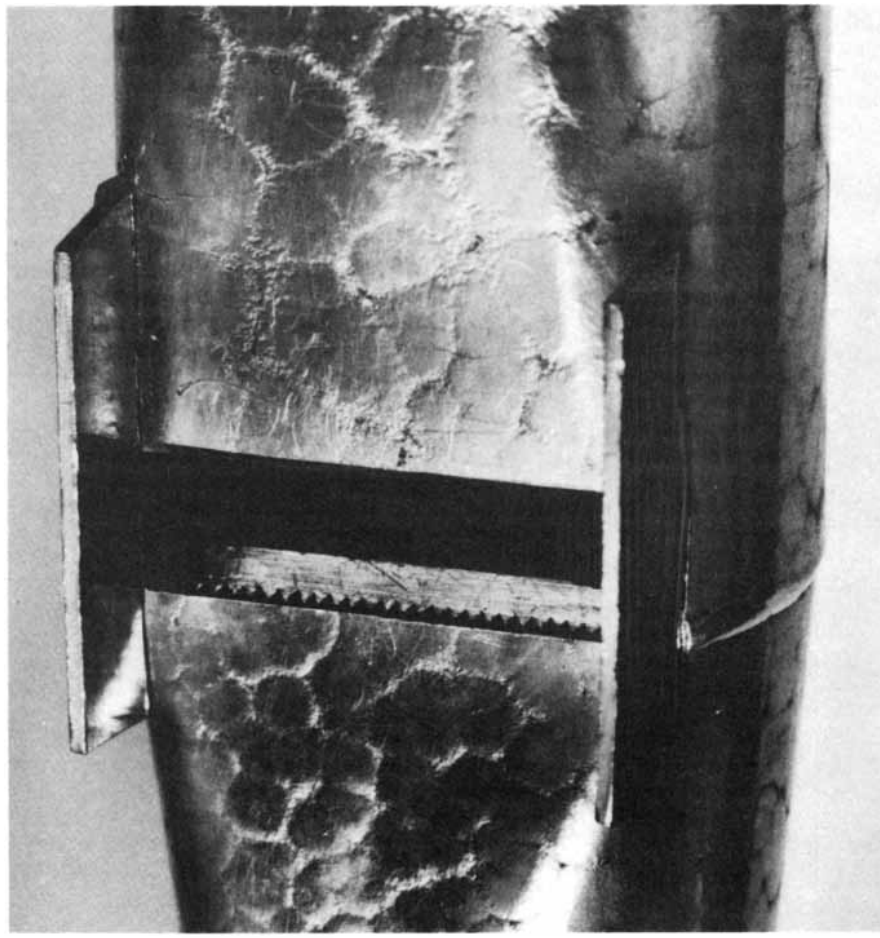
maximum at the mouth. This leads to the conclusion that the resonances of a pipe open at its far end occur at frequencies for which the pipe length is an integral number of half wavelengths for sound in air. If the fundamental frequency is f_1, the upper resonances are $2f_1$, $3f_1$ and so on. (Actually all the upper resonance frequencies are stretched a little beyond these ideal values, as we noted above.)

For a pipe that is stopped, or closed, at its far end, resonances occur when the pipe length is an odd number of quarter wavelengths. Therefore a stopped pipe need be only half as long as an open pipe to produce the same note; its resonances are at f_1, $3f_1$, $5f_1$ and so on.

Returning to the organ-pipe jet, we have seen that high-frequency disturbances tend to die out along the jet as the width of the jet increases, so that, almost independently of the higher-frequency components in the acoustic field at the pipe mouth, the tip of the jet at the upper lip moves back and forth almost sinusoidally at the frequency of the fundamental component of the pipe sound. A sinusoidal flow of the jet, however, does not lead to a sinusoidal jet flow into the pipe, because the flow "saturates" by flowing completely inside or outside the upper lip at each end of the jet's deflection range. More than that, the lip is usually offset so that it does not cut the jet exactly along its central plane, thus making the saturation asymmetric. The waveform of the jet flow into the pipe therefore has all the harmonics of the fundamental frequency, locked into precise frequency and phase relation, and the relative amplitudes of the high-frequency harmonics increase rapidly as the amplitude of the jet deflection increases.

In a typical organ pipe the distance the jet is deflected in the mouth is on the same order as the width of the jet at the upper lip. As a result there is a wide spectrum of harmonics in the jet flow. If the jet were to strike the lip exactly symmetrically, the even-numbered harmonics would be absent from the excitation. The jet is usually given a small offset so that all the harmonics are present.

As one might expect, the quality of sound delivered by an open pipe is rather different from that delivered by a stopped pipe. The frequencies of the harmonics in the driving force produced by the jet are exact integer multiples of the fundamental frequency of the wave in the jet. The air column in the pipe will respond strongly to a particular harmonic only if the pipe's acoustic admittance is large, which corresponds to a sharp resonance peak close to the frequency of the harmonic. A stopped pipe, which has only odd-numbered resonance peaks, therefore suppresses all the even-numbered harmonics of its fundamental. The result is a characteristic "hollow" sound in which the even-numbered harmonics are weak, although they are not totally absent. An open pipe, on the other hand, achieves a "brighter" sound because it responds to all the harmonics of its fundamental frequency.



MOUTH OF AN ORGAN PIPE OF THE DIAPASON FAMILY reveals how the languid of the pipe has been "nicked" in order to achieve homogeneous turbulence in the air jet. The pipe is made out of "spotted metal," a tin-rich alloy of lead and tin that solidifies to the surface pattern seen here when it is cast in a metal sheet. The sheet is then formed into a pipe.

The resonance characteristics of a pipe are determined largely by its energy losses. These losses are of two types: (1) viscous and heat-conduction losses to the pipe walls, and (2) radiation losses from the mouth and the open end. The viscous and heat-conduction losses are more important for narrow pipes than they are for wide pipes and turn out to be more important at low frequencies than they are at high frequencies. The reverse is true for radiation losses.

The upshot is that, for a given pipe length and hence for a given fundamental frequency, wide pipes are efficient and well-tuned resonators for only the fundamental and its first few harmonics, which results in a dull, "fluty" sound. Narrow pipes are good resonators for a large range of harmonics, and because higher frequencies are radiated more efficiently than lower ones the tone is thin and "stringy." Between these two extremes are pipes with the bright, ringing sound characteristic of a good organ. These are the pipes called principals or diapasons.

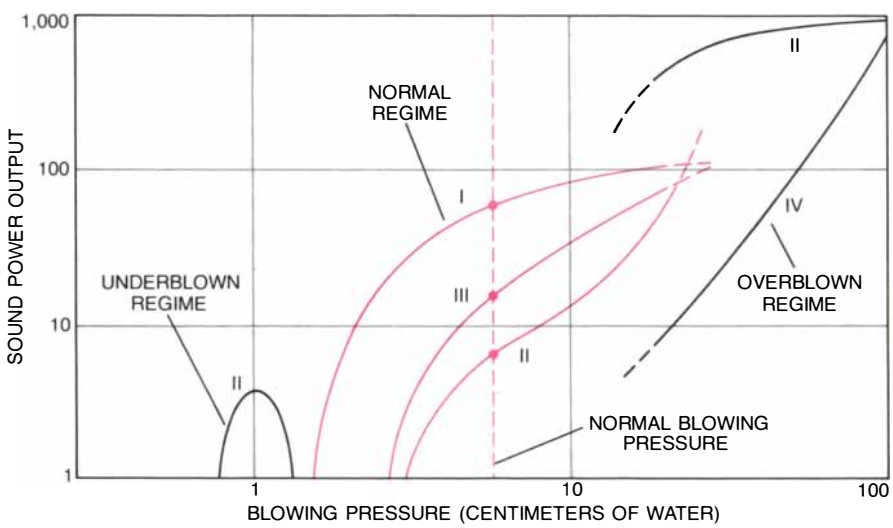
A large organ may in addition have ranks of pipes with a conical body, a perforated stopper or other geometric variations. These features are designed to modify the resonance frequencies of the pipe and in some instances to enhance a narrow range of the higher harmonics to produce a particular tone color. The material from which the pipe is made is of only minor importance.

The fact that the air in a pipe can vibrate in any one of an entire set of possible modes introduces additional complications in the acoustic behavior of pipes. For example, if the blowing pressure in an open pipe is increased so that there is just a quarter of a wavelength of the first mode, f_1, on the jet, the point in the admittance spiral for this first mode will move into the right half and the jet will cease to maintain the mode. At the same time the second mode frequency, $2f_1$, is such that it corresponds to half a wavelength along the jet, so that it can be strongly driven. The sound of the pipe

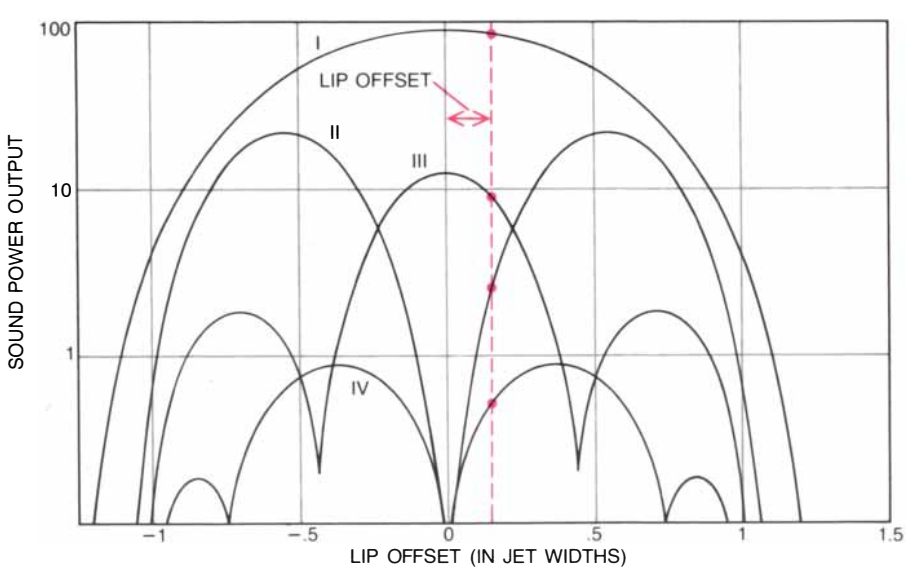
will therefore jump to this second mode, almost an octave above the first, the exact frequency depending on the resonance frequency of the pipe and on the blowing pressure.

A further increase in the blowing pressure can excite the next mode, at $3f_1$, if the cut-up of the pipe lip is not too great. Conversely, it often happens that a low blowing pressure inadequate to drive the fundamental will softly drive one of the higher modes in the second loop of the admittance spiral. These overblown or underblown pipe sounds, although they are interesting in the laboratory, are not exploited in pipe organs except for special effects.

The main problem facing the organ builder who has created a single pipe of the appropriate sound quality is to design a rank of pipes covering the musical compass of the keyboard and matching one another in loudness and harmonic content. A set of geometrically similar pipes, simply scaled up or down in size, does not meet this requirement: the pipes differ in how wall and radiation losses affect different frequencies. In order to achieve constant behavior across the acoustic spectrum one must apply a scaling rule. The diameter of the pipe is varied with the length of the pipe raised to a suitable power, k, whose value is less than 1. The result is that the long bass pipes are made somewhat narrow. The theoretically determined value of k is 5/6, or .83, but when the psychophysics of human hearing is taken into account, a value of .75 proves to be more satisfactory. That value turns out to be close to the empirical rule developed by the great organ builders of the 17th and 18th centuries.



EFFECT OF CHANGING BLOWING PRESSURE on sound output is depicted for a typical organ pipe. The Roman numerals represent the first few harmonics. The central regime (*color*) covers the range of normal speech, with a good balance at normal blowing pressure. At higher pressures the pipe overblows to its second mode; at low pressures there is a weak second mode.

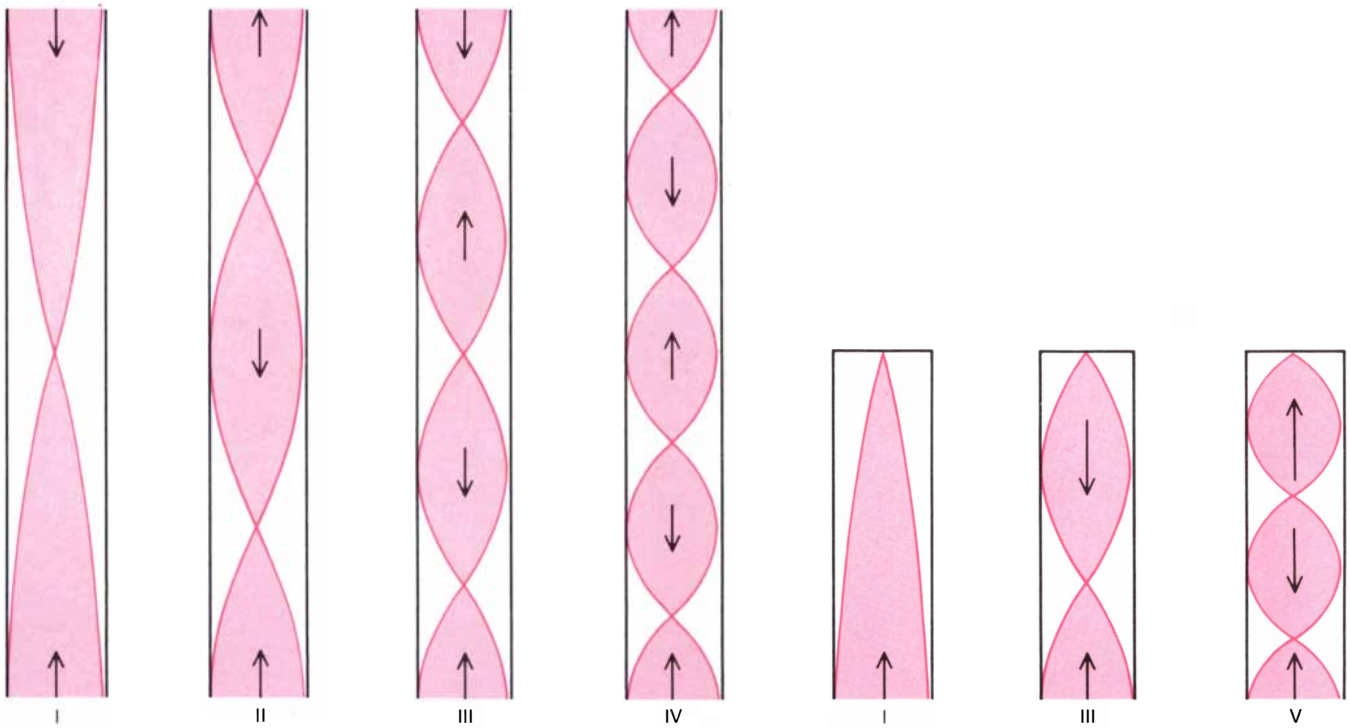


EFFECT OF LIP LOCATION on harmonic components shows the desirability of an offset. If the lip were exactly centered on the jet, the pipe would deliver only the fundamental frequency (*I*) and the third harmonic (*III*). With the lip offset at the location shown by the broken line, the second and fourth harmonics are introduced, greatly enriching the quality of the sound.

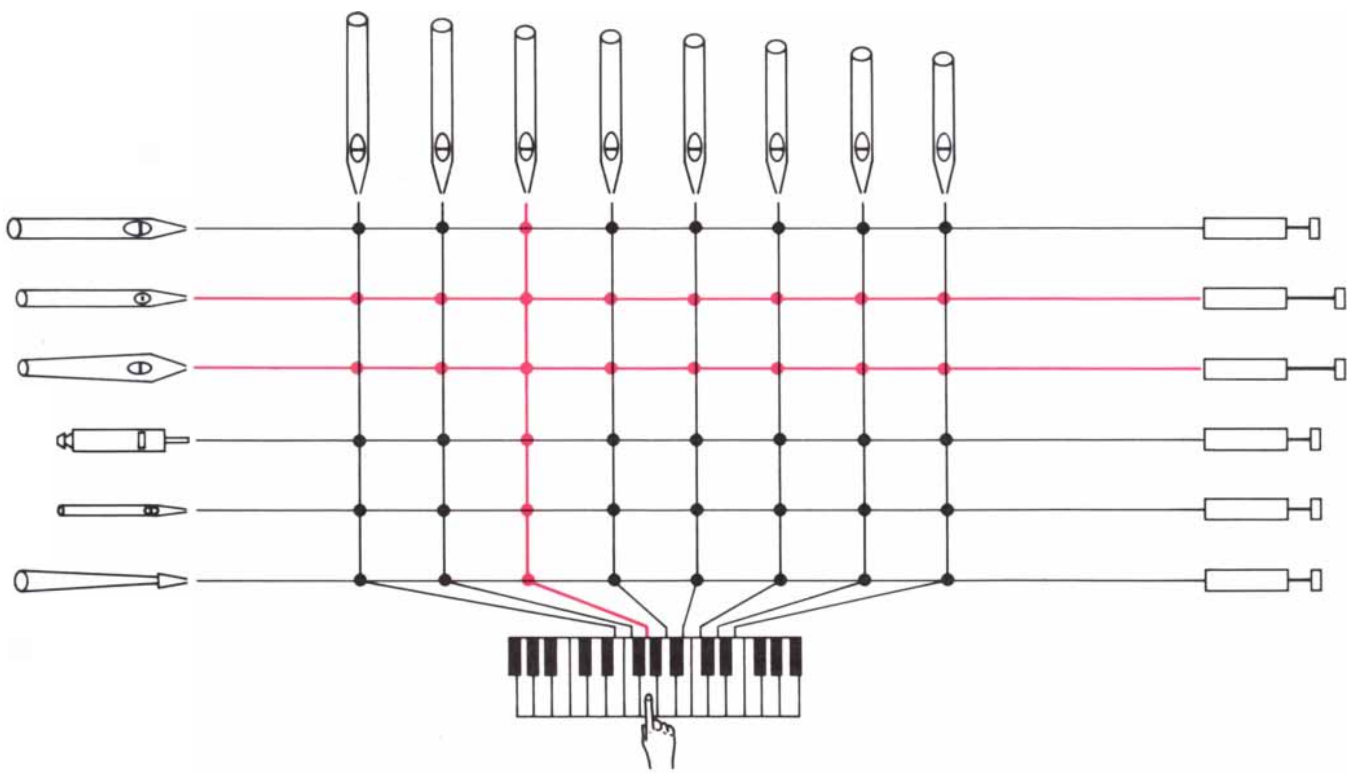
Finally, and of course important from the viewpoint of the player, comes the mechanism by which the speech of the multitude of pipes in a large organ is controlled. The basic design is delightfully simple and resembles the rows and columns of a matrix. The pipes are laid out in ranks, corresponding to the rows of the matrix. All the pipes in each rank have one tone quality, with one pipe for each note of the keyboard or pedal board. The air supply to each rank is controlled by a drawstop knob with the name of the rank written on it. The air supply to the pipes associated with each note, or column of the matrix, is controlled by the keyboard key for that note. A given pipe will sound if and only if it is in a rank for which the drawstop is pulled and on a key channel for which the key is depressed.

In these days one can imagine a host of ways of implementing such a scheme based on digital logic and electrically operated valves under each pipe. Early organs had simple mechanical levers and pallet valves to let air into the key channels, and perforated mechanical sliders to control the entry of air into an entire rank of pipes. Apart from having the virtues of simplicity and durability, this mechanical system gave the player precise control over the rate of opening of the key-channel valve and a feeling of closeness to the operation of an otherwise rather mechanical instrument.

In the 19th century and the early 20th large organs were built with all manner of electromechanical and electropneumatic mechanisms, but recently the emphasis has shifted back to mechanical action for the keys and the pedals, with sophisticated electronic controls for setting stop combinations as the instrument is being played. The organ in the concert hall of the Sydney Opera House, for example, is the largest organ in the world to have a mechanical action. Completed in 1979, it has 10,500 pipes in 205 ranks distributed across five keyboards and a pedal board. The key action is mechanical but is duplicated by an electrical action so that electric couplers can be used. Thus a live performance can be recorded digitally and the tape can be subsequently used to operate the organ, thereby re-creating the original performance. The stops and combination pistons are all electric or electropneumatic and are under the control of microprocessors having versatile stored-program capabilities. In such ways the best resources of modern technology, combined with traditional design and the skills of individual craftsmen following principles known for centuries, create the flexible and glorious sound of a large organ.



STANDING-WAVE RESONANCE PATTERNS are represented schematically for pipes open or closed at their upper end. The width of each colored pattern corresponds to the acoustic vibration amplitude in various parts of the pipe. The actual air motion is parallel to the axis of the pipe. Arrows show the direction of air motion during one half of the vibration cycle; during the other half the direction is reversed. The Roman numerals are the harmonic numbers and are proportional to the vibration frequencies. An open pipe has resonances for all the harmonics of its fundamental. A stopped pipe need be only half as long to produce the same note but has resonances for only the odd harmonics. Vibration patterns actually extend slightly beyond the ends of the pipe (calling for the end correction), and the complex geometry of the pipe mouth distorts the patterns somewhat at the lower end of the pipe without altering their "basic" nature.



PIPES OF AN ORGAN are arranged like the rows and columns of a matrix. In this simplified diagram each row, which is called a rank, consists of pipes of the same type with one pipe for each note (*top*). Each column, which is associated with one note on the keyboard (*bottom*), provides access to all the pipes of different kinds (*left*). The drawstops of the organ console (*right*) admit air to all the pipes of a rank; the keys of the keyboards admit air to all the pipes in the key channel for that note. The mechanism is arranged so that air reaches a pipe only if both its row and its column are activated. In the situation that is depicted in the diagram just two pipes will sound.

The Origin of the Cosmic X-Ray Background

What accounts for the diffuse glow of X-radiation that appears to fill the universe uniformly in all directions? New findings suggest that one source may be a multitude of distant quasars

by Bruce Margon

A fundamental observation about the structure of the universe can be made with the unaided eye on any clear night: The space between the stars is black. The fact that most of the visible light in the night sky emanates from individual stars, and that these light sources are in turn grouped into galaxies, is one of the most basic findings in cosmology. An observer surveying the heavens at wavelengths in the ultraviolet, infrared or radio regions of the electromagnetic spectrum finds this same clumping of the constituents of the universe. At the much shorter wavelengths of X-radiation, however, the situation is quite different. Since 1962 it has been known that every part of the sky emits a uniform glow of X rays. After two decades of intensive study the origin of this diffuse X-ray background is still a subject of controversy.

The sky is uniformly bright at only one other range of wavelengths: in the microwave region of the spectrum. It is a curious accident of astronomical history that the microwave background radiation was discovered in 1965, soon after the first observation of the X-ray background. The origin of the microwave background has in the meantime been identified, and the phenomenon is widely recognized as a profound clue to the evolution of the universe. Most astrophysicists regard the microwave glow as a remnant of the big bang, the massive explosion that began the current expansion phase of the universe.

X rays and microwaves differ in wavelength by a factor of about a million, and in nature the two kinds of electromagnetic radiation are normally generated under very different physical conditions. Accordingly there is no reason to expect any direct connection between the two kinds of cosmic background radiation. Nevertheless, the microwave background stands as both a tantalizing hint and a nagging frustration to students of the X-ray background. The hint is that a better understanding of the origin of the X-ray background, although still elusive, may eventually reveal clues to the large-scale structure of the universe that are just as profound as the inferences drawn from the microwave background. The frustration is that although the study of both phenomena began at roughly the same time, the microwave background is now understood in some detail, whereas there is no general agreement on even the basic physical mechanism underlying the X-ray background.

This uneven progress in deciphering the two kinds of cosmic background radiation can be attributed in part to the different observational techniques employed at the two wavelengths. Microwaves, which range in length from about .1 centimeter to 10 centimeters, can readily penetrate the earth's atmosphere and hence can be detected by ground-based antennas. In the spectral region where the X-ray background was discovered, between one angstrom and 10 angstroms (that is, between a 100-millionth and a 10-millionth of a centimeter), the earth's atmosphere is totally absorbent. Astronomical observations at these wavelengths therefore require that the detector be placed completely above the atmosphere. German V-2 rockets captured by the U.S. at the end of World War II gave American astronomers their first opportunity to lift astronomical instruments into space. A rather primitive X-ray detector carried on one of those early flights revealed that the solar corona, the extremely hot, dilute outermost layer of the sun, is an identifiable source of X-radiation.

In one sense the discovery of X rays from the sun was a triumph. It marked the advent of a new observational technique in astronomy. In another sense it was a disappointment. The intensity of the solar X rays is extremely feeble: about a millionth of the sun's visible-light output. Armed with this ratio, and with the knowledge that the sun is a normal star, one can easily compute the expected X-ray emission from the roughly 100 billion other stars in our galaxy. The calculation shows that even the nearest of the sun's neighboring stars would be expected to deliver 40 billion times less X-radiation to the earth than the sun does. In other words, X rays are detected from the sun not because it is a remarkable object but only because it is a nearby one. Thus the first achievement of X-ray astronomy—the detection of X rays from the sun—seemed to indicate that X-ray emission would be a property of very limited usefulness in stellar astronomy, at least for the study of the vast majority of normal stars.

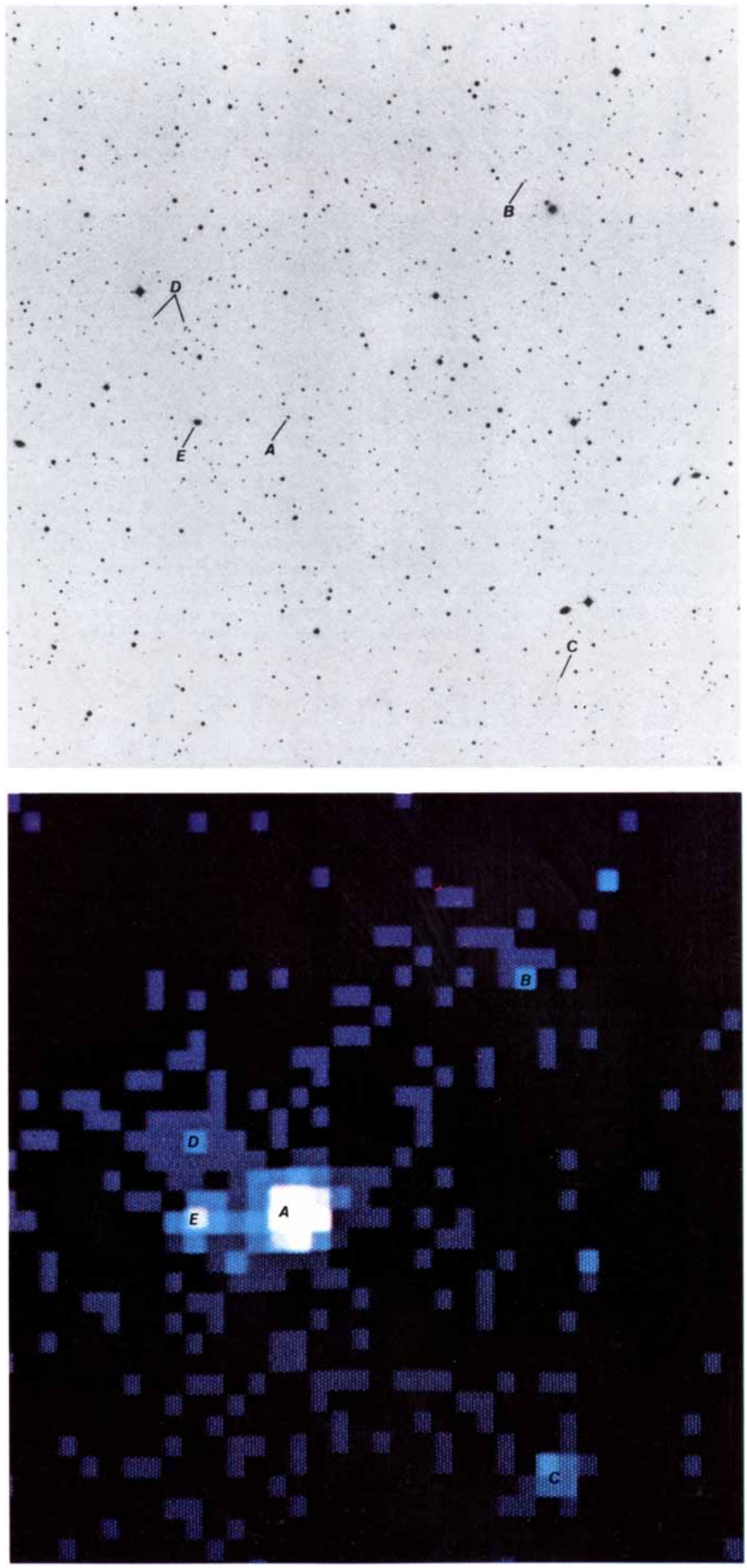
Fortunately this argument did not deter Riccardo Giacconi and his colleagues at American Science & Engineering, Inc., from exploring the potential role of X-ray observations in the study of abnormal celestial objects. In 1962 they built an improved X-ray detector that was briefly lifted above the atmosphere on board a small rocket. The new detector was substantially more sensitive than the X-ray detectors that had been carried aloft to observe the sun; it was the first such device flown that had sufficient sensitivity to detect X-ray sources outside the solar system, if indeed any such sources existed.

Giacconi's experiment resulted in two spectacular discoveries. First, a discrete point source of X rays was found in the direction of the constellation Scorpius. Although the X-radiation from this source was a million times fainter than that from the sun, it was still much more intense than the predictions had indicated the X-radiation from even the nearest normal star could be. The newly discovered object, designated Sco X-1, was only approximately located in the sky by the initial observation, and so it was not possible to determine whether the source of the X-ray emission corre-

sponded to a visible object. Successive generations of experiments flown by groups of astronomers from a number of countries have since found several hundred such X-ray-emitting objects in the galaxy and have analyzed them in increasing detail, to the point where these galactic X-ray sources are now understood quite well.

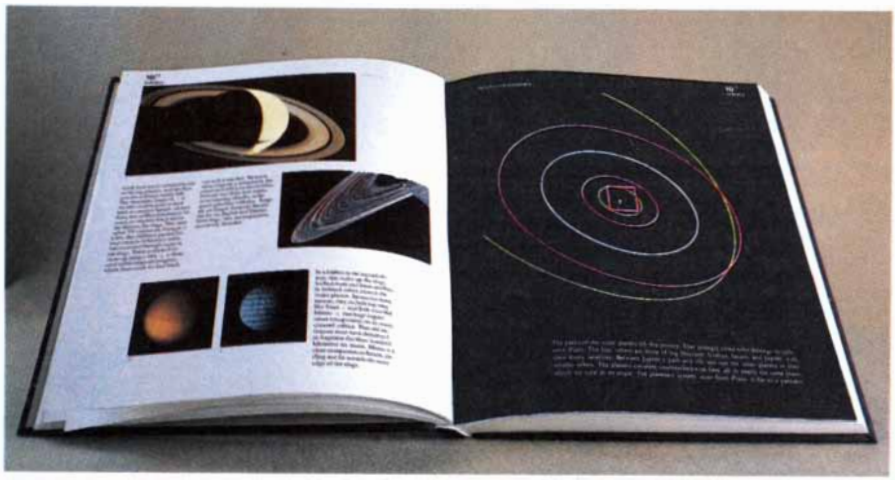
The prediction that it would not be possible to detect X rays from more than a few of the nearest normal stars turned out to be correct. X-ray sources such as Sco X-1 are much more distant and distinctly abnormal stellar systems. The X-radiation from these sources is generated by a process that involves the exchange of stellar material between a fairly ordinary star and a highly evolved, compact star orbiting nearby; the compact companion in such an X-ray-emitting binary system may be a white dwarf, a neutron star or possibly even a black hole.

The few minutes of data from this first "extrasolar" X-ray experiment provided a surprise in addition to the dis-



QUASARS AND NORMAL STARS appear quite similar in an ordinary, visible-light photograph (*top*), but they can be readily distinguished with the aid of a corresponding X-ray image (*bottom*). The photograph, reproduced here in negative form to help preserve fine details, was obtained by means of the 1.2-meter Schmidt telescope on Palomar Mountain. It shows a region of the sky measuring about half a degree on a side (approximately the angular diameter of the full moon) in the constellation Hercules. The photograph includes about 1,000 starlike images, the vast majority of which are normal stars in our galaxy. Accordingly any selection technique designed to identify the small number of quasars likely to be found in any such region must reject the foreground stars with very high efficiency. The false-color X-ray image of the same region was made from data recorded by an X-ray detector on board the orbiting satellite known as the Einstein Observatory. Only five X-ray sources are clearly discernible, four of which can be associated with quasars. The source labeled *A* is a previously known quasar that served as the target of this particular X-ray exposure; designated 3C 345, it is estimated to be some 7.2 billion light-years from the earth. The other four X-ray sources were discovered on this image. Based on spectroscopic observations made with large telescopes, the author and his colleagues were able to determine that the serendipitously discovered X-ray sources labeled *B*, *C* and *D* are in fact previously uncatalogued quasars. Sources *B* and *C* are respectively 9.5 and 6.7 billion light-years away; source *D* is the combined emission from two separate quasars, one at a distance of 7.9 billion light-years and the other at 7.2 billion light-years. Source *E* coincides with the comparatively nearby but abnormal galaxy NGC 6212; it is only about 560 million light-years away. The optical counterparts of the X-ray sources are indicated on the photograph. The X-ray image was provided by Gary A. Chanan of Columbia University.

SCIENTIFIC AMERICAN LIBRARY



MAN'S CORNER OF THE UNIVERSE lies open here at the pages in *Powers of Ten* that display the universe at the 10^{13} meters dimension (10 billion kilometers), about the diameter of the solar system.

POWERS OF TEN

A wondrous journey across the universe from its largest dimension to its smallest

Here is an atlas of the universe and a handbook to its exploration by the disciplines of science. In vivid color illustration—from the informed imagination and from the instruments of science—*Powers of Ten* shows how the universe looks from the largest to the smallest dimensions now in the grasp of human understanding.

This spectacular journey, with its sweeping changes in perspective, proceeds in ordered steps, each a multiple of ten of the last step or of the next.

There are 42 of these equal logarithmic steps from the outermost regions to the innermost recesses of the known universe. More than half were taken in the lifetime of scientists now alive.

The visual magic of this first-time-in-history journey is the bold invention of Charles and Ray Eames, who have here turned their mastery of multi-media technology to book design. Your tour guides are Philip and Phylis Morrison. They supply instructive and poetic pages of commentary, answering questions and showing how the sciences get those answers at each of the 42 steps.

The SCIENTIFIC AMERICAN LIBRARY publishes *Powers of Ten* as its premier volume because it sets out the agenda of the books that are to come. Each succeeding book further illuminates the ideas and images displayed, step by step, on the pages of *Powers of Ten*.

The SCIENTIFIC AMERICAN Library opens an entirely new approach to understanding and sharing in the work of science. Each specially commissioned volume

. . . comprehends a large topic that engages more than one discipline and, typically, recurs in many articles in SCIENTIFIC AMERICAN;

. . . is written by a scientist who has made his own pivotal contribution to the work on the topic;

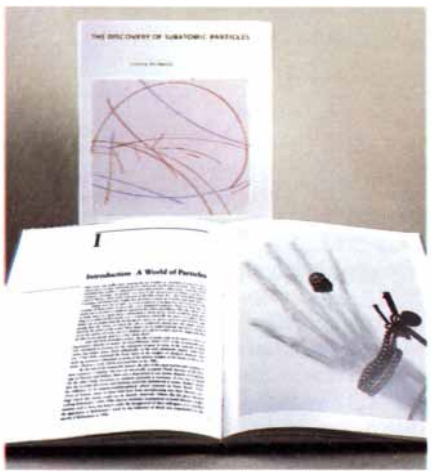
. . . rewards its reader not with "information" but, thanks to the sympathetic collaboration of the author and editor in the preparation of text and illustration, with command of the concepts central to understanding of the work and appreciation of its meaning.

Readers of this magazine are invited to accept membership in the SCIENTIFIC AMERICAN Library. Members are entitled to receive for inspection, without any charge, each volume as it is published. Membership, cancellable at any time, will gather in time a reference library of the member's own choosing, with a reach and accessibility unprecedented in the literature of science.

A SAMPLING OF THE DIMENSIONS of the known universe explored on the journey through the Powers of Ten is shown here. At center, the Taj Mahal adorns the 10 meter world in reach of our unaided senses. Dimensions grow smaller to the left: first, crystalline skeletons of single-celled diatoms at 10^{-4} meter (.1 millimeter) and at far left, tungsten atoms, 10^{-10} meter in diameter, brought into visibility by a field-emission microscope. At right, looking across the Indian Ocean toward the South Pole through the window of Gemini II, the curvature of the Earth, 10^{7} meters in diameter. At far right, a quasar, 10^{25} meters away (one billion light-years), caught by a radio telescope and imaged by computer graphics.

SUBATOMIC PARTICLES

by Steven Weinberg



DISSECTION OF ATOM began with accidental discovery of X-rays by Roentgen in 1895.

As everyone knows, the once irreducible atom has been shown to be composed of still more fundamental particles: the electron, proton and neutron. But how do we know this? In *The Discovery of Subatomic Particles* the reader comes to know the answer just as Steven Weinberg and other physicists did, thereby sharing in the intellectual enterprise that has reshaped physics in the twentieth century.

We do not accept scientific truth on faith or authority. We know truths of this kind are grounded on experiment and observation—operations we might have performed ourselves, but in any case can perform inside our heads. With Weinberg the reader of this book reenacts the historic succession of experiments that disassembled the atom. From each experiment to the next, the reader will gain a confident understanding of the laws of physics that govern the events encountered.

HUMAN DIVERSITY

by Richard C. Lewontin

The relevance of science to human values finds compelling demonstrations in the proper understanding of the genetics of our species. In *Human Diversity* Richard Lewontin shows that each human being differs from all others because of the interaction of genetic differences, environmental differences, and chance events that occur during development. There is no unique "normal" genetic constitution.

The genetic markers that supposedly divide the species into races are only a trivial fraction of the total spectrum of diversity in which each person finds singularity. The genetic endowment held in common across racial, national and cultural boundaries is a consequence of the biological history of our species. The segregation of people into social castes can never again invoke genetics as its rationalization.

FOSSILS

by George Gaylord Simpson

One of the architects of the modern evolutionary synthesis here establishes the nature—and the true splendor—of the solid evidence upon which much of the theory and the facts of evolution rest. Simpson declares: "The primary record of the history of life is written in the successive strata of rocks as in the pages of a book. Fossils may be called the writing on those pages. Traces of organisms living at successive geological times, they represent once living things and should be seen as such. They must be put in their sequence in time. They were influenced by and bear witness to geographic and geological changes on the earth. They are basic materials for the study of organic evolution. Their study combines historical geology and historical biology into one great synthesis."

THE SOLAR SYSTEM

by Roman Smoluchowski

The fine-grained picture of the Earth and its immediate cosmic neighborhood now in the possession of mankind is here comprehensively assembled. The author has brought together the latest graphic images from the satellites and from earthbound telescopes as well. (We know the terrain of cloud-wrapped Venus from radar echoes returned to the 600-foot radio telescope at Arecibo in Puerto Rico.) He shows us our solar system in time as well as in space. In the wealth of images of the Sun, of the planets, of their satellites, of the swarming asteroids and of the cometary messengers from the primordial ice clouds on the outermost circumference of the system, we comprehend the origin, the history and the fated future of our Sun and its planets.

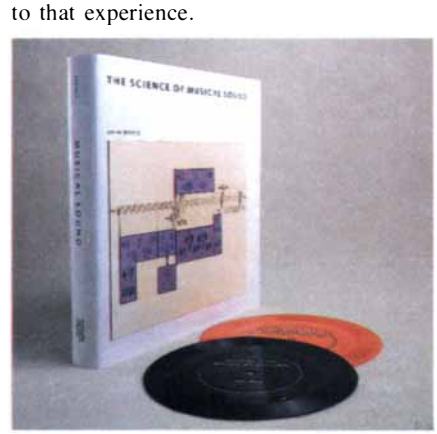
MUSICAL SOUND

by John R. Pierce

In this volume, the author will share with you his joy in the discovery that understanding from physics and mathematics enlarges and enhances one's appreciation of music. Thereupon you will find yourself in the company of a long succession of natural philosophers who have pursued just that strategy.

John R. Pierce is the latest of these natural philosophers. To his work in music he brings his study of sound, its transmission and reproduction, at Bell Telephone Laboratories plus the insights of such contemporary composers as Pierre Boulez, John Cage and Edgard Varèse.

In the live recordings that come with this volume Pierce demonstrates something of what the psychology of acoustics has learned about the perception, the illusion and the effect of sound. He imbues the reader with his confidence that rational enquiry into this most intensely subjective of all aesthetic experiences will open up new realms of sound to that experience.



SOUND SHEETS, for 33-r.p.m. turntable, demonstrate acoustic illusions discussed in book.

covery of Sco X-1. The rocket carrying the X-ray detector spun quickly during its brief ascent and descent, thereby sweeping out a substantial fraction of the celestial sphere. Contrary to expectations, the detector registered an apparently constant intensity of X-radiation at all times during the observations above the atmosphere, regardless of the viewing direction of the rocket. Simultaneous with the beginning of stellar X-ray astronomy, therefore, was the discovery of the diffuse X-ray background.

Like the stellar X-ray sources, the X-ray background has been the object of study by almost all subsequent experiments in X-ray astronomy. The first of these efforts called for additional rocket flights; for the past 10 years or so the detectors have been carried on board orbiting satellites. Several fascinating features of the X-ray background have emerged from these studies. Perhaps the most remarkable is the uniform intensity of the radiation from different parts of the sky. Every region of the sky exhibits an X-ray brightness that is precisely equal to that of neighboring regions, to within the best measurement accuracies currently available. (Current observations would have been able to detect fluctuations in brightness of about 1 percent between adjacent regions of the sky with an angular extent about the size of the full moon.) In other words, the X-ray background is exceedingly isotropic: whatever generates the background radiation produces an equal number of X rays from every direction in the sky, to an accuracy exceeding 99 parts in 100. (The microwave background radiation is also known to be highly isotropic, to an even greater measurement accuracy: about 2,999 parts in 3,000.)

The near-perfect isotropy of the X-ray background is clearly a vital clue to its origin. The clue is usually interpreted as indicating that the source (or sources) of the radiation is either very near or very far but not at some intermediate distance. The possibility that the source is very near is easy to understand; an observer completely immersed in the radiation from a nearby source would perceive a perfectly isotropic glow. In the case of the X-ray background, however, it is difficult to postulate any highly uniform source of X rays in the vicinity of the solar system.

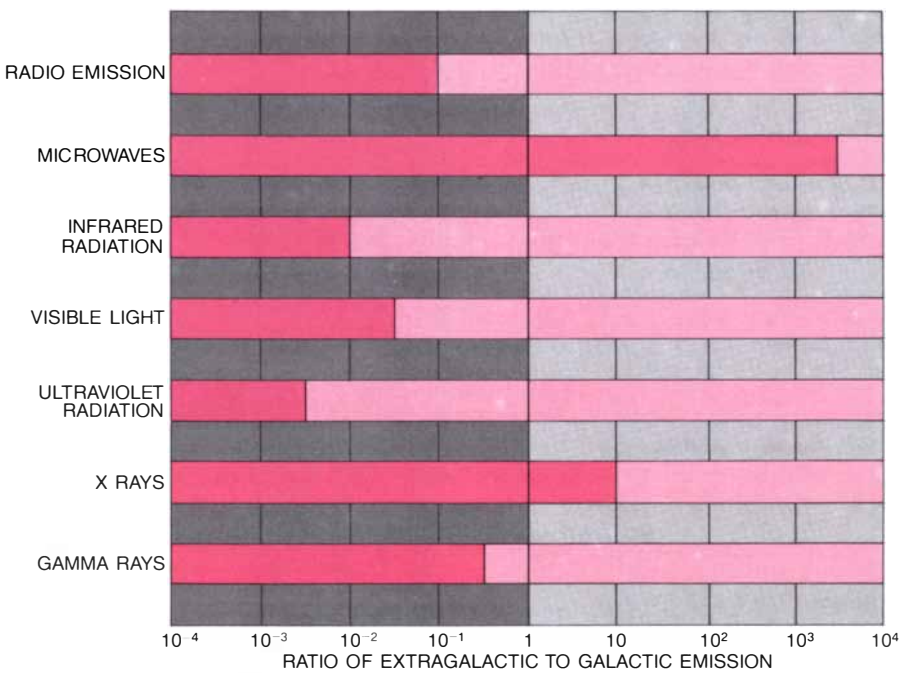
To understand why the alternative is an extremely distant origin for the radiation, consider instead a hypothetical distribution of X-ray emitters at some intermediate range of distances, for example at distances comparable to the spacing between the stars in our galaxy or even between nearby galaxies. In that case there would have to be one or more emitters in some direction that were closer to the solar system than all the other emitters. When one looked in that direction, one would expect to see a bright patch corresponding to the nearer source of radiation.

This conclusion, however, would contradict the original observation that the X-ray background is highly isotropic: there are apparently no lumps of X-ray-emitting matter in any part of the sky. The only reasonable explanation is that most of the radiation comes from enormous distances, distances that are in fact comparable to those of the most distant known objects in the universe. As with the microwave background, the eventual explanation of the diffuse X-ray background will almost surely have some bearing on the cosmological question of the large-scale structure and evolution of the universe.

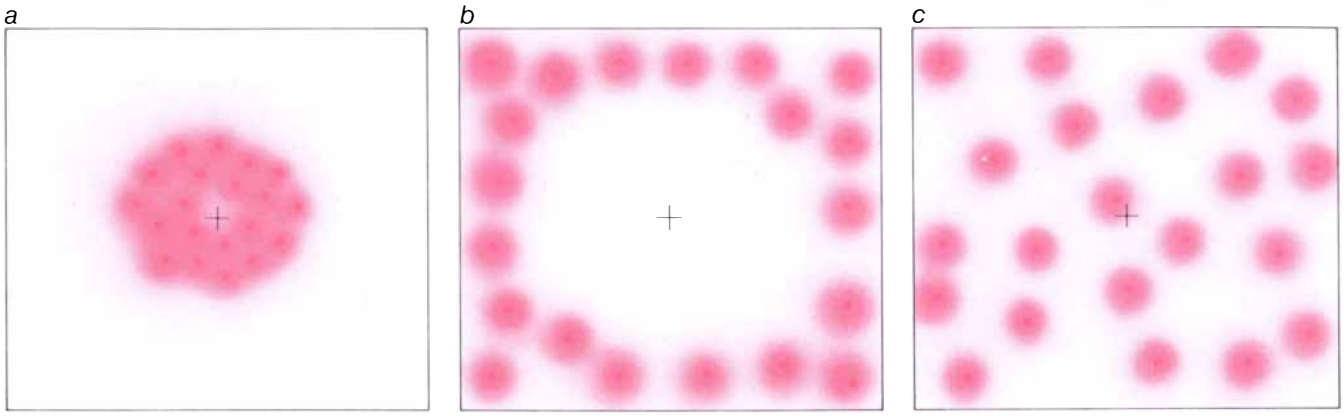
A second clue to the origin of the X-ray background arises from the fact that it is X-radiation that is generated, rather than some longer-wavelength radiation. X-radiation is one of the most energetic forms of electromagnetic radiation; hence the source of the observed emission must be some highly energetic process. Furthermore, the process must be one that operates in conditions prevalent on the scale of the universe.

One such process for the generation of X rays has been familiar to physicists for decades; it involves the extremely hot, dilute gas called a plasma. In such a gas the electrons and the protons, which are normally bound together to form atoms, are present as an unbound mixture of rapidly moving particles. Since an electron and a proton have opposite electric charge, there is an attractive force between nearby pairs of electrons and protons in the plasma, and near-misses of free electrons attracted to free protons but not colliding with them are common. Since the advent of quantum physics early in the century it has been known that a free electron passing close to a free proton loses some energy, and that the lost energy is carried away by a discrete packet of electromagnetic energy: a photon. The radiation created by the interaction of an electron and a proton is called bremsstrahlung (German for "braking radiation").

The exact wavelength (or range of wavelengths) of the photons created by the bremsstrahlung process is sensitively related to the temperature of the plasma. For a plasma with a temperature in the range between 10 and 500 million degrees Kelvin bremsstrahlung is expected to generate X rays with wavelengths



RATIO OF ELECTROMAGNETIC ENERGY measured by a hypothetical observer outside our galaxy to that known to be present within the galaxy is represented by the colored bars for different parts of the spectrum. At five of the seven spectral ranges shown the ratio is less than 1, indicating that the energy per unit of volume at these wavelengths is greater within the galaxy than it is outside. In other words, the outside observer would perceive that the dominant source of the radiation at these wavelengths is concentrated into galaxies, with little radiation being emitted from the space between galaxies. At two wavelength ranges, however, the reverse is true: the energy of the radiation that fills the universe is greater than the energy of the radiation within the galaxy at the same wavelengths. In the latter case the outside observer would have difficulty distinguishing galaxies, since they would be overwhelmed by the diffuse background radiation. Unlike the microwave background radiation, which is generally interpreted as being a remnant of the big bang, the X-ray background is still a mystery. The chart was adapted from one devised by Joseph Silk of the University of California at Berkeley.



EXTREME ISOTROPY, or uniformity, of the X-ray background implies that it originates either very close to the earth or very far from it but not at some intermediate distance. If the sources of the radiation are all nearby (*a*), an observer would be completely immersed in the radiation and would therefore see a uniform glow. If the sources are all extremely far away (*b*), the distance from the observer to various sources may differ slightly, but the difference would be small compared with the enormous average distance and therefore no bright patches would be observed. If the radiation originates at a variety of intermediate distances (*c*), however, the difference between a nearby source and a distant source would become important; one would then expect to see some part of the sky, corresponding to the nearer source, as a bright patch. Various lines of evidence point to an extremely distant origin for the diffuse X-ray background, although it is still uncertain whether the source is actually a diffuse gas, an ensemble of discrete objects or some combination of the two.

predominantly between one angstrom and 10 angstroms. Is there any reason to expect that the universe is more or less uniformly filled with a very hot plasma that generates by the bremsstrahlung process the radiation observed today as the X-ray background? There are a number of reasons to suspect the existence of just such a hot gas.

The space between the stars in the galaxy, although it is an excellent vacuum by terrestrial standards, contains a significant amount of dilute gas. The interstellar medium accounts for only about 10 percent of the mass represented by all the stars in the galaxy, but it plays a crucial role in galactic evolution: it is the raw material out of which new stars form. Astronomers have long sought to understand how, by analogy, the enormous agglomerations of matter that ultimately become individual galaxies form. It seems likely, although it is by no means certain, that galaxies condensed from a dilute gas that filled the entire universe shortly after the big bang: the intergalactic medium.

The hypothetical nature of this scenario must be recognized; although the interstellar medium can be observed directly and star formation is known to be a continuing process in our galaxy, the intergalactic medium has not been observed directly and the formation of galaxies apparently stopped throughout the universe billions of years ago. Nevertheless, there is reason to suspect that some intergalactic medium may still exist. Few processes in nature work with an efficiency of 100 percent, and so it would be surprising if the process of galaxy formation swept away every atom of the original medium.

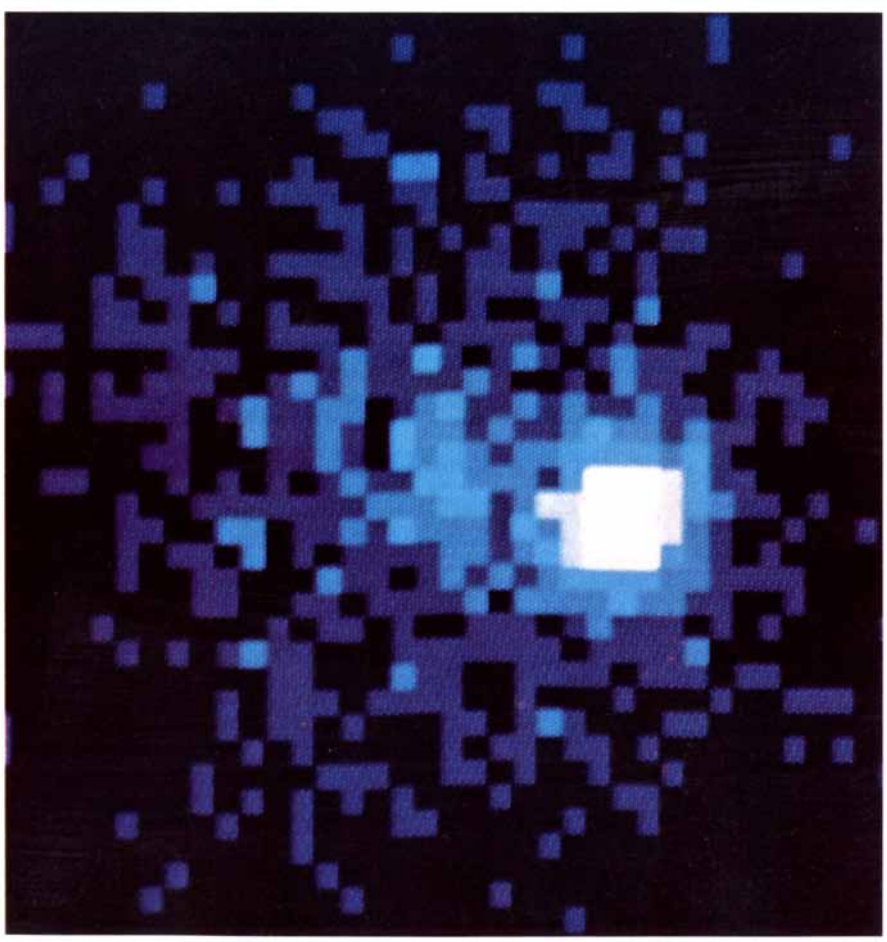
If the diffuse X-ray background is to be explained as bremsstrahlung from an intergalactic medium, the question arises of whether it is reasonable to suppose such a plasma would have the extremely high temperatures required to generate X rays. There are two considerations that indicate this assumption may be plausible. The first is an argument by elimination. If the intergalactic medium exists, and if the gas were as cold as the interstellar medium, then such a gas, intervening between our galaxy and very distant objects, would give rise to noticeable absorption features in the visible-light spectrum of those distant objects. Such absorption features, which can be shown to have a very specific signature, have been thoroughly searched for, and they are normally absent. If there is any cold intergalactic gas, it is certainly not ubiquitous enough to fill the universe isotropically. Hence by default one is led to suspect a hot intergalactic gas; if such a gas exists in any significant quantity, there is strong observational evidence that it is not cold.

The second plausibility argument concerning a hot intergalactic gas involves a calculation of how long the gas could remain hot. If such a medium were expected to cool rapidly compared with current estimates of the age of the universe, one would be faced with the awkward situation of having to explain not only how the gas reached such a prodigious temperature originally but also what mechanism continues to reheat the gas throughout the lifetime of the universe, making the X-ray background observable today. Actually it is easy to calculate that in such a dilute plasma the cooling time would be comparable to or greater than the age of the universe.

Bremsstrahlung is simply not an efficient way of extracting energy from such a thin medium, and there is no other likely cooling mechanism. Assuming there was a way of heating the intergalactic gas once, earlier in the evolution of the universe, one would therefore expect the gas to still be hot today. A number of such possible heat sources are thought to have been present in the violent early stages of the universe. Although the temperature of the gas is very high, it can be shown that the amount of the gas in any one small volume of the universe is likely to be minuscule. Thus the total energy content of the gas, and hence the energy needed to be supplied by this ancient heating mechanism, is large but not overwhelming.

There is even existing observational precedent for significant amounts of very hot cosmic plasma, although not in a form as omnipresent as would be required for an intergalactic medium. Throughout the universe it is commonly found that individual galaxies tend to gather into large, gravitationally bound clusters. The richest clusters can have hundreds of, or even a few thousand, member galaxies. One of the early discoveries of X-ray astronomy was that rich clusters tend to be detectable sources of X rays. Detailed studies of the X rays from clusters have shown that the emission is not due to the sum of the individual sources within the member galaxies. Instead the emission is extended and diffuse, filling the space between the member galaxies and often having a spatial extent comparable to that of the entire cluster. This X-ray emission is now attributed to bremsstrahlung from a hot plasma (typically between 20 and 60 million degrees K.) that fills the volume of the cluster and coexists with the member galaxies.

The amount of matter in the intracluster medium can be estimated from the observed X-ray brightness to be comparable to the total mass of all the galaxies in the cluster. Hence at least in



these rich clusters hot plasma is hardly an exotic state of matter but rather one of the dominant forms. Over the past few years X-ray observations of the intracluster gas have yielded good evidence that it originated in the member galaxies; the gas is found to bear the signature of chemical elements such as iron that are thought to be manufactured only in stars. Thus the intracluster medium probably followed rather than preceded the existence of galaxies, and as such it is not a true representative of the hot intergalactic medium. Nevertheless, it stands as a concrete example of the fact that the universe does contain, in at least certain places, large amounts of very hot, X-ray-emitting plasma.

Suppose one accepts for the moment the hypothesis that the diffuse X-ray background can be attributed to bremsstrahlung from a hot intergalactic medium. Apart from resolving an observational curiosity of two decades' standing, would such a plasma have any other interesting properties? The answer is most emphatically yes. In the late 1920's Edwin P. Hubble and his colleagues at the Mount Wilson Observatory established through observations of the spectral features of galaxies that all the galaxies immediately outside the local group are moving away from it; in short, the universe as a whole is expanding. This finding led to the idea that at least the current phase of the evolution of the universe had a discrete beginning in a dense, violent phase; subsequent confirmation of this idea has come from the observation of the microwave background radiation.

Will the present expansion phase last forever? If it does, the universe will be a rather dull place in another 10 or 20 billion years. Not only will all the galaxies have moved far away from any neighboring galaxies but also because

RICH CLUSTER OF GALAXIES serves as proof of the existence on a cosmic scale of an X-ray-emitting plasma: a hot gas of charged particles. The visible-light photograph reproduced in negative form at the top shows the cluster Abell 2151 as viewed by the Palomar Schmidt telescope. The cluster, which is in the constellation Hercules, is about 670 million light-years from the earth. The false-color X-ray image at the bottom, again provided by Chanan, shows the same region of the sky as viewed by the Einstein Observatory. Instead of the individual galaxies seen in the visible-light image, the X-ray image reveals an extended cloud of hot gas, measuring some three million light-years across, entrained by the combined gravitational pull of the assemblage of galaxies. The intracluster medium has an angular diameter of about .25 degree, or half the angular diameter of the full moon. A plasma of this type, estimated from its X-ray emission to have a temperature on the order of 20 million degrees Kelvin, has been found in many similar clusters of galaxies.

stellar evolution is a finite, one-way process that eventually converts galactic matter into faint compact stars such as white dwarfs, neutron stars and possibly black holes, the individual galaxies will eventually lack internal sources of luminosity and will slowly fade from view. This gloomy prediction depends crucially, however, on the assumption that the expansion will continue unabated. The existence of the intergalactic medium can potentially alter the fate of the expansion.

There is a force working to counteract the universal expansion, namely the gravitational attraction of every atom in the universe for every other atom. If the self-gravitation of the universe is sufficiently strong, the expansion will eventually stop and reverse into a gravitationally driven contraction. How can one resolve this important question? Observationally astronomers today know only what Hubble realized half a century ago: the current phase of the universe is one of expansion. Remarkably, a prediction of its future fate—continued expansion or eventual contraction—is in itself not a complicated procedure. Because the magnitude of the gravitational force depends only on the amount of matter involved and the distance between the matter, and the expansion rate and the scale of today's universe can be derived at least in principle from observations of the expansion of galaxies, the question of whether or not the expansion will reverse is in fact equivalent to a calculation of the total amount of matter in the universe.

More rigorously, the mass density of the universe (that is, the amount of matter in any given small volume) is the quantity that needs to be estimated. Above a certain threshold, known as the critical density, the expansion will eventually cease and contraction will begin; alternatively, if the average mass density in the universe falls below the critical density, the expansion will continue forever. The best current estimate of the value of the critical density is exceedingly small, partly because the volume of the current universe is so large; the value is equivalent to an average of about one hydrogen atom per cubic meter throughout the universe.

Apart from the possible existence of the intergalactic medium, most of the matter in the universe is thought to exist within galaxies, in the form of stars and, to a lesser extent, gas and dust. Estimating the mass of the galaxies is a complex and indirect process, but the result in the context of this discussion is so striking that the inaccuracies are probably not critical. The conclusion is that galaxies account for only 1 or 2 percent of the critical density. Thus even allowing for uncertainties in the tally of galactic material, it seems unlikely that normal galaxies provide enough matter to stop the expansion.

The existence of the diffuse X-ray background and its possible origin in an intergalactic plasma, however, demand that one reconsider this calculation. If the mass density in the intergalactic medium exceeds the average mass density in the universe attributable to galaxies by a factor of 50 or 100, the universe would indeed have its critical density, and the prediction of the future evolution of the universe would change drastically. Furthermore, one would have discovered the bizarre fact that astronomers had previously been neglecting as much as 99 percent of the matter in the universe; the familiar, luminous galaxies and their contents would be relegated to the role of a trace impurity in the intergalactic medium!

If the X-ray background is due to the intergalactic medium, one can readily estimate the mass density of the medium by measuring the observed intensity of the background radiation (that is, how much X-radiation arrives at the earth each second from a given direction) and then calculating theoretically how many atoms per volume of space are required to create this much radiation through the bremsstrahlung process. The result of such a calculation proves to be tantalizing. If the X-ray background is due to bremsstrahlung from a hot plasma, then depending on the exact values of several rather poorly measured quantities the inferred density of hot intergalactic plasma is comparable to the critical density of the universe.

Has the discovery of the X-ray background therefore changed ideas about the future expansion of the universe? Unfortunately the background radiation is evidence, but by no means proof, of the existence of a critical density of intergalactic matter. There are three major uncertainties in this chain of argument. First, the value of the critical density depends sensitively on the expansion rate of the universe: the faster galaxies recede from one another, the more matter is needed to restrain the expansion. Estimates of this rate, embodied in the quantity known as the Hubble constant, are derived from a complex sequence of astronomical measurements subject to various uncertainties. Indeed, since Hubble's time the prevailing estimate of this constant has changed considerably. Only for a highly specific (and today somewhat unpopular) value of the Hubble constant does the inferred amount of the intergalactic medium provide the critical density of matter.

A second uncertainty is that the foregoing determination of the density of the plasma based on the observed intensity of X rays calls for a knowledge of the average lumpiness of the emitting regions, because smooth and lumpy mediums emit radiation with slightly different efficiencies. Only if the intergalactic medium is highly uniform is its inferred density equal to the critical density. Although no direct measurement of the degree of lumpiness exists, a smooth medium seems unlikely on theoretical grounds: most constituents of the universe, for example the luminous matter in galaxies, are distributed in a highly nonuniform fashion.

A third uncertainty in the conclusion that a hot intergalactic medium binds the universe is perhaps the most profound one of all: there is no direct proof that the observed diffuse X-ray background is indeed due to such a hot plasma, as opposed to some entirely unrelated mechanism. Unfortunately the X-ray emission bears no unambiguous signature of its origin as bremsstrahlung from a dilute gas, and there are other familiar laboratory processes that can also generate X-ray emission. An explanation of the X-ray background in terms of a hot plasma is consistent with current observations, and perhaps even elegant in its simplicity, but is it unique? Given the cosmological importance of the existence of the plasma, at least if it is near the critical density, one certainly should demand more than a consistent explanation. For example, it would be comforting to know there are no other existing explanations of the X-ray background that are also consistent with the data.

That, however, is not the case. There is an entirely different explanation of the diffuse X-ray background that is completely independent of the possible existence of the intergalactic medium. The competing explanation of the background has been receiving increasing attention among workers in the field in the past few years. The observation that the X-ray background is highly isotropic is a statement not only of a characteristic of the source of the radiation but also of the sophistication of the available instrumentation. Consider the possibility of a large number of discrete objects each of which emits X rays. If the objects were far enough from the observer and sufficiently numerous so that every region of sky held a large number of them, then the observer might misinterpret these discrete sources as a hazy, uniform glow. No matter how accurate a given instrument may be, there would be some hypothetical density of the individual objects in the sky such that the instrument could not resolve them as discrete sources and so determine that an ensemble of points is under observation rather than a genuinely diffuse medium.

Is the diffuse X-ray background really from a diffuse source? Or does it originate from a large number of distant but distinct emitters of X rays, quite unrelated to a dilute intergalactic gas? An ini-

tial step toward settling this issue is obviously to ask whether there is any known constituent of the universe that is both very common and a powerful source of X rays. Galaxies, the basic building blocks of the universe, do not satisfy the requirement. Although they are quite numerous, they simply do not emit enough X-radiation to add up to the observed intensity of the X-ray background.

This conclusion in turn is a consequence of the original conclusion arrived at after the first observation of X rays from the sun: normal stars do not emit a substantial fraction of their energy as X rays. Therefore the sum of the X-ray emission from all the stars in a galaxy is still feeble compared with its visible-light output. I have mentioned a second constituent of the universe that is a powerful source of X rays: the gas in rich clusters of galaxies. This source seems at first to be a promising possibility, since clusters of galaxies are relatively common. Recent observations of the X-ray emission from nearby clusters, together with an estimate of their density elsewhere in the universe, however, tend to indicate that clusters can contribute at most a few percent of the diffuse X-ray background.

In the past three years X-ray observations from the orbiting satellite known as the Einstein Observatory have revealed another class of objects as a strong candidate for the source of the X-ray background. The data show that quasi-stellar objects, or quasars, are strong sources of X-ray emission. The exact nature of quasars is still an enigma. On optical photographs they appear as points, indistinguishable from normal stars in the galaxy. Two decades ago, however, it was realized that unlike normal stars many of these seemingly ordinary starlike objects are intense sources of radio waves.

Detailed studies of the optical emission from quasars prompted by this observation soon revealed that they are receding from the earth at enormous velocities, in some cases at more than 90 percent of the speed of light. If one interprets the recession velocities as being due to the general expansion of the universe, which also causes the recession of external galaxies (a point accepted by most workers in the field but not by all), then one is forced to conclude that quasars are not stars in our galaxy but are the most distant observable objects in the universe. The most extreme examples are far beyond the most distant known galaxies, and they are emitting light that began its journey shortly after the big bang. In spite of the prodigious distances of these objects, many are bright enough to be viewed with an amateur's backyard telescope. Thus the intrinsic luminosity of quasars is enormous, far beyond that of normal galaxies, and their energy-generating mechanism is unlikely to resemble the thermonuclear fusion that powers normal stars.

An examination of carefully selected samples of faint objects on photographic plates has revealed that although few quasars are close to our galaxy, they are a fairly common constituent of the very distant universe. Apparently some evolutionary effect over the past few billion years has largely extinguished the quasar phenomenon.

Before the launching of the Einstein Observatory a handful of the closer quasars were known to be exceedingly powerful sources of X-radiation, based on observations from rockets and earlier satellites. In the most extreme cases a single quasar is known to generate 1,000 times as much energy at X-ray wavelengths as the sum of the visible-light output of all 100 billion stars in our galaxy! Yet because of the vast distances to most quasars even this staggering output of energy is so feeble by the time it reaches the earth that the majority of quasars were not accessible to X-ray observation before the launching of the Einstein Observatory.

The Einstein satellite detected X rays with a sensitivity that exceeded previous experiments by a factor of about 1,000. Furthermore, in contrast to previous detectors of the Geiger-counter type, which essentially recorded just crude X-ray intensities, the Einstein detector carried a novel focusing system that was capable of forming X-ray images, which were then radioed to the ground in digital form. Over the satellite's two-year lifetime it was ordered to obtain X-ray images of a large number of interesting objects, including several hundred quasars. Most of the quasars observed in this way were found to be powerful sources of X-radiation.

Given these X-ray observations of a limited sample of quasars, together with information from the more complete optical studies on the total number of quasars per area of sky, it is possible in principle to estimate the contribution of quasars to the diffuse X-ray background. Such a calculation yields the fascinating conclusion that a large fraction, and possibly all, of the X-ray background is in fact due to distant, individually invisible quasars. As with any survey that relies on a small fraction of a population to extrapolate to the properties of a vastly larger group, however, there are serious uncertainties in it.

In the case of the Einstein Observatory survey of X-ray-emitting quasars the difficulties are compounded. First, the optical data relied on to infer the number of quasars per unit area of sky are still somewhat controversial, and these data are crucial to link the X-ray emissivity of the small sample of objects observed to an estimate of the total contribution of the ensemble of objects. Second, as astronomers have recognized since the discovery of quasars, these strange objects seem to evolve in their properties over the lifetime of the universe. In fact, the X-ray data also provide tentative evidence for the evolution of the X-ray properties of quasars with time; according to workers at the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory, the ratio of the X-ray output to the visible-light output of the quasars they observed seems to depend on one or more other properties, including perhaps how far the object is from our galaxy.

Evidence for such an evolution in X-ray properties is dismaying, particularly because any contribution quasars make to the X-ray background must originate as a mixture of emissions from quasars throughout the universe at a variety of different evolutionary stages. Therefore the exact nature of the evolutionary process must be precisely understood to properly compute the quasars' contribution to the background. (The problem is analogous to that of computing the compound interest on a long-term bank account. If the initial amount of funds and the interest rate are known precisely, it is easy to predict the balance at any given time. If one assumes a slightly different interest rate, however, the predicted balance will be quite different.)

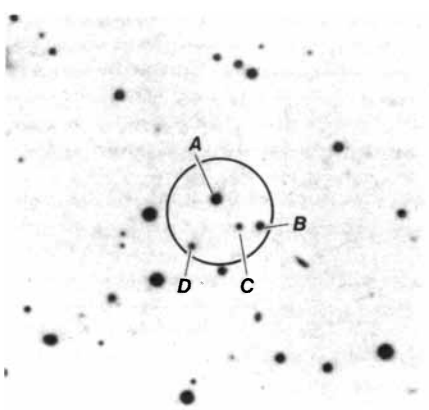
The final problem with the current data is one that is familiar to all statisticians and census takers: How normal is the small fraction of the population selected for the survey? The Einstein Observatory was commanded to point at quasars discovered over the past few years because they are extremely luminous in either their visible-light output or their radio output; these objects represent the large majority of the previously known quasars. The question is: Will such objects yield an unbiased survey of the X-ray properties of quasars?

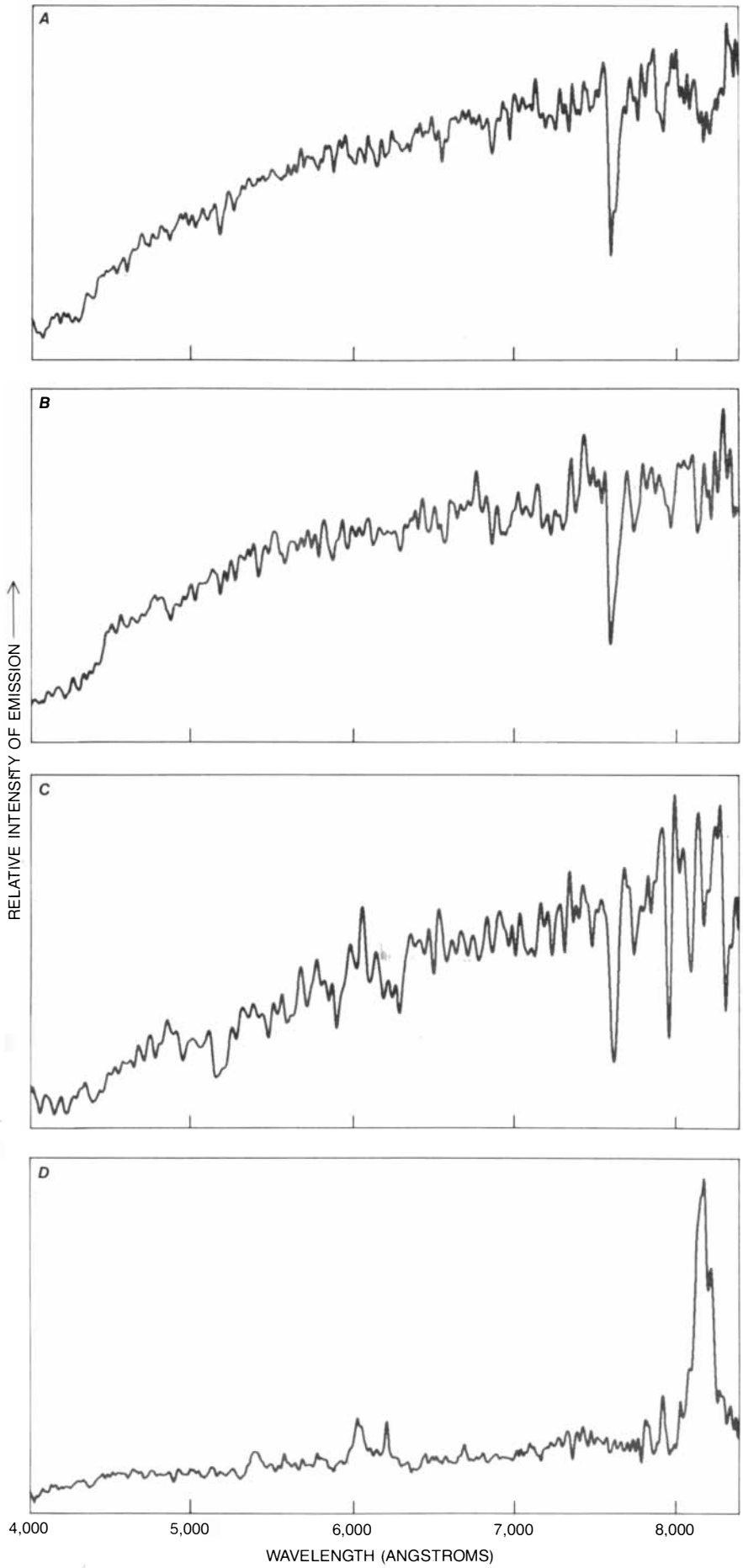
At the beginning of the Einstein program there was obviously no alternative to the practice of observing the quasars from existing catalogues, but the remarkable sensitivity of the Einstein X-ray detectors has rather unexpectedly provided an alternative. A typical observation consisted of orienting the satellite so that the astronomical object of interest was in the center of the focused X-ray image prior to the start of the exposure. As with any camera some surrounding area unrelated to the subject of interest was included in the image. An examination of the Einstein X-ray images reveals that a large fraction of these records show one or more additional faint X-ray sources at the edges of the image, unrelated to the original subject of the exposure. These newly dis-

covered X-ray sources have come to be called the serendipitous Einstein sources because of the accidental nature of their discovery.

The longer the X-ray exposure is, the more serendipitous sources tend to accumulate on a given Einstein image. The finding provides strong evidence that if point sources of any type are a significant contributor to the diffuse X-ray background, then the serendipitous sources are probably representative of this population. In other words, the serendipitous sources appear in the data unpredictably and without the intervention or advance knowledge of the observer; they are the pollster's dream of a truly unbiased sample.

What is the nature of the serendipitous sources? In particular, do they cor-





SERENDIPITOUS DISCOVERY of a previously uncatalogued X-ray-emitting quasar is depicted. In the course of a study of the X-ray emission from a binary stellar system in the constellation Aquarius a second X-ray source was noted nearby on an Einstein Observatory X-ray image. On the basis of the X-ray data the position of the new source could be established only approximately. On a visible-light photograph of the same region of the sky (*above*), made with the Palomar Schmidt telescope, a circle is drawn around the region where the X-ray source appeared to be. Four starlike objects are visible within the circle, all of which were treated as candidates for identification with the X-ray source. Spectra of each of these extremely faint objects were obtained by the author and his colleagues with the three-meter Shane reflector of the Lick Observatory (*curves at right*). The first three spectra have prominent dips where cool atoms in the outer layers of the object have apparently absorbed radiation from the interior. Such spectral absorption features are characteristic of normal stars, indicating that these three objects are probably not related to the X-ray source. The fourth spectrum has a completely different appearance: prominent peaks that correspond to excessive emission at certain wavelengths. This is the telltale signature of a quasar, in this case one that can be shown from its spectral features to be some 3.8 billion light-years away. Designated 2216–043 for its celestial coordinates, the newly discovered quasar is the optical counterpart of the serendipitous X-ray source.

respond to other objects already well studied, for example by optical methods, or are they some entirely different constituent of the universe? The question is not easy to answer because the limited focusing ability of X-ray optics results in images that are significantly blurred by the standards of optical astronomy. Just as it is difficult to measure precisely the characteristics of a blurred image on a visible-light photograph, so it is hard to assign an exact position in the sky to each of the serendipitous X-ray sources. Instead one establishes a small region, about a thirtieth of a degree across, from within which one can estimate that the X rays emanate. This precision is not sufficient to identify the object uniquely with any known visible source; a visible-light photograph of the same region of the sky where the X rays originate will typically show between four and six faint objects, any one of which (or none of which) could correspond to the X-ray source.

Further observations are clearly necessary to reveal the nature of the serendipitous X-ray sources and to determine if they have any bearing on the problem of the diffuse X-ray background. Several groups around the world have conducted such observations, including one consisting of Gary A. Chanan of Columbia University, Ronald A. Downes of the University of California at Los Angeles and me. Our technique has been to examine ordinary visible-light photographs of the small region of the sky that corresponds to the region of positional uncertainty of each of the serendipitous sources. As I have noted, this procedure generally yields a handful of faint objects that are candidates for the visible-light counterpart of the X-ray source. We then examine these visible objects in detail, searching for some anomalous property. (Bear in mind that normal stars are rarely detectable X-ray sources.) Because the candidates are typically 100,000 times too faint to be seen with the unaided eye, large telescopes are needed for the purpose. We worked primarily with the 2.1-meter reflector at the Kitt Peak National Observatory, the three-meter Shane telescope at the Lick Observatory and the four-meter reflector at the Cerro Tololo Inter-American Observatory in Chile.

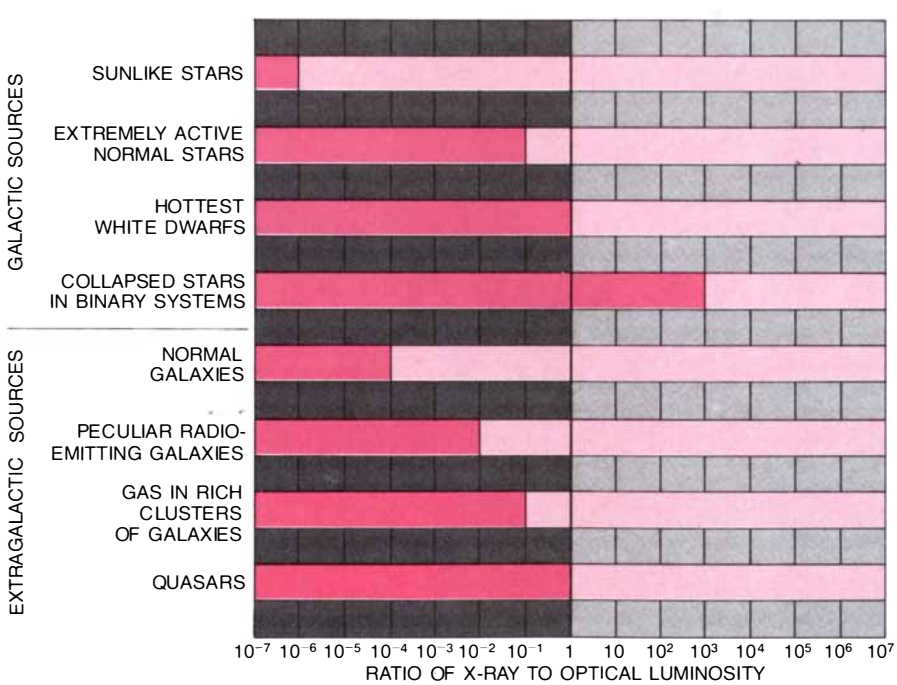
Our technique is based on spectroscopy, the division of radiation into its component wavelengths. The spectrograms of normal stars have gaps at characteristic wavelengths, where cool atoms in the upper layers of the star have absorbed light before it could reach the detector. Typically the spectra we obtained of almost every stellar object near the location of the serendipitous X-ray source showed this familiar pattern; apparently all these objects are normal, comparatively nearby stars, completely unrelated to the source of the X rays.

In most instances, however, we have found another object in the region that has a radically different spectrum, with a discrete set of wavelengths at which there is an excess of radiation. This pattern of emission features, precisely the opposite of the absorption features found in normal stars, indicates a highly unusual object, evidently one with copious amounts of hot gas. Furthermore, the precise wavelengths where the emission features appear correspond to those of familiar atoms such as hydrogen, helium and carbon, but with a large, systematic displacement in wavelength toward the red end of the spectrum. The red shift indicates that the source of the light is receding from the observer. This peculiar combination of spectral characteristics—strong, red-shifted emission features associated with common chemical elements—is the signature of a quasar. In short, what we, along with several other groups working on the problem, have found is that a large fraction of the serendipitous X-ray sources are actually previously undiscovered quasars.

The finding obviously has profound implications for the question of the origin of the diffuse X-ray background, and in particular for those interpretations in which the background radiation is attributed to an ensemble of distant, pointlike sources. If one peers through a fog and can just barely discern the nature of the brightest of a large number of distant objects that fade into a blur, it seems a reasonable assumption that similar objects constitute a significant fraction of the more distant, individually unobservable population. Similarly, the identification of a large number of the serendipitous X-ray sources with quasars (about 150 such cases are currently known) strongly hints that quasars are responsible for a significant fraction of the background radiation.

How large a fraction? Are quasars the predominant contributor to the background, thus solving the 20-year-old problem of the origin of this radiation, or are they merely a curiosity, a trace contaminant of a background due instead to a hot intergalactic plasma? Because there is no estimate independent of the intensity of the X-ray background of how much hot intergalactic medium exists (if any), there is a real possibility that both sources contribute to the observed X rays.

The properties of the ensemble of serendipitously discovered quasars are the key to the resolution of this issue. A study of the properties of the objects in



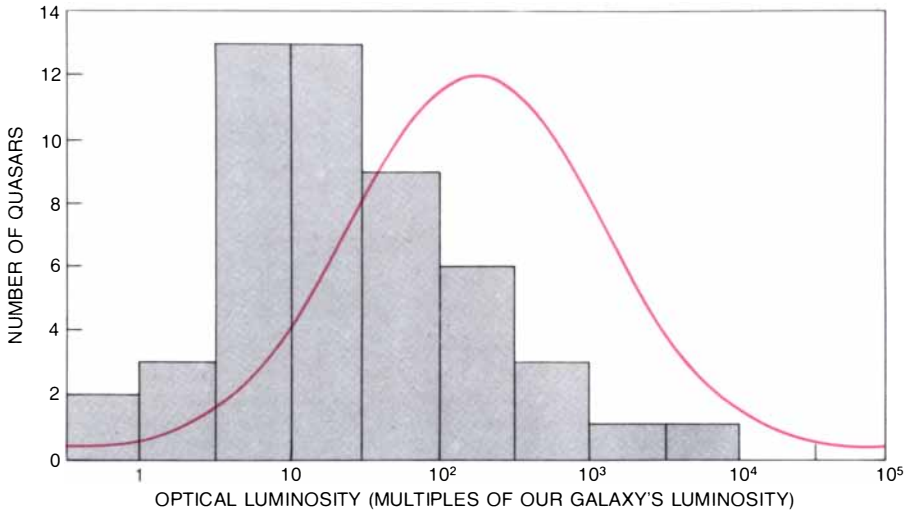
MOST ASTRONOMICAL OBJECTS generate only a small fraction of their total output of radiation at X-ray wavelengths. If the X-ray background is attributed to a large number of very distant individual sources, the objects must be both powerful X-ray emitters and fairly common in order to account for the observed isotropy. As these bars indicate, this condition is difficult to meet: the powerful X-ray emitters also tend to be rather exotic objects. Individual stars whose ratio of X rays to visible light is 1 or more are so rare that the sum of the X-ray emission from a normal galaxy is typically only a 10,000th (10^{-4}) of its visible-light output. Hence normal galaxies, although they are numerous, cannot contribute appreciably to the diffuse X-ray background. The hot plasma trapped in rich clusters of galaxies does generate an abundance of X rays, but such clusters are too rare to account for more than a few percent of the X-ray background. Quasars come closest to the required combination of properties: their X-ray output is prodigious, and they are particularly numerous at great distances from earth.

this sample, the first quasars discovered by their X-ray emission rather than their optical or radio emission, should provide a new and perhaps less biased insight into the X-ray properties of these enigmatic objects. Indeed, a comparison of the properties of the X-ray-selected sample with previous samples of optically selected or radio-selected quasars does reveal interesting differences. The intrinsic amount of light energy emitted by the X-ray-selected group seems to be systematically and significantly less, by a factor of about 10, than the corresponding quantity for the ensemble of previously known quasars.

A precise interpretation of the difference, however, is at the moment unclear. Some workers argue that the ensemble of all quasars has a broad range of possible X-ray and visible-light luminosities. Optical surveys would then find the most conspicuous visible-light objects with rather normal X-ray characteristics; the X-ray surveys would find exactly the opposite. Then a comparison of the two samples might well reveal significantly different properties, as are in fact observed. Alternatively, certain of the evolutionary effects cited above could preferentially suppress the X-ray emission of a subset of quasars, again creating different properties for X-ray-selected and optically selected samples. Worse yet, both effects may be important simultaneously.

The possible evolutionary effects are particularly troublesome. Although any isotropic radiation must be due to the sum of constituents at various distances from the observer, the dominant contribution can be shown to be made by the most distant objects. If the properties of the objects evolve and change progressively with distance, however, these most distant, extremely faint objects, for which the poorest data are available, can be very different from the closer, brighter objects. For the latter group the data may be good, but these data are not typical of the population of interest. This effect is almost certainly important at some level for the X-ray-emitting quasars. Even the extraordinary sensitivity of the Einstein detector was not sufficient to examine the objects that must be responsible for the bulk of the X-ray background, if indeed quasars are the main source of the background. Therefore extrapolations of the properties of the X-ray-selected quasars are needed to assess the total contribution of these objects to the diffuse radiation. At the moment the accuracy of the extrapolations is not known. Indeed, depending on the precise nature of the procedure, one can conclude that quasars are either a dominant component of the background or a negligible one.

Although the serendipitously discovered X-ray-emitting quasars provide new insights into the contribution of



FIFTY-ONE QUASARS discovered by the author and his colleagues as the optical counterparts of serendipitous X-ray sources appearing in the images recorded by the Einstein Observatory are plotted in this bar chart according to their visible-light output. Prior to this work virtually all the known quasars were discovered because of their unusual optical or radio emission rather than their X-ray emission. The colored curve indicates the optical-luminosity distribution for a representative sample of 51 optically selected and radio-selected quasars. The X-ray technique evidently selects quasars with an optical luminosity lower than that of the quasars selected by the earlier methods. The difference in luminosity between the X-ray-selected quasars and the previously discovered quasars is a clue to estimating the X-ray output of a typical quasar and hence to determining the contribution of quasars to the diffuse X-ray background.

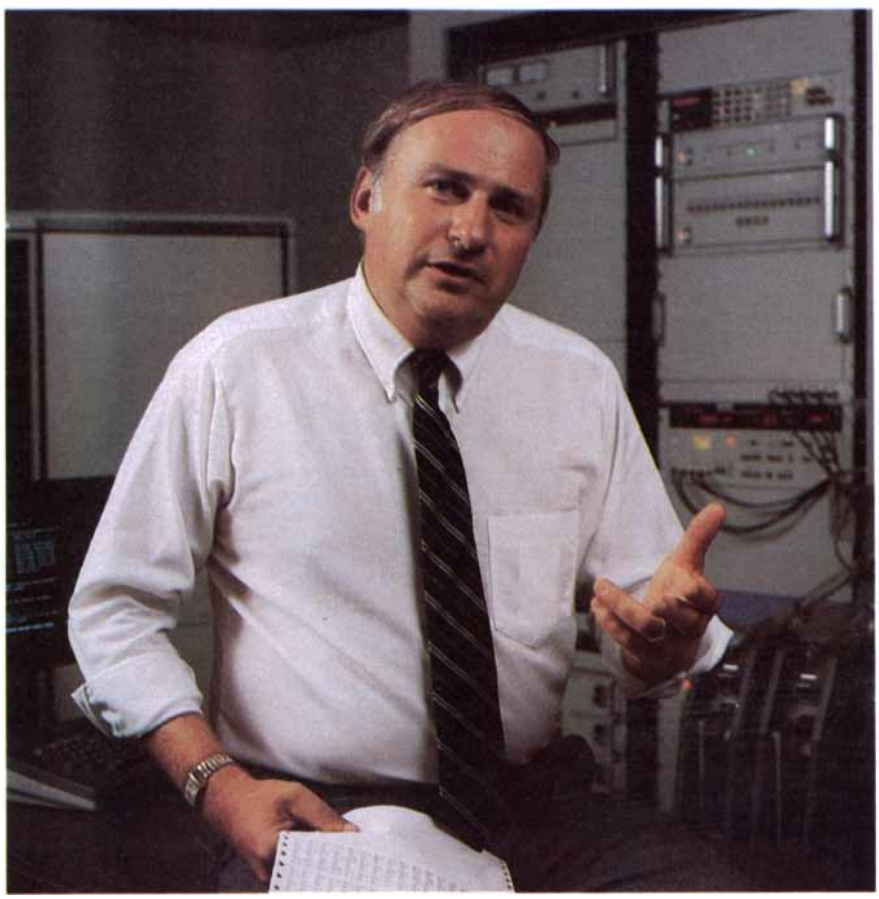
quasars to the background radiation, the exact interpretation of these insights is clearly complex and at the moment ambiguous. As often happens in observational astronomy, a large increase in sensitivity has brought not only new answers but also many new questions. Are there any independent clues to aid in separating the contribution of quasars to the X-ray background from that of the intergalactic plasma? One possibility lies in determining the spectrum of the observed radiation: How much X-radiation is detected at each wavelength compared with neighboring wavelengths? This technique is potentially a powerful diagnostic tool, because the different fundamental physical processes that generate X rays often result in different, recognizable spectra.

The experimental techniques for spectroscopy at X-ray wavelengths are more difficult than those at the wavelengths of visible light. Nevertheless, satellite-borne instruments have determined the spectrum of the X-ray background with increasing accuracy over the past few years. Much of the work in this area has been done by a group at the Goddard Space Flight Center of the National Aeronautics and Space Administration. Workers there have pointed out that the observed spectrum of the background radiation closely resembles the one expected from bremsstrahlung from a hot plasma at a temperature of about 500 million degrees K. This result can be taken as a powerful point in favor of the argument that a hot plasma dominates the background radiation, and that contributions from quasars (or any other type of distant discrete source) are quite small. As usual, however, there are complexities in the interpretation of the result that weaken the conclusion.

One such difficulty is in understanding what spectrum might be expected if an ensemble of quasars, rather than a hot gas, constitutes most of the background. In that case the spectrum would be the sum of the individual spectra of a large number of objects. X-ray spectra have been individually determined for only a few of the brightest quasars, and they do not particularly resemble the observed spectrum of the background. Perhaps the X-ray spectra of distant quasars, however, are different from the spectra of nearby quasars, again because of evolutionary effects. Furthermore, the spectrum of the diffuse background is most accurately determined at a range of X-ray wavelengths slightly different from the wavelength where the Einstein imaging observations were made. If nature were sufficiently mischievous to change the emission mechanism responsible for the bulk of the radiation from one regime or the other, workers determining the spectral properties of the radiation and workers concentrating on the imaging observations might in fact be unknowingly attempting to solve two slightly different problems.

There is one reasonably unambiguous measurement available to help distinguish the contribution to the X-ray background of distant, discrete X-ray sources from that of a hot plasma. As I have mentioned, the longer the exposure obtained by the Einstein Observatory,

What if you chose as a technical



"At M/A-COM, HP computers increase test speed 12-fold."

M/A-COM, Inc., is a major manufacturer of electronic products for commercial telecommunications and defense markets. Dr. Alan Carlson, Manager of the Automatic Test Equipment Group in Burlington, Mass., says: "We build automatic test systems and conduct final test on hundreds of M/A-COM products. To automate, we needed precision instruments plus computer power.

"HP was the only vendor to offer us a complete solution. With 20 HP 1000 computers driving more than 200 HP instruments, we've completely automated several product tests. And we can depend on HP's support teams to keep our entire testing network up and running.

"Now we can test sophisticated products 12 times faster than before. With these kinds of savings, each HP 1000 has paid for itself in less than one year."

What should you expect from a computer partner?

Quality—from researching customer needs to product development, manufacturing, marketing, after-sales service, and support!

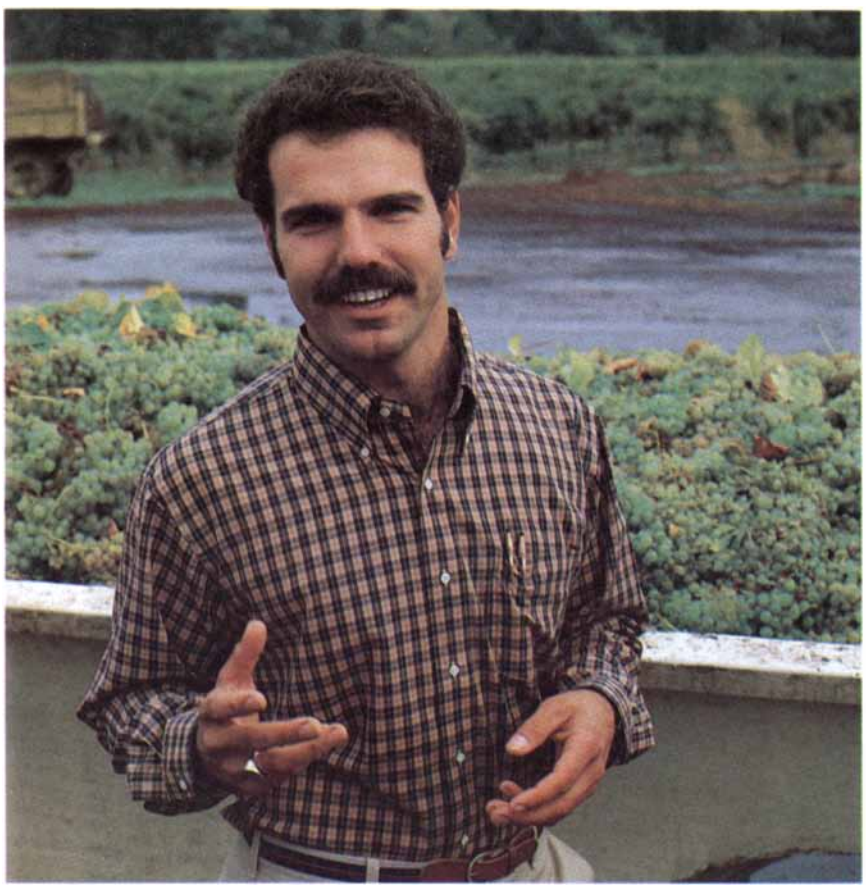
Quality, by HP's definition, encompasses more than the product.

This high quality recently was recognized by the Union of Japanese Scientists & Engineers, who awarded a coveted Deming Prize to Yokagawa-Hewlett-Packard, HP's joint venture company in Japan.

YHP offers the entire spectrum of HP products and in 1981 accounted for Japanese sales of more than $200 million. Nearly 60 percent of the products YHP sells are imported from HP manufacturing operations

The Deming Prize: *for outstanding merit in the control of quality.*

Hewlett-Packard computer partner?



"At Charles Krug Winery, an HP computer tests temperature 150 times faster than manual methods, improves quality control, and doubles production."

Charles Krug and CK Mondavi premium wines are produced at the Charles Krug Winery in St. Helena, California. The winery uses an HP 1000 computer system to monitor and control the temperature of its fermentation tanks.

Treasurer Peter Mondavi, Jr. explains: "The HP 1000 tests the temperature of each tank 150 times faster than manual methods. And because it automatically adjusts fluctuating temperatures we know each bottle of wine will meet our quality standards.

"Improved temperature control has enabled us to double our production while maximizing quality control. As a result, we've saved $50,000 in production costs this year alone."

in the US and abroad. Conversely, YHP manufactures and exports various products to HP customers throughout the world.

The total quality control approach for which YHP was awarded the Deming Prize is a point of pride throughout HP. Customers in every part of the world should expect *quality* as a fundamental element in a computer partnership with HP.

You should consider a working partnership with HP. Now.

For a free copy of our brochure explaining HP's practical, proven approach to meeting your long-term computer and information needs, write to A.P. Oliverio, Senior Vice President, Marketing, Dept. 207, Hewlett-Packard Company, P.O. Box 10301, Palo Alto, CA 94303.

When performance must be measured by results

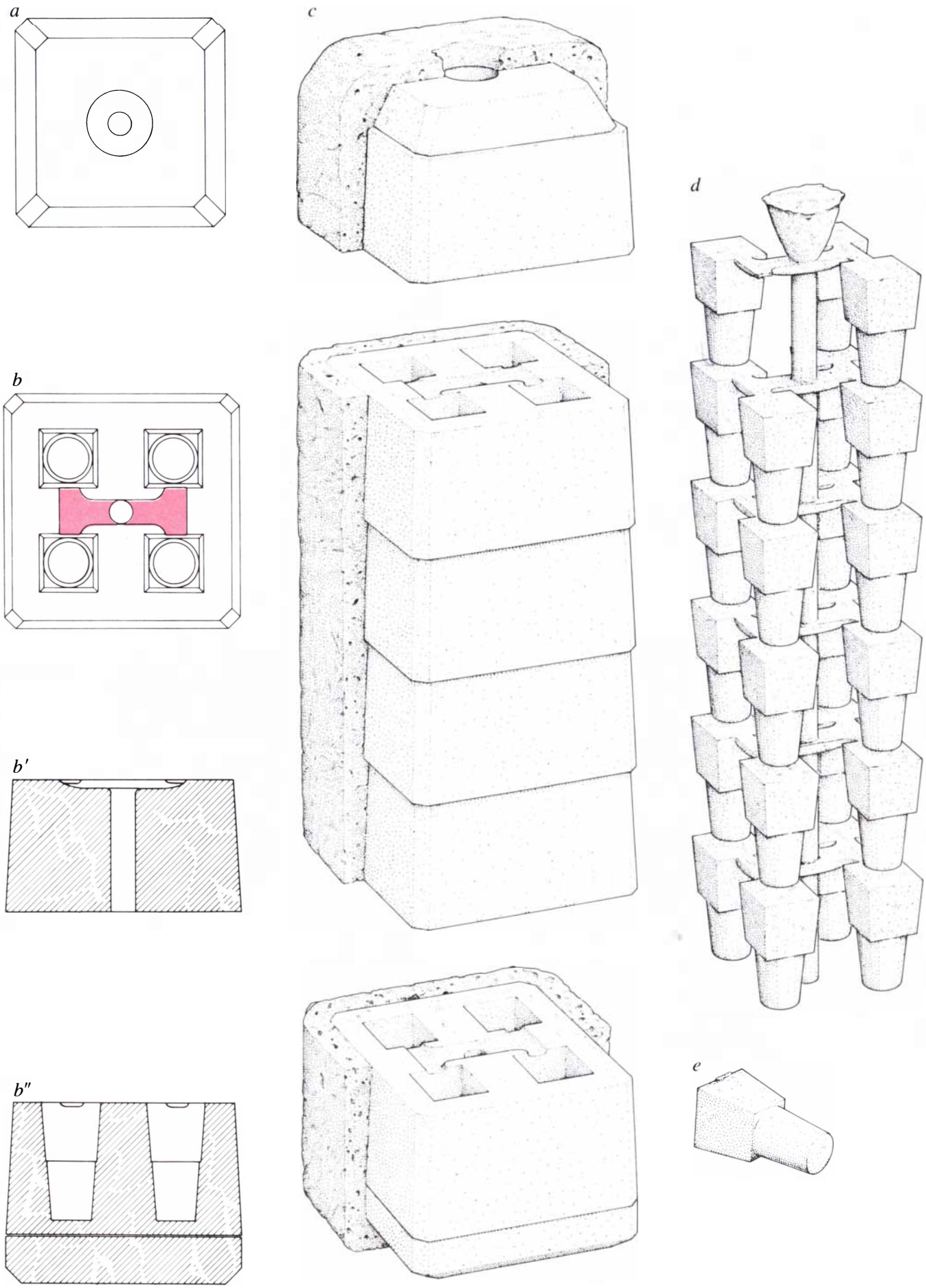
HEWLETT PACKARD

0002340

the greater the number of serendipitous X-ray sources detected. Therefore the fractional contribution of X rays represented by these discrete sources to the total number of X rays in an entire image is certainly a conservative lower estimate of the fraction of the background due to such discrete sources. On the deepest exposure this fraction is about a fourth. Clearly objects such as quasars do make a significant contribution to the background radiation. This measurement, however, is well short of the 50 percent mark, where one could say with confidence that discrete objects are responsible for the dominant contribution to the background radiation, thus essentially resolving the issue.

It should be clear from this discussion that a more powerful X-ray imaging instrument than the ones carried by the Einstein Observatory has the potential of settling the question in a direct and elegant manner, simply by individually resolving enough weak sources to account for half or more of the total emission in any small patch of sky, or alternatively discovering that at very low levels of X-ray brightness the radiation remains truly diffuse and does not break up into separate sources. The requisite technology is available, and a suitably powerful instrument has been proposed to NASA for flight in a satellite. The development of this project, known as the Advanced X-ray Astrophysics Facility (AXAF), would require approximately 10 years once approval for funding was obtained. A committee of astronomers headed by George B. Field of Harvard University recently prepared a list of important priorities for the next decade of astronomical research, and the list ranked AXAF first among its recommendations.

It is unclear at the moment whether further studies of the serendipitously discovered X-ray quasars may make it possible to understand with reasonable accuracy the possible effects of the evolution of quasars. Such an understanding, coupled with more sophisticated optical data on the number of quasars typically found in any small region of the sky, might then lead to a precise calculation of the importance of quasars to the background and by implication to an estimate of the contribution of the hot intergalactic plasma. The necessary higher-sensitivity optical data will almost surely be provided by the Space Telescope, which is now scheduled to be launched into orbit in 1985. Either at that point, or perhaps in the early 1990's with the launching of the AXAF satellite, it may finally be possible to know with confidence whether the diffuse X-ray background and quasars, both discovered independently 20 years ago, are actually two separate manifestations of one phenomenon.



The Mass Production of Iron Castings in Ancient China

By 500 B.C. the Chinese had developed stack casting: a high technology in which multiple castings are made by pouring iron into multiple molds designed to be stacked one on top of another

by Hua Jue-ming

The casting of bronze in China, going back at least to the Xia Dynasty (21st to 16th centuries B.C.), is one of the best-known traditions of ancient technology. Less familiar outside China is a development of the first millennium B.C.: the mass production of iron castings. Chinese foundrymen made coins, harness buckles, vehicle parts and other small articles in huge numbers by the technique of stack casting. The term refers to the fact that individual molds are stacked one on top of another and share a single "gate," or point of entry for the molten metal. The technique made it possible to fill as many as a dozen molds at the same time, thereby economizing on both metal and fuel. Stack casting began in China between 800 and 500 B.C. and is still practiced today.

Recent archaeological investigations have uncovered the remains of stack-casting foundries in four provinces: Henan, Shandong, Shaanxi and Jiangsu. Perhaps the richest site, found almost a decade ago by Lee Jing-hua and Tang Wen-xing of the Henan Provincial Museum, is a kiln west of the city of Wenxian in Henan. During the Eastern Han Dynasty (A.D. 24–220) the kiln was a key part of a major stack-casting foundry. Excavations at the site revealed that the kiln still held several hundred intact sets of molds, some of them designed to produce more than 80 castings at a time. Many of the ancient molds were so well preserved that the investigators have been able to use them to create objects identical with those first produced in cast iron nearly 2,000 years ago.

Before describing what has been learned at the Wenxian kiln site it will be useful to acquaint the reader with the basic technology of metal casting and the steps involved in it: first the making of the molds and then the melting, pouring and freezing (that is, the solidification) of the metal. Most of the work in casting technology has to do with the making of the molds. Throughout antiquity and into modern times the ideal mold material has been clay. It is abundant in all parts of the world, and little more than screening and washing to rid it of larger particles and foreign matter is needed to prepare it for the making of molds.

When dry clay is mixed with water, the material's exceedingly fine platelike particles become coated with a film that allows the plates to slide past one another on the application of very little force. At the same time the particles cannot easily be separated from one another because of the surface tension of the water film. Depending on the proportion of water to clay the wet mixture can vary from a stiff paste to a thin slurry. When drying removes the water, the clay reverts to a solid of modest strength. That strength can be greatly enhanced by firing (baking in a kiln); the heat initiates processes that bond between the clay particles. Like most other peoples of the world, the Chinese were familiar with these properties of clay as a result of making pottery for millenniums before the rise of metallurgy.

Clay pastes and slurries both play a role in mold making. A component of a mold can be made out of a paste and then formed into the desired shape by pressing into the soft surface of the clay a model made of wood or metal (a "pattern"). Conversely, a pattern having the desired shape can be repeatedly painted with a slurry and allowed to dry until the pattern can be lifted away from what is by then a thick clay shell that perfectly duplicates the pattern's detail. Either way the pattern can serve repeatedly for making a very large number of identical mold segments.

The mold segments must next be dried (to evaporate most of the water the clay still holds) and then fired. This not only increases the strength of the mold but also rids the clay of any remaining water and eliminates the hazard of steam bubbles' causing surface defects when the metal is poured or otherwise distorting the form of the casting. Firing the molds can make another contribution to casting. If, as was true at the Wenxian foundry, the stacked molds are taken from the oven while they are still quite hot and the molten cast iron is poured into the hot stack, the freezing process is retarded, allowing the fluid metal to fill all the mold interstices.

This retardation of freezing was a factor of considerable importance to the workmen at the Wenxian foundry because their stack molds had particularly narrow "runners": the connecting slots that conducted liquid iron from the central sprue, or vertical channel, to the various mold cavities. Premature freezing of the iron in these narrow slots

SIX-MOLD STACK for the production of 24 key wedges from a single pouring of high-carbon "white" cast iron appears in the illustration on the opposite page. At the left, shown in plan view only, is the "gate" (*a*) at the top of the stack, which provides a central pouring point for the molten metal. Below the gate, also in plan view, is a single mold with four casting cavities (*b*) and "runners" (*color*) leading to them from the central "sprue." The elevation view below it, (*b′*), running through the midline, shows the runners and the continuation of the sprue. The elevation view at the bottom (*b″*) runs through two of the four casting cavities. At the center (*c*) is the stack assembly shown with the top and bottom components exploded. The dark outer coating, half cut away, is a fired layer of straw and mud that made it possible to handle the six molds as a unit. At the right (*d*) is the casting as it appeared after the iron had solidified and the clay molds had been broken away. One wedge, runner removed, appears below the casting (*e*).

would have resulted in incomplete castings. At the same time the mold designers could not simply eliminate the problem by making the runners larger. The reason was that the metal they used was not the ordinary gray cast iron of the Western world but white cast iron, a form of the metal that is very much harder. Indeed, white cast iron is so hard that the labor of breaking thick runners off the finished castings and then grinding any rough areas smooth would have seriously cut the rate of production.

Bronze, the first metal cast on a large scale in China, consists mainly of copper. Iron ores are much commoner than copper ores, so that it is not surprising iron became the workhorse metal in China, as it did elsewhere in the world. In Europe and as far east in Asia as India the first iron implements were made by smelting and smithing. The product of the smelting process is a spongy, slag-filled mass of almost pure "bloom" iron. The bloom is converted into usable metal by hot hammering that squeezes most of the viscous slag out of it. Further heating in a forge and hammering on an anvil yields finished iron objects such as tools, weapons and ornaments.

In China iron production followed a different track. The Chinese, using more fuel and pumping more air into their smelters than the metalworkers of Europe and West Asia did, attained conditions that completely melted the iron ore, yielding a metal free of slag and rich in carbon. The carbon was important because the greater its concentration in the iron, the lower the metal's melting point. Pure iron does not melt below 1,535 degrees Celsius, a temperature that was not attained in furnaces on a practical basis until the 19th century. When iron contains 4.3 percent of carbon, however, it becomes completely liquid at 1,130 degrees C., a temperature only about 80 degrees higher than the melting point of bronze. Once this property of white cast iron was discovered in China the Chinese foundrymen, already masters of casting bronze, recognized they could do the same with the carbon-rich metal, thereby utilizing a much more abundant metallurgical resource.

By the middle of the first millennium B.C. cast iron was common in China. Weapons, ax and adz blades, hammerheads, hoe blades, spade tips and plow tips were being cast in iron; so were decorative items such as belt buckles and rings, household articles such as pots and braziers and equipage such as horse bits and the axle fittings of chariot wheels. A period of more than a millennium, from about 300 B.C. to perhaps A.D. 1100, saw a great growth of all technology in China. It was probably also a period of population growth. As one factor in this dual growth consider the plow. It harnesses human and animal power not only for routine cultivation but also for breaking new ground. For example, cast-iron plow tips broke the virgin prairie in the 19th-century march of agriculture across North America. In China as elsewhere wood plows preceded iron-tipped ones. It is not hard to imagine that with the availability of cast-iron plow tips toward the end of the first millennium B.C. there came a revolution in the increase of arable land, in food production and therefore in food for an expanding Chinese population. With the growth of agriculture and handicrafts came the need for transportation, and the demand for harness and vehicle parts was great. Such was the setting for the work at Wenxian.

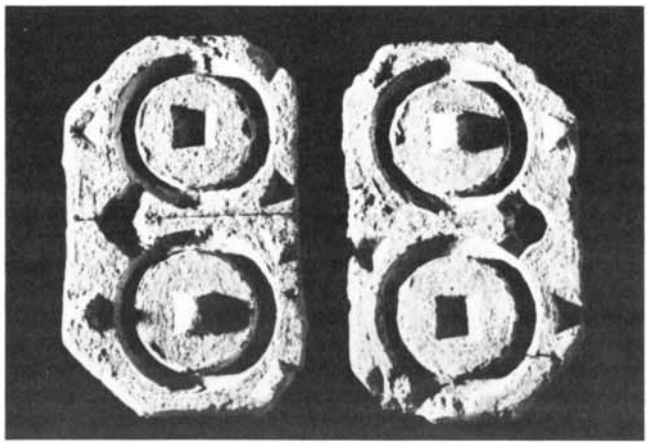
The Wenxian kiln was a part of a specialized foundry mainly producing iron castings for horse-drawn vehicles. The 500 sets of molds found at Wenxian could turn out 16 different kinds of castings in a range of 36 sizes. The castings included horse bits, harness rings, harness buckles and vehicle fittings such as bearings and axle end pieces. Among the molds for other kinds of castings were those for the weights of the steelyard, a type of balance found in many markets even today.

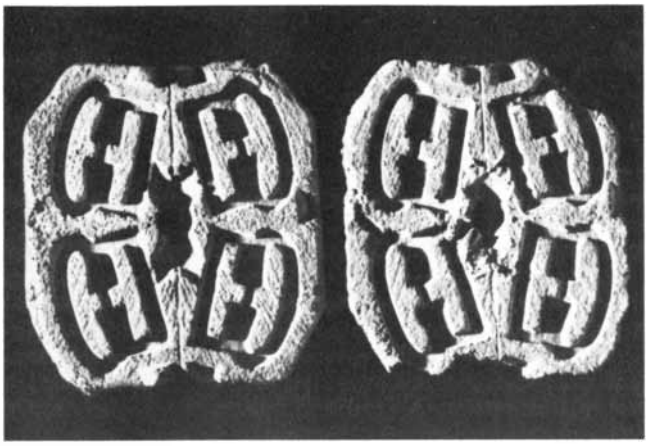
The uniformity of the molds suggests they were formed in master mold boxes made of metal, although no such masters were found at the site. One advantage of having such master patterns, in addition to the uniformity it ensures, is that any two halves of the mold made from the master can be mated to form a complete pair. This appears to have been particularly useful for the manufacture of buckles; some of the stack molds for these objects, with 14 levels and six buckle impressions per level, yielded 84 castings in a single pouring. An even greater advantage is that with many master patterns available many mold makers can work at the same time, which greatly increases productivity.

Analysis of vegetable debris at the kiln indicates that the metal masters were dusted with a fine bran before clay paste was packed into them and pounded down to acquire the negative impression. The bran facilitated the clean parting of the clay mold from the master; the molds were then left to dry for a week or so in a cellar or some other shaded place.

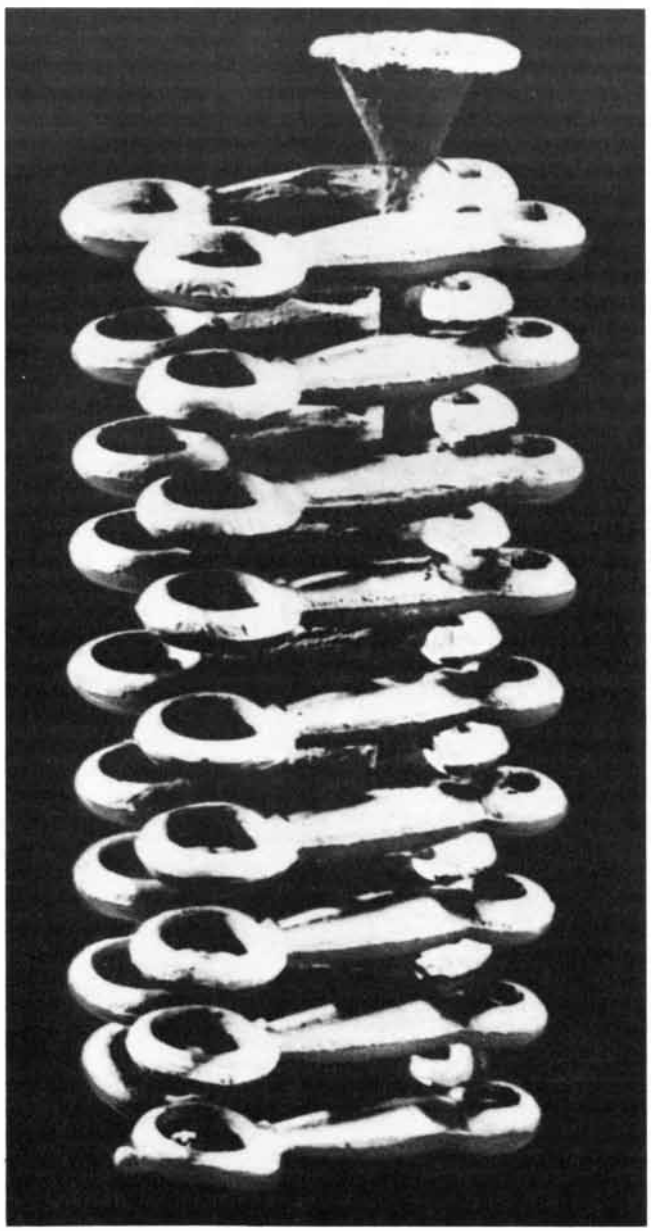
When the foundrymen had accumu-

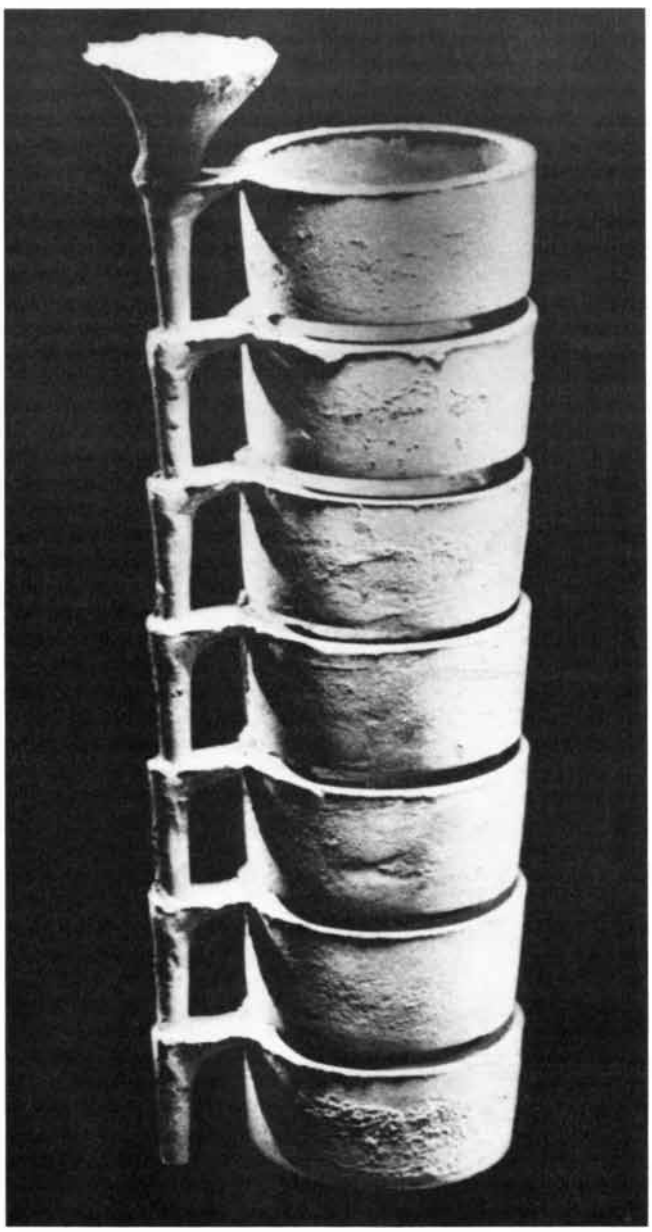
RECENT ARCHAEOLOGICAL FINDS in China include stack molds for castings (*open circles*) near Zibo in Shandong and near Nanjing and in Jurong county in Jiangsu. Molds and kilns (*dots*) have been found near Xi'an in Shaanxi and near Nanyang and Wenxian in Henan.





MATCHING HALVES of two different stack molds are shown in these photographs. The mold at the left produces two iron rings per unit. The mold at the right produces four iron buckles. Because both molds are shallow 10 or more could be stacked and filled together.

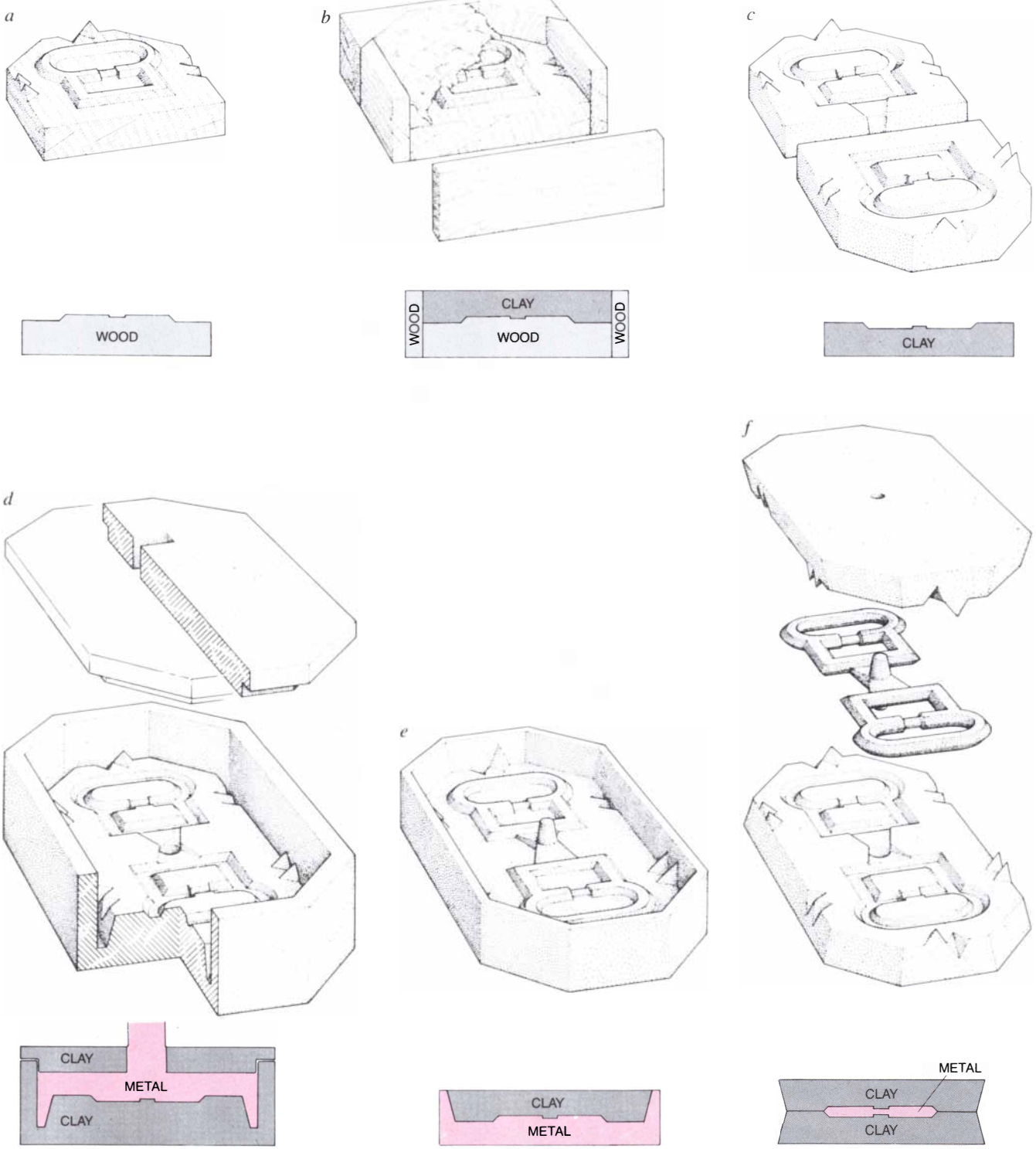




MOLDS MADE 2,000 YEARS AGO were stacked up, joined and filled by the excavators of the Wenxian kiln. The results were, at the left, 18 horse bits from a single pouring and, at the right, seven axle sleeves. Fewer of the deep axle sleeves could be poured at a time.

lated enough air-dried sets of molds to load into the kiln, they stacked the various kinds of molds together, taking care that the sprue hole in each successive mold was aligned with the hole in the one under it. Some molds were aligned by means of matching tongues and grooves; others had a hole for a guiding pin that extended through the stack. Once the molds were assembled the entire stack was coated with a layer of mud mixed with chopped straw (the standard raw material for making mud brick) and the stacks were again set aside to let the covering dry. This mud-brick covering made it possible to handle the stacked molds as a single unit. The molds were now ready for the kiln.

The Wenxian kiln proved on excavation to be three meters wide and 7.4 meters long. Along its front wall was a shal-



METAL MASTERS speeded the mass production of clay molds. The making of such a master begins at the left (*a*) with an exactly half-thick replica of a metal buckle (*top*) being carved out of wood. Next (*b*) the wood replica is reproduced in negative form by packing a box with clay. At *c* are two such negative clay reproductions. The two are then boxed together (*d*), forming the finished mold for the metal master, which is then cast. The finished metal master (*e*) is then used to make hundreds of clay molds that are paired (*f*) to cast the buckles.

Medicine is always changing, now there's a text that changes right along with it.

SCIENTIFIC AMERICAN *Medicine*.

The answers that you get from SCIENTIFIC AMERICAN *Medicine* are current, because each month new chapters replace the old. You will know where to look for the answers, because this advanced text follows the familiar structure of medical knowledge, changing as medicine changes.

Comprehensive. Leading authorities from Harvard and Stanford cover the full range of medicine for you the seasoned clinician.

Complimentary CME. At no extra cost, we offer a CME program in which you can earn 32 top credits* per year.

Trial Offer. We invite you to try SCIENTIFIC AMERICAN *Medicine* for two months with our compliments. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, you will owe us $220 for the first 12 months (current renewal is $170); otherwise, return the books; there is no charge.

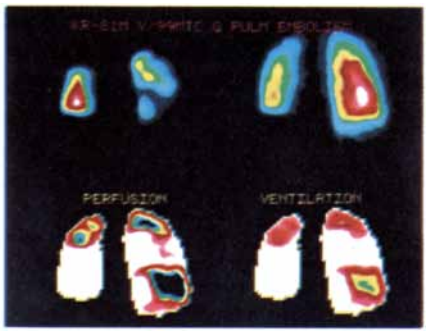
Please mail the coupon today so that we can put medical progress to work for you.

You may also order your trial copy by calling toll-free 1-800-345-8112 (in Pennsylvania, 1-800-662-2444).

*As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions.

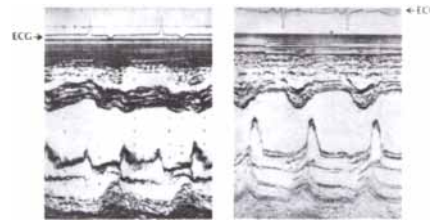
This program has been reviewed and is acceptable for 32 prescribed hours by the American Academy of Family Physicians.

This program has been approved by the American College of Emergency Physicians for 32 hours of ACEP Category 1 credit.

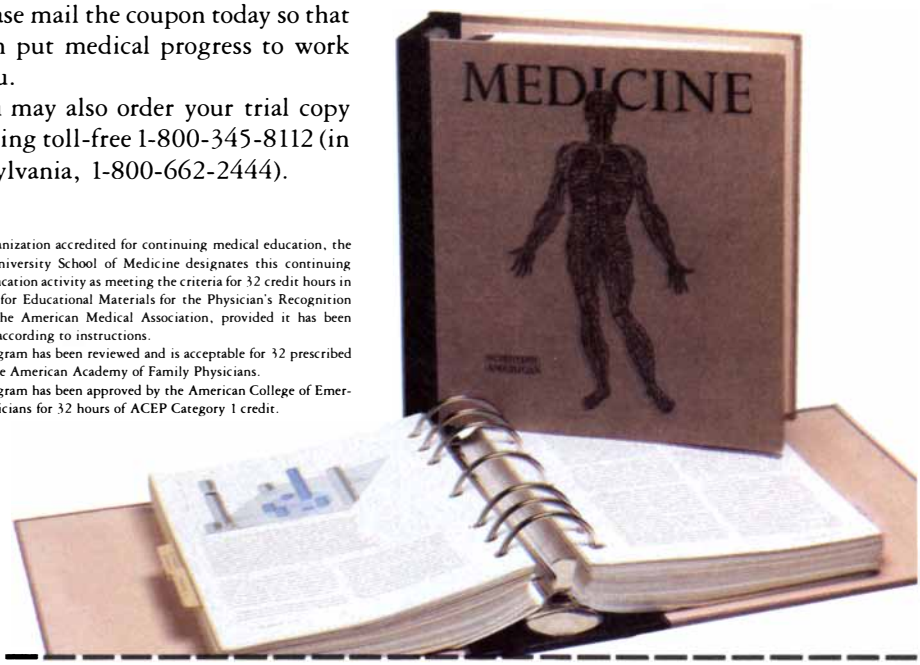


Computerized scintigraphy reveals pulmonary thromboembolism.

Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.



Echocardiograms from patients with aortic regurgitation.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, New York 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the annual subscription of $220 for SCIENTIFIC AMERICAN *Medicine* is tax deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund*.

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $220*
☐ Bill me ☐ VISA ☐ MasterCard
Expiration Date_____ Account Number_____
*Please add sales tax for California, Illinois, Massachusetts, Michigan and New York

Name_____ MD Specialty_____

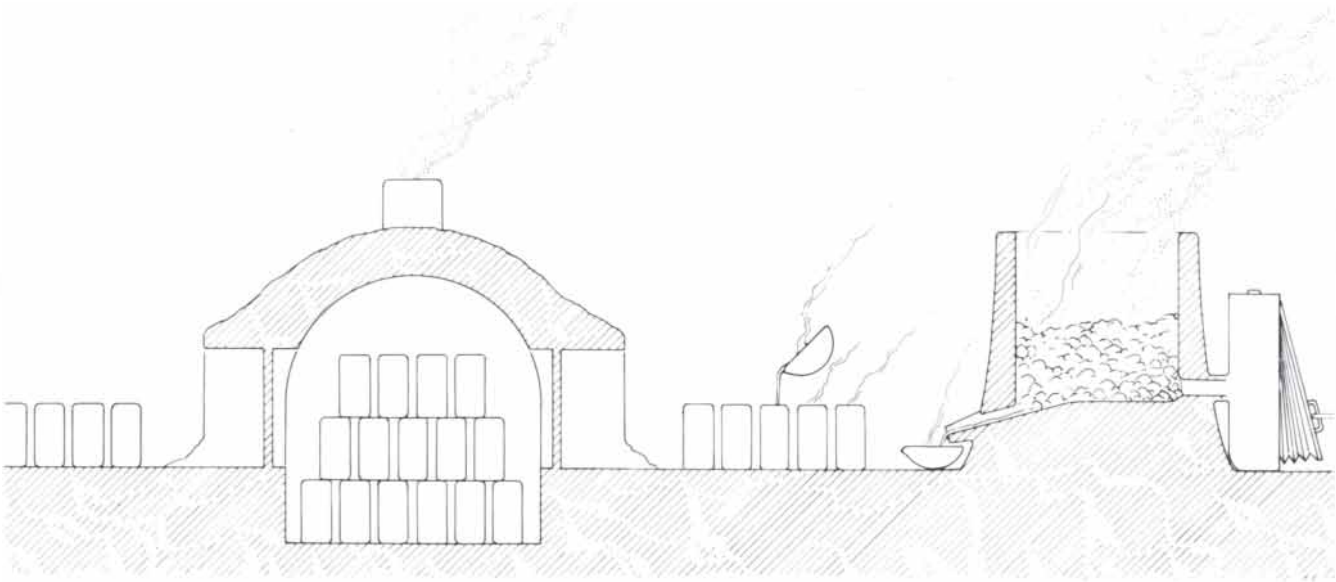
Address_____

City_____ State_____ Zip_____

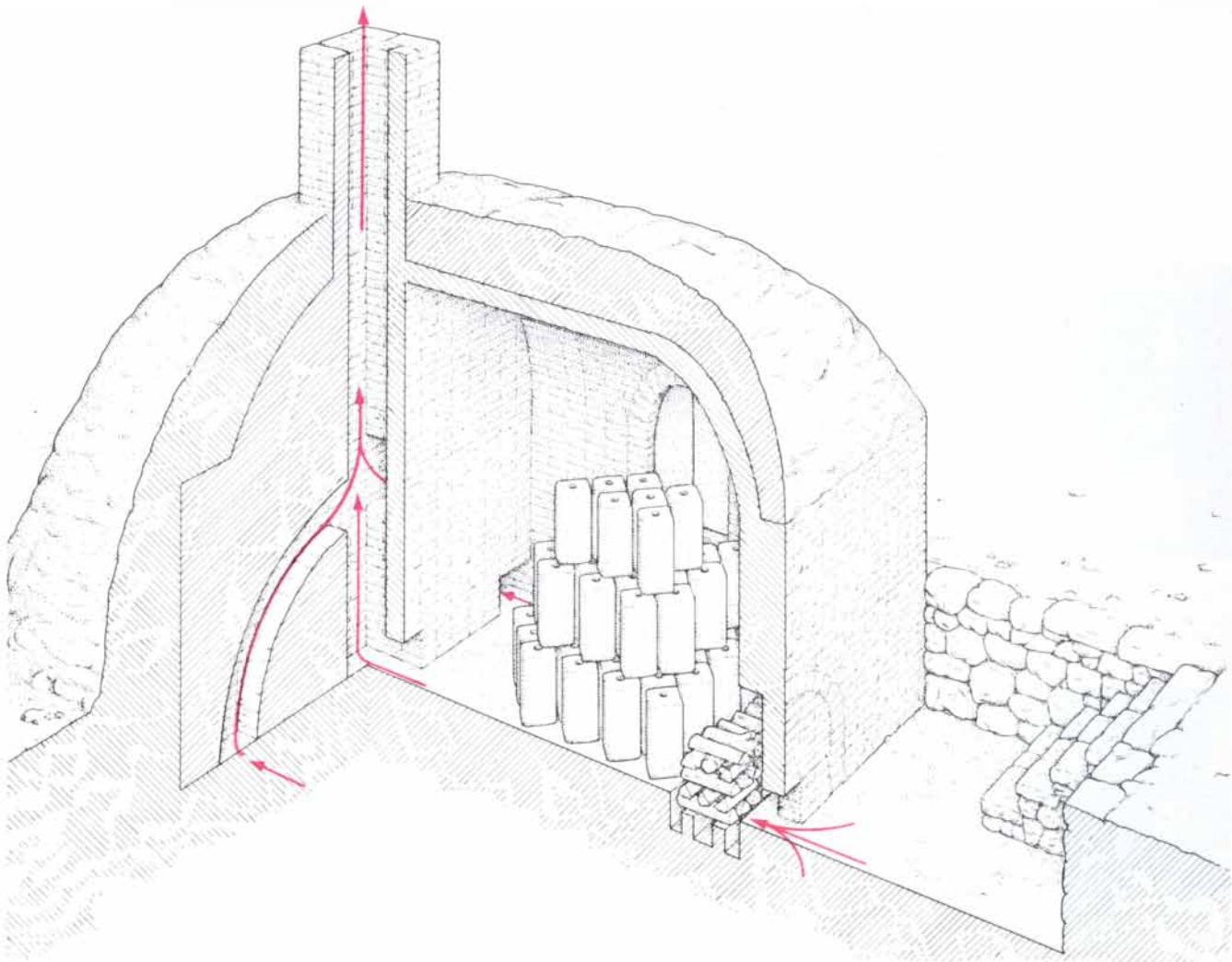
Signature_____

Please allow six to eight weeks for delivery. All payments must be in U.S. dollars. CME is available outside the U.S. and its possessions for an additional $65. Surface routes used unless airmail delivery is prepaid

W1



POSSIBLE PRODUCTION LINE at Wenxian could have combined the kiln for heating stacked molds, at the left, with an open smelter, at the right, where the white-iron pigs could be remelted, using a bellows for forced draft. No smelter, however, was found at Wenxian.



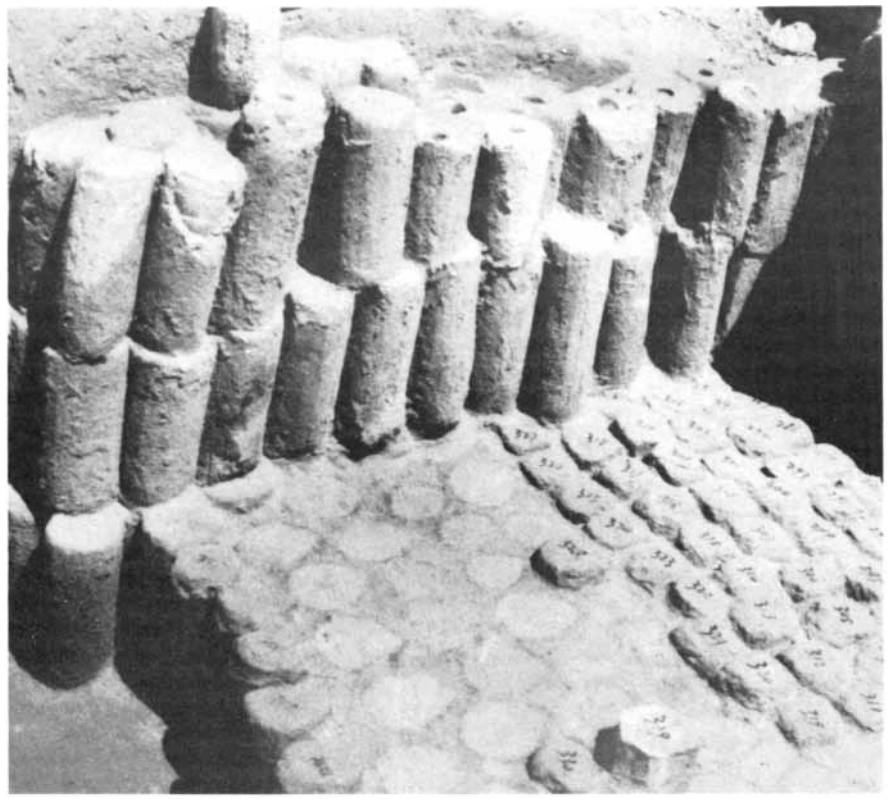
KILN LOAD OF STACKED MOLDS is seen ready for firing in this cutaway view of the Wenxian kiln as it was restored by its excavators. The colored arrows trace the passage of hot gases from the wood fire at the right through the stacks, ensuring even heating, and out through the chimney. The kiln was three meters wide and some seven meters long. Interior temperature evidently exceeded 700 degrees Celsius.

low trench where the fire was built. The kiln floor was lined with brick; three flues in the back wall provided an outlet for the hot gases. The mud-coated mold assemblies were piled one stack on top of another and carefully positioned on the kiln floor so that the gases from the fire would flow through the spaces between the molds before reaching the exhaust flues. This evenly distributed the heat in the kiln. To judge from the color of the fired clay the temperature must have exceeded 700 degrees C.

In Han Dynasty times and for many centuries afterward the foundry fuel was wood: kindling and logs for the kilns and charcoal for the smelters. The demand eventually denuded the countryside of its forests (as it also did in Europe). As a result coal mining had become a major Chinese activity by the end of the first millennium A.D. and thereafter coal supplied most of the heating needs of industry.

Since the Han foundrymen poured the molten cast iron into their stacked molds while the molds were still hot in order to facilitate the flow of the metal, it is quite possible that a smelter or smelters operated in tandem with the mold kiln at Wenxian. In that way crucibles of liquid iron would have been available for pouring when the freshly fired stacks were ready to leave the kiln. The investigators of the site have demonstrated, however, that although this would have been economical of fuel, it was also quite practical to reheat a cooled stack some time after it was first fired: they reheated stacks nearly 2,000 years later! In the process they also learned something about the procedures probably followed by the Han foundrymen.

This information was acquired when the investigators decided to exploit some of the intact molds in order to reenact the stack-casting process. Sets of molds were cleaned and then assembled, smeared with straw mud and allowed to dry. For the first experimental casting sets of molds were kept at a temperature of 300 degrees C. for some five hours before the molten metal was poured. After cooling, the castings were found to be low in quality; the defects included large voids and excessive shrinkage. Evidently the molds had not been adequately heated. For the second experiment other sets of molds were heated to 600 degrees C. and kept at that temperature for a full six hours. By then the stacks glowed red. They were left in the oven and allowed to cool to 300 degrees C. before they were removed. Then the metal (either copper or iron) was poured into them. All the castings except one set were of high quality. The exception was a stack of key wedges. These wedges were quite thick and their molds lacked risers. Their top surface was dimpled, apparently as a result of shrinkage in the mold.



THREE TIERS OF STACKED MOLDS were still standing when the excavators uncovered the site near Wenxian. A total of 500 molds were found; 16 items could be cast in 36 sizes.

Stack casting as a mass-production technique arose in China before the middle of the first millennium B.C. The sophisticated molds found at the Wenxian site therefore represent a culmination of more than five centuries of technological progress in iron metallurgy. For example, since all the molds at the foundry were intended to be stacked for casting, their "parting" (and assembly) design was horizontal. This put a natural limit on how many objects higher than they were wide, such as the key wedges, could be cast at one pouring. Nevertheless, by stringently controlling the height of each individual mold the Wenxian foundrymen, working with a stack six molds high and with four cavities in each mold, were able to cast 24 key wedges at a time. For objects such as buckles that were longer and wider than they were high, of course, a stack could have many more individual molds and each mold could have multiple cavities. One such buckle mold, although it was not uncovered intact, carried the negative impressions of six buckles and was only a tenth as high as it was long. A stack of 20 such molds would have yielded 120 buckles per pouring.

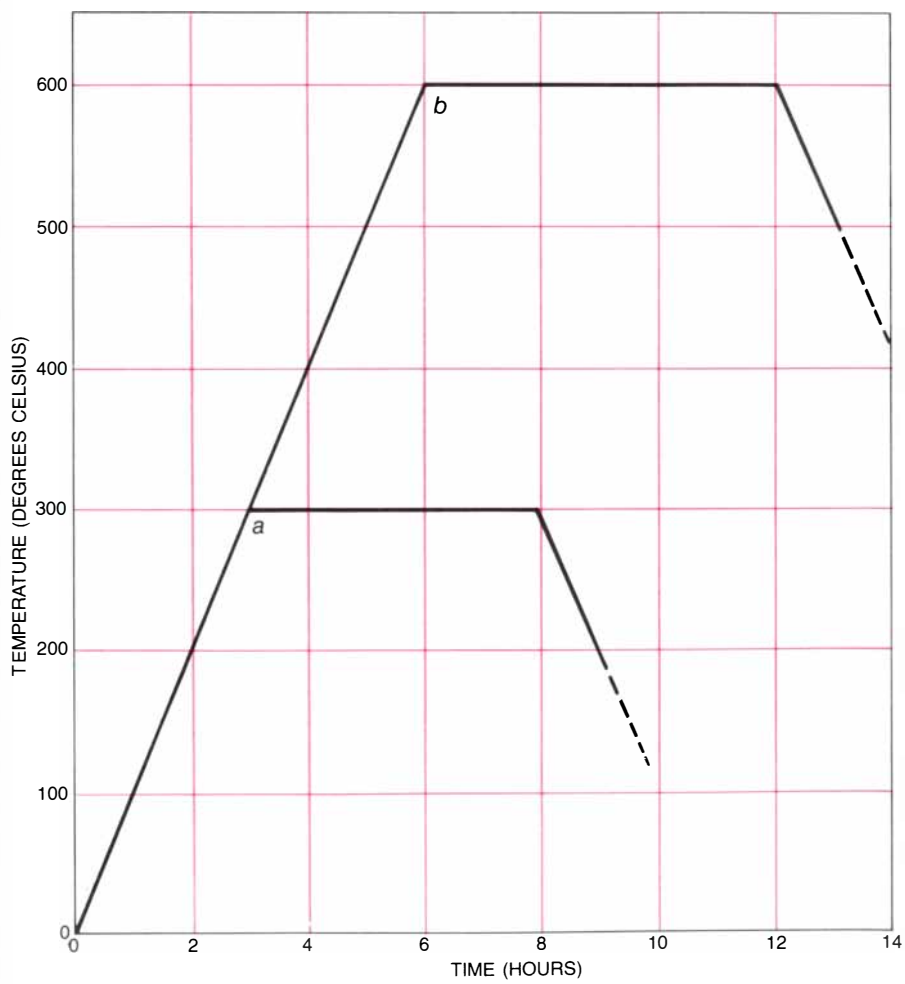
Among the molds at Wenxian there is one of rather complex geometry that could have been intended for casting a metal master pattern. The master in turn would have yielded clay molds for casting two buckles. Making such masters was time-consuming but nonetheless efficient; it is estimated that a single master might yield as many as 10,000 clay molds before it showed signs of serious wear.

Elsewhere in China at this time simple metal molds were serving not to make clay molds but to make castings directly. Their use was confined to the casting of utilitarian objects with uncomplicated shapes: plow tips, ax and adz blades and the like. When these simple iron molds were clamped together, they formed a mold cavity complete with a pouring gate. Such "permanent" molds were thick-walled and sturdy, and their thermal conductivity was much higher than that of clay. As a result the mold could be assembled, the liquid iron could be poured, the iron could solidify, the hot casting could be removed and the mold could be reassembled for another pouring—all in a few minutes. Indeed, both the productivity and the unit cost of casting with metal molds were probably better than those of stack casting. Metal-mold casting, however, was limited to relatively simple forms. Although no such molds were found at Wenxian, the fact that castings were being made with both metal molds and clay molds in China some two millenni-

ums ago shows that more than one good idea was being put into practice in the interests of increasing the output of the Han foundries.

What accounts for this early flowering of metallurgical ingenuity and sophistication in China? The findings of archaeologists consistently indicate that the craftsmen of the past were both intelligent and resourceful. Under the appropriate conditions any foundryman anywhere and at any time could have invented stack casting. What, then, were the conditions in China? First, the basic technologies were in place: in China both pottery manufacture and bronze casting were already ancient arts when cast-iron metallurgy began. Second, the social circumstances provided a strong stimulus for innovation: in China the technological and economic advantages of mass-producing tools, vehicle fittings and ornaments of cheap and plentiful cast iron must have been self-evident in the first millennium B.C. or there would have been no demand for them. In the light of these conditions the introduction of stack casting, although it is interesting enough in itself, perhaps more significantly suggests it would be rewarding to examine the social circumstances that nurtured such an innovation.

Stack casting not only became an important element of foundry practice early in the development of Chinese metallurgy but has continued as part of foundry practice to this day. For example, in the modern foundry at Foshan in Guangdong Province stack casting is used to produce gears and other precision metal parts notable for their accurate dimensions and fine finish. The ancient technologies uncovered at Wenxian and elsewhere not only help the Chinese of today to appreciate their heritage but also should encourage them to advance further. As the historian of technology A. B. Wilder has written: "We study the past because it is a guide to the present and a promise for the future. As we look backward our attention is directed forward."



REENACTMENT OF HAN CASTING was attempted at first (*a*) by heating the stacked molds to 300 degrees C. and heat-soaking them for somewhat less than six hours. The castings showed many defects. On a second attempt (*b*) the temperature was raised to 600 degrees and the molds were heat-soaked for a full six hours. When they had cooled to 300 degrees, the metal was poured. The quality of the castings in the second batch, with one exception, was excellent.

我们与创造明天的人接触

KE XUE is the most recently launched of SCIENTIFIC.AMERICAN's six translated editions. KE XUE's paid circulation of 25,000 includes the People's Republic of China's emerging scientific, engineering and industrial leadership: the people who will make China's future happen. KE XUE is the only American originated publication on the newsstands in the PRC.

China's ability and inclination to buy will increase. They plan to accelerate progress in virtually all industrial categories: energy, aerospace, data/processing, communications, manufacturing, transportation, heavy machinery, agriculture, chemicals and mining. Now is the time to present your corporate identity, your product, your service to the People's Republic of China.

We speak the language of more than 3,000,000 world shakers in 140 countries each month. We reach the people who make the future happen in Chinese, Japanese, French, Spanish, Italian, German and English.

For more information contact:

John Kirby
Vice President
Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue
New York, New York 10017
Telephone (212) 754-0262

SCIENTIFIC AMERICAN
54 Pall Mall
London SW 1Y 5JH, England
Telephone (011-441) 839-2873

SCIENTIFIC AMERICAN

We reach the people who make the future happen.

THE AMATEUR SCIENTIST

Simple and vivid demonstrations of advanced concepts in physics

by Jearl Walker

One might suppose that demonstrations of advanced concepts in physics would be beyond the reach of the amateur experimenter. Three such experiments have been devised, however, by Richard E. Crandall of Reed College. One experiment demonstrates the Doppler shift of light, converting the phenomenon into sound. A second measures Planck's constant, which figures importantly in quantum mechanics. The third measures the universal gravitational constant, which does the same in Newton's theory of gravitation. In each experiment Crandall has minimized the difficulties.

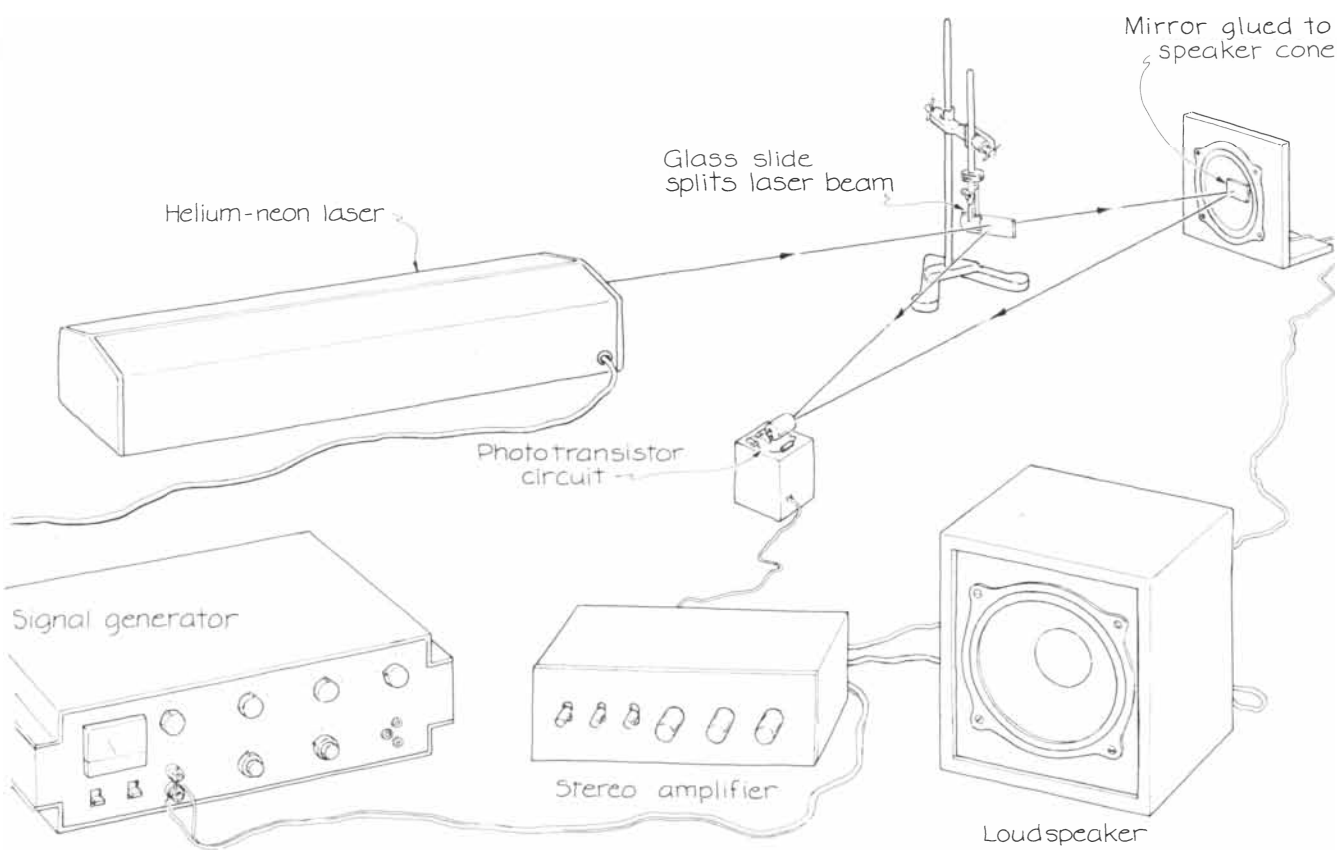
Interference patterns of light are standard fare in physics classrooms and are familiar to many amateurs. Normally the pattern is seen directly by the experimenter or is scanned with a light detector. Crandall and his student Edward H. Wishnow devised a method in which the pattern can be heard.

A beam from a two-milliwatt helium-neon laser is sent through a glass slide. Part of the beam reflects from the slide and illuminates a phototransistor (Motorola MRD3052 NPN or an equivalent type) that serves as a light detector. The rest of the beam passes through the slide and then reflects from a mirror (or anything shiny) mounted on the cone of a small loudspeaker. Part of the reflected light illuminates the phototransistor and interferes with the light arriving directly from the glass slide.

The transistor is biased (given a reference level) with a 15-volt source in line with a 30,000-ohm resistor. When the transistor is illuminated by light from the slide, it generates a voltage that is detected by a standard stereo amplifier. The amplified signal goes to a second loudspeaker that makes it audible. The other output channel of the amplifier is connected to the speaker that reflects part of the laser beam.

When the mirror moves, the signal at the transistor varies and so does the output of sound from the sound-producing speaker. Even small disturbances of the air in the room provide sufficient motion for a noticeable variation in the sound. More control of the variation is achieved by feeding a sine-wave signal into the second of the input channels of the amplifier. This signal is sent to the reflecting speaker, forcing it to oscillate steadily. The sound output from the apparatus then varies continuously.

The interference of the light at the transistor can be explained in two ways. In one view the critical factor is the difference between the distance the light travels after it is reflected from the slide and the distance it travels after it is re-



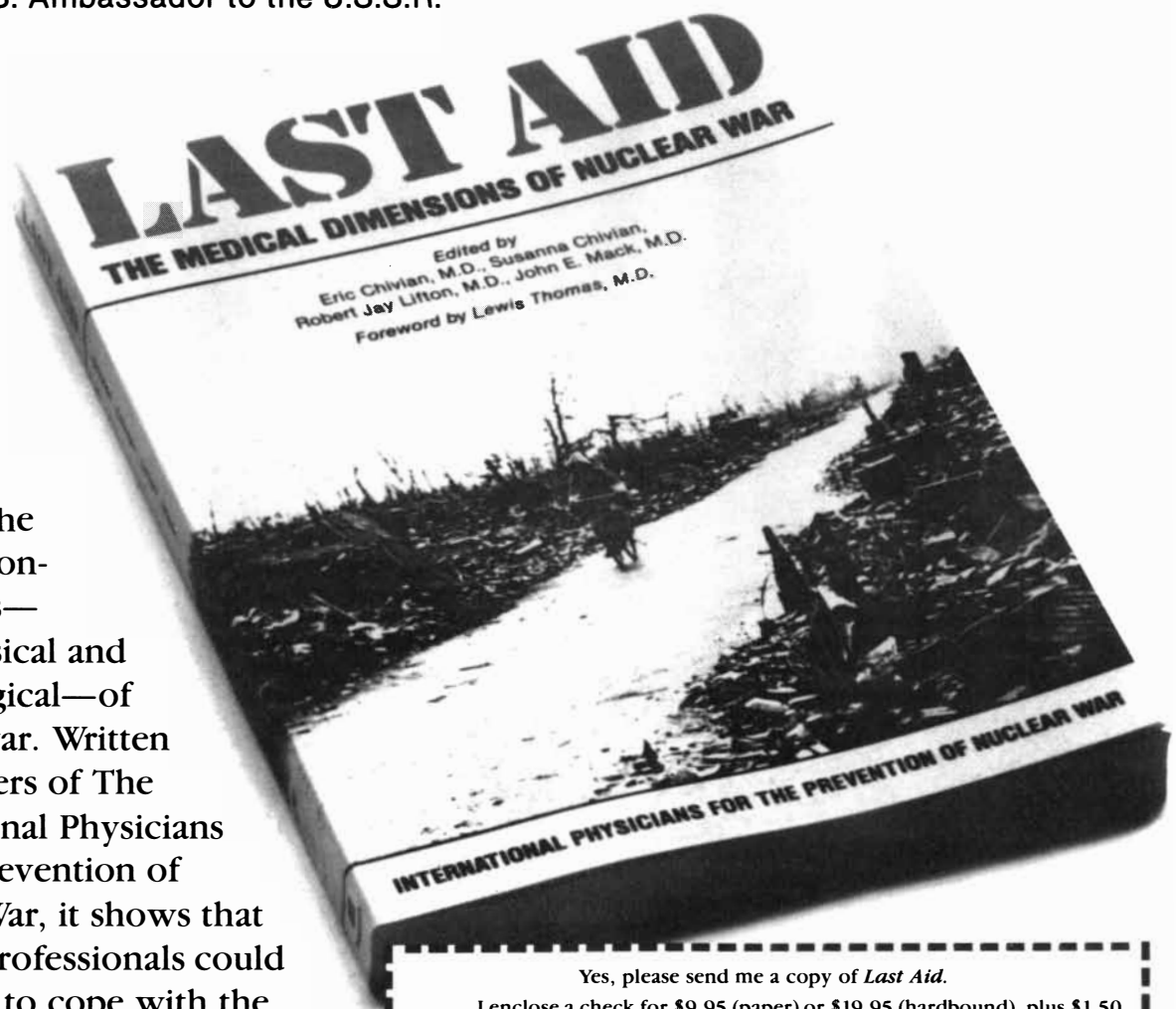
Richard E. Crandall's setup for converting the Doppler shift of light into sound

"To ignore this passionate presentation is to turn our backs on the most urgent issue of our time."

Walter Cronkite

"The facts revealed and documented in this book are fundamental for any constructive thinking about the problems of nuclear weaponry, and deserve the widest possible attention."

George F. Kennan, Ph.D.
Former U.S. Ambassador to the U.S.S.R.



Last Aid presents the medical consequences—both physical and psychological—of nuclear war. Written by members of The International Physicians for the Prevention of Nuclear War, it shows that medical professionals could not begin to cope with the overwhelming problems that would ensue in the aftermath of a nuclear holocaust. A reaction to heads of government who speak rationally of a conflict that can have no rational basis, ***Last Aid* is must reading for every intelligent, thinking human being.**

Yes, please send me a copy of *Last Aid.*

_____ I enclose a check for $9.95 (paper) or $19.95 (hardbound), plus $1.50 to cover postage and handling.
(California residents add appropriate sales tax.)

_____ Please charge my ☐ MasterCard ☐ VISA

Account # _____ Expires_____

Signature _____
(All credit card orders must be signed.)

Name _____

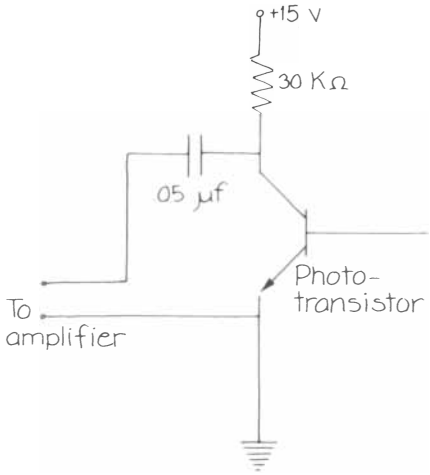
Address _____

City _____ State _____ Zip _____

W. H. Freeman and Company
660 Market Street
San Francisco, CA 94104

ISBN hardbound 0-7167-1434-5
ISBN paper 0-7167-1435-3

SA



The detection circuit for the experiment

flected from the mirror. When the light waves emerge from the laser, they are all approximately in phase. Some of them reflect from the slide and travel to the transistor. The rest of the waves travel a different path to reach the mirror and the transistor.

The difference between the length of the path traveled by one set of waves and the length of the path traveled by the other determines the interference at the transistor. If the path lengths are such that the two sets of waves arrive in phase, the interference is constructive: the illumination of the transistor is bright and the transistor puts out a strong signal to the amplifier. If the path lengths are such that the two sets arrive out of phase, the interference is destructive: the illumination of the transistor is dark and a weaker signal is sent to the amplifier.

Suppose the first light waves arrive in phase. Also suppose the mirror moves away from the laser by a distance equal to one-fourth of the wavelength of the laser light. The light reflecting from the speaker then has an additional distance to travel that is equal to half of the wavelength of the light. (It must travel through an extra one-fourth wavelength to reach the speaker and then back again to reach the transistor.) The additional distance puts the light waves at the transistor out of phase. Thus as the cone of the reflecting speaker moves parallel to the light beam it varies the strength of the signal from the transistor.

One can hear the random motion of the mirror as it is disturbed by slight movements of air in the room. When a sine-wave signal drives it in a more controlled oscillation, the variation in the interference at the transistor generates a repetitive pattern. The mirror moves through much more than one-fourth of the wavelength of the light, so that during each oscillation the interference at the transistor reaches many strong and weak levels.

A different but related explanation of the experiment is based on the Doppler shift of light. Wavelength is related to frequency: the number of waves that pass per second. The frequency of the light perceived by an observer can depend partly on his motion with respect to the source of the light. Assume the observer is stationary with respect to the light source. He then perceives light at a certain frequency. If the source moves toward the observer, the perceived light is shifted to a higher frequency. This shift is often called a blue shift because it is toward a higher frequency and so toward the blue end of the spectrum of visible light. If the source moves away from the observer, the perceived light is shifted to a lower frequency. This shift is often called a red shift. Both blue and red shifts are Doppler shifts.

A Doppler shift in frequency also results, of course, if the observer moves toward or away from the source. The degree of the shift depends on the relative speed of the source and the observer. The closer that speed is to the speed of light, the larger the Doppler shift is. This fact is crucial in inferring the motion of a distant astronomical object. The object's Doppler-shifted light can be measured and its speed toward or away from the earth can be calculated.

In the experiment done by Crandall and Wishnow the mirror moves relative to the transistor at a speed that is low compared with the speed of light. The resulting shift in the frequency of the light is quite small. The light is actually shifted twice in frequency, once because the mirror moves relative to the laser and again because it moves relative to the transistor. In the first instance the light source is the laser and the mirror is the observer. In the second the mirror acts as a source (since it is sending out light) and the transistor is the observer.

Normally the special theory of relativity is required to calculate the Doppler shift of light. The speed of the oscillatory motions of the loudspeaker is so low compared with the speed of light, however, that an approximation can be made on the basis of classical physics. The shift in frequency is equal to twice the unshifted frequency multiplied by the ratio of the speed of the mirror to the speed of light.

The light arriving at the transistor from the glass slide is unshifted. The light from the moving mirror is shifted up or down in frequency depending on which way the mirror is moving when the light reaches it. These two signals interfere with each other at the transistor to create what is called a beat frequency. The net illumination at the transistor varies between constructive and destructive interference with a beat frequency that is equal to the difference between the unshifted frequency of the light and the shifted frequency. The experimenter hears this beating as a variation in the sound emerging from the sound-producing speaker. In other words, he hears a pattern that is related to the Doppler shift of light.

Stefan's law ——— $P = \epsilon A \sigma T^4$

Stefan's constant ——— $\sigma = \dfrac{2\pi^5 k^4}{15 h^3 c^2}$

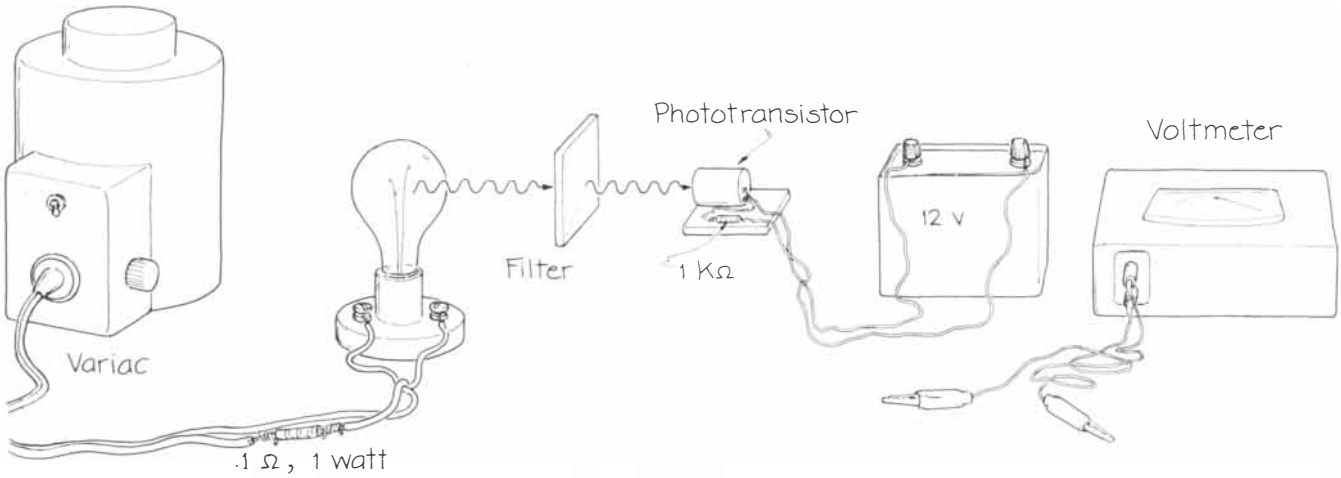
Planck's equation ——— $I = \dfrac{Nf^3}{\exp(hf/kT) - 1}$

Approximate equation for Planck's constant ——— $h \simeq \dfrac{15c^2}{2\pi^5 A f^4} \cdot \dfrac{P_1 P_2}{(P_1^{1/4} - P_2^{1/4})^4} \cdot \mathrm{Log}^4 \left(\dfrac{I_1}{I_2}\right)$

Equations for determining Planck's constant

At the end of the 19th century the subject of thermal radiation presented a serious problem in physics. Every body radiates electromagnetic waves according to the temperature of its surface. Some of the radiation might be in the visible range of the electromagnetic spectrum, so that it can be seen, but every surface also emits radiation over a broader range of the spectrum. For example, a poker heated in a fireplace gives off infrared radiation that can be felt long before the surface is hot enough to glow in the visible range.

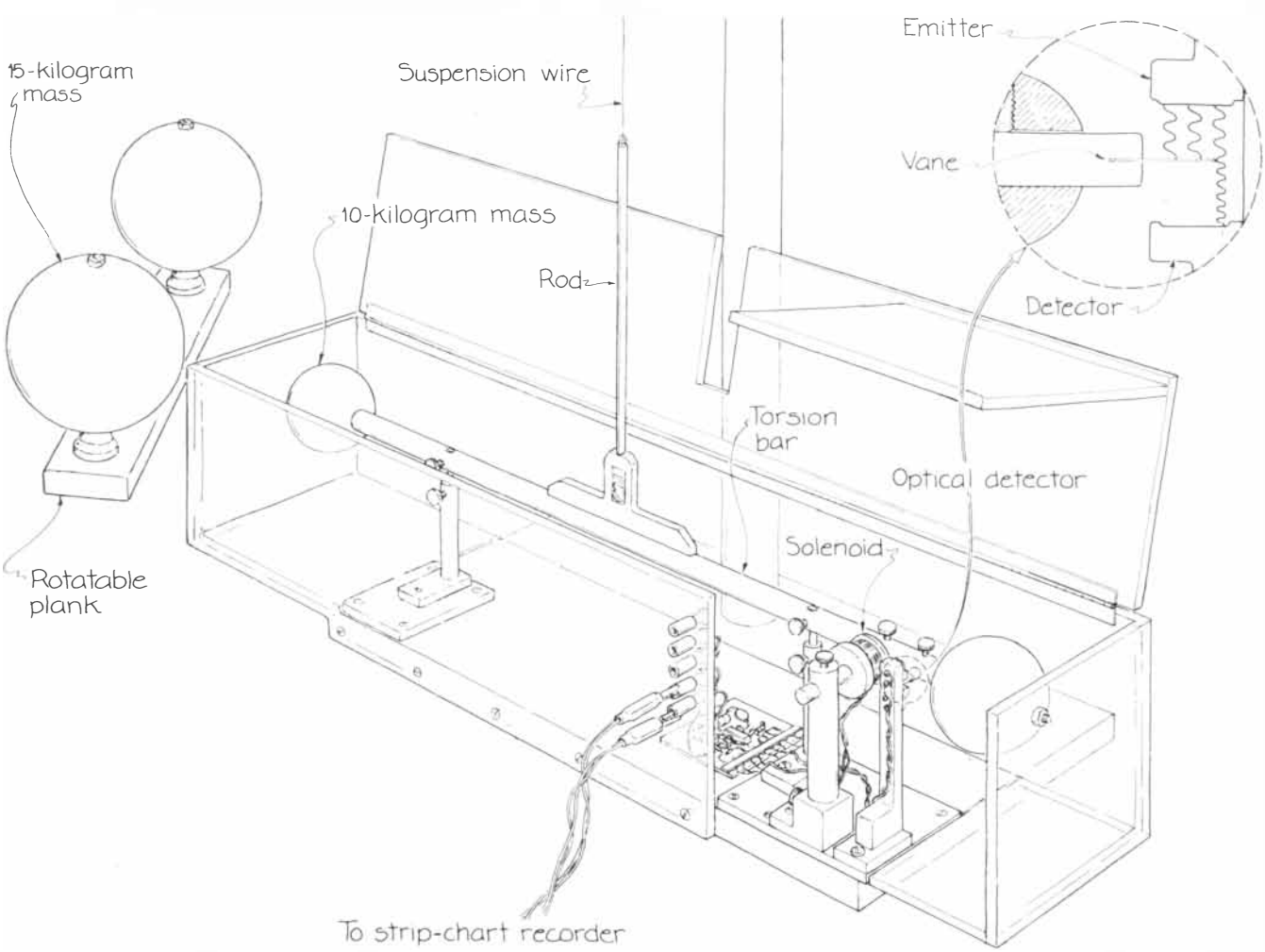
Armed with classical thermodynamics and the fresh brilliance of James Clerk Maxwell's theory of electromagnetic radiation, physicists attempted to derive an expression for the intensity of the thermal radiation at a particular frequency. Repeated efforts yielded only approximations. At low frequencies a mathematical expression derived by Lord Rayleigh approximated the ex-



How Crandall measures Planck's constant

perimental results, but at high frequencies the expression failed. Indeed, it predicted that the power radiated at high frequencies would be infinitely large, which was clearly wrong. This mathematical disaster was known as the ultraviolet catastrophe. Earlier Wilhelm Wien had devised an expression that worked well at high frequencies but was obviously wrong at low frequencies.

The problem was resolved in 1900 when Max Planck presented an equation that closely agreed with the experimental evidence at all frequencies. The equation was almost a product of luck, because it was not until several weeks later that Planck understood why it worked. In his theoretical model of a radiating surface he imagined that the radiation was emitted from many small oscillators. The intensity of the radiation at a given frequency was simply the combined output from a distribution of these oscillators. The surprising feature



Crandall's arrangement for measuring the universal constant of gravity

of the model was that the oscillators could not have just any energy on a scale of energies. Instead each oscillator was limited to particular energy values that were integral multiples of a fundamental value given by the product of a constant (now called Planck's constant h) and the frequency f of the oscillator.

Planck had no information about the atomic nature of a radiating surface, and his imaginary oscillators certainly did not exist. With his crucial discovery, however, he was able to justify his equation for the thermal radiation as a function of frequency. It was more important that Planck's work ushered in the age of quantum mechanics. On the microscopic scale energy values were limited to certain distinct quantities that were integral multiples of hf.

Although Planck's constant is an important value in the most advanced quantum mechanics of atoms and fundamental particles, it can be measured in a surprisingly simple experiment devised by Crandall and his colleague Jean F. Delord. The apparatus consists of a color filter, a light detector and an unfrosted 60-watt tungsten-filament light bulb. Light from the filament of the bulb is filtered so that only a narrow range of frequencies reaches the detector. The intensity of the light in that range is measured for two levels of power delivered to the light bulb, and the results are inserted in an equation that yields a value for Planck's constant.

The filament of the light bulb is heated by the passage of current and emits light across a broad band of frequencies in the visible and infrared ranges. The total power radiated from a surface was computed by Josef Stefan in 1879 as being equal to the product of four factors: (1) the surface area, (2) a constant known as Stefan's constant, (3) the absolute temperature of the surface raised to the fourth power and (4) a value known as the emissivity of the material.

At given values of surface area and temperature a material can theoretically radiate a certain maximum amount of power. Such a material is assigned an emissivity of 1 and is called a black-body radiator. No such ideal material exists; all real surfaces radiate less power. Each surface is assigned an emissivity smaller than 1 according to how well it radiates. The filament in the light bulb employed in the experiment runs hot enough to be almost a black-body radiator. Hence Crandall and Delord have made the approximation of assigning it an emissivity of 1.

Planck's constant appears in the modern expression of Stefan's constant, so that it is related to the total power radiated by the filament. It also appears in the equation devised by Planck for the intensity of the light radiated at any given frequency. Crandall and Delord perform two sets of measurements in order to get a value for Planck's constant. In each set they find the total power radiated by the filament by measuring the electric power delivered to the bulb. They also measure the intensity of the light passing through the color filter. This set of measurements is then repeated for a different amount of power delivered to the bulb. The data are inserted into an equation that is the combination of Stefan's law (the total power radiated) and the equation from Planck (the intensity of the light at a certain frequency).

The combined equation is shown in the bottom illustration on page 132. The equation incorporates an approximation in order to simplify the mathematics, but the accuracy of the final value of Planck's constant is hardly altered by it. The symbols I and P represent the intensity and power Crandall and Delord measure in the experiment. The equation also requires a value for the speed of light c, which is 3×10^8 meters per second. The symbol f is the frequency of the light passed by the filter. Since the filter actually passes not one frequency but a narrow range of frequencies, the value for this factor is chosen as being the center of the range. If the range were narrower, less error would be introduced by that approximation.

The symbol A in the equation is the surface area of the filament. To measure it Crandall and Delord project a shadow image of the filament onto a wall. The dimensions of the image are measured and then scaled down by the magnification of the projection. You can determine the magnification easily by placing a transparent ruler near the bulb and measuring its image on the wall. The value for A in their experiment was 5.2×10^{-5} square meter.

A Variac controls the power delivered to the bulb. To calculate the power Crandall and Delord measure the voltage across the .1-ohm resistor that is in line with the bulb. Dividing the voltage by the resistance gives the current flowing through the resistor and thus also through the filament.

Next Crandall and Delord measure the voltage across the bulb. Multiplying this voltage by the current gives them the power delivered to the bulb. Crandall says that if a Variac is not available, a 12-volt battery can be substituted, but one would then have to adjust the current with an appropriate resistor in order to deliver two different levels of power to the bulb.

The filter placed between the bulb and the photodetector is a green cellophane one bought from Corion Instruments, Inc. (73 Jeffrey Avenue, Holliston, Mass. 01746). The frequency entered in the calculation of Planck's constant was 5.3×10^{14} hertz. Similar filters can be obtained from other scientific-supply companies. Different colors can be substituted. The band pass, that is, the frequency range, of the filter should be as narrow as possible, but even an inexpensive filter will serve for determining a good value for Planck's constant.

The detection circuit consists of a phototransistor (Motorola MRD3052 or its equivalent), a 1,000-ohm resistor and a 12-volt battery. The spectral response of the phototransistor is not important because the frequency of the light illuminating the transistor does not vary. The current in the detection circuit is determined by measuring the voltage drop across the resistor. The current need not be converted into units of light intensity. Since a ratio of only two intensities is required in the equation, the units cancel.

Crandall sent me the following example of the data. With a power input of 19.3 watts to the bulb the current in the detection circuit was .81 milliampere. When the power input was changed to 36.5 watts, the current changed to 7.57 milliamperes. From these data Crandall and Delord calculated Planck's constant to be 4.9×10^{-34} joule-second, which is close to the accepted value of 6.6×10^{-34} joule-second. Their value is off somewhat because of the finite band pass of the filter and because the emissivity of the filament is not exactly 1.

Of the several fundamental forces in the universe the most noticeable in everyday life is gravitation. The gravitational pull on one's body from the presence of the earth is appreciable only because of the large mass of the earth. The gravitational attraction between any pair of common objects around the house is so small that it goes unnoticed. It is overwhelmed by much larger forces such as friction.

How does one demonstrate and measure the gravitational force between two masses in a laboratory? This question was first tackled by Henry Cavendish in 1798. His apparatus consisted of a six-foot wood arm suspended horizontally by a thin wire. At each end of the arm he mounted a lead ball about two inches in diameter. The suspension was extremely sensitive to any force tending to rotate the arm about the suspension wire.

To present such a force a third lead ball was brought near one of the mounted balls. Its gravitational pull on the mounted ball exerted a force tending to rotate the arm about the suspension wire. To increase the force a fourth lead ball was brought on the other side of the ball at the other end of the arm.

Cavendish was careful to shield the apparatus from air currents. He also arranged for the third and fourth balls to be moved into position by remote control so that the mass of his own body

would not interfere with the experiment. The angle of rotation was measured by monitoring a light beam reflected from a small mirror on the wire from which the beam was suspended.

With the third and fourth lead balls in position the arm rotated slightly until the gravitational pull was countered by the torque from the twisted wire. This position of equilibrium was measured. Then the third and fourth balls were repositioned so that they exerted forces in the opposite direction. Again the angle of equilibrium was measured. The angular difference between the two positions of equilibrium was divided by two to give an accurate measure of the rotation in each case.

With these numbers Cavendish was able to compute the strength of the gravitational force between the pairs of lead balls at each end of the arm. From work by Isaac Newton the mathematical form of the gravitational force between two masses was already known. Its strength depends directly on the two masses attracting each other and inversely on the square of the distance between them. It also depends on the factor called the gravitational constant. Newton was effectively stymied on how to compute the constant because his only demonstration of gravitation was based on at least one large object whose mass was not well known (such as the earth). The Cavendish experiment was exciting because it was based on smaller masses that could be measured easily. Cavendish computed the gravitational constant as being $6.754 \times 10^{-11} N(m^2/\text{kg.}^2)$, close to the currently accepted value of $(6.668 \pm .005) \times 10^{-11} N(m^2/\text{kg.}^2)$. ($N$ is newtons, m is meters, kg. is kilograms.)

The Cavendish experiment is not an ordinary demonstration in physics because it is arduous to do. The arm takes a good deal of time to settle down into its equilibrium position. Any perturbation from the environment (for example from the experimenters' moving around the laboratory) adds further to the settling time. As a result the demonstration often takes hours. Working from a prototype built by Greg Eibel of Reed College, Crandall has devised a clever electronic version of the Cavendish experiment that takes only minutes and yields a value for the gravitational constant of $(7.5 \pm 1.5) \times 10^{-11} N(m^2/\text{kg.}^2)$.

Whereas the classical Cavendish experiment depends on the rotation of an arm, the experiment devised by Crandall depends on preventing such a rotation. The strength of the gravitational force is measured from the magnetic force that is required to prevent the rotation of a torsion bar when an additional mass is brought close to it. The position of the torsion bar is monitored by an optical sensor.

When the additional mass is brought near one of the masses mounted on the bar, the start of rotation by the bar triggers a servomechanism that exerts enough magnetic force on the bar to stop the rotation. By calibrating the servomechanism Crandall gets a measure of the gravitational pull on the mass mounted on the bar. The entire experiment can take as little as 20 minutes.

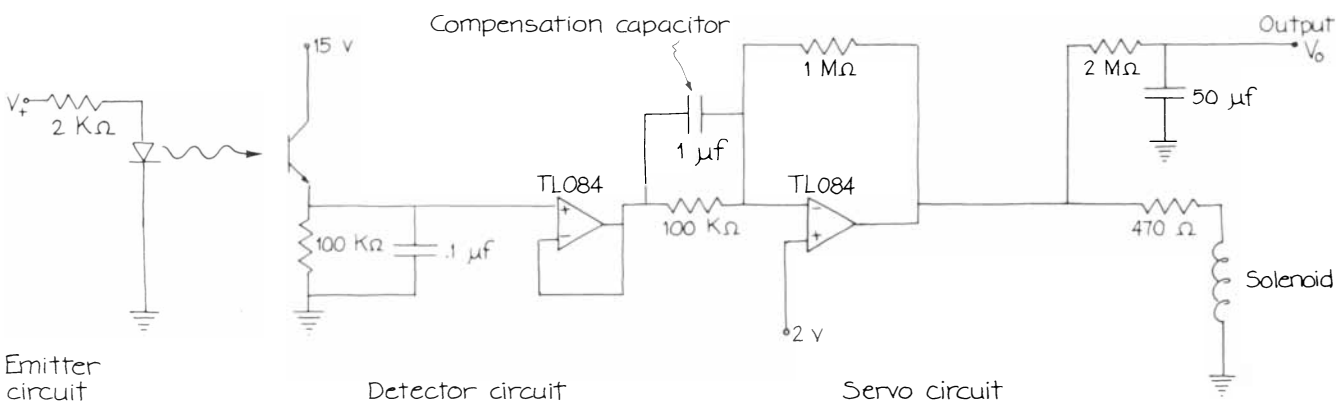
The bar is suspended by a length of piano wire that runs from an overhead mount down to a rigid rod attached to the middle of the bar. The rod is not necessary, but it helps to damp out vibrations of the wire. Masses of 10 kilograms are mounted at both ends of the bar as in the classical apparatus. Near one end a magnet is glued into a hole extending through the bar. When the bar begins to rotate, a solenoid in the servomechanism pulls magnetically on the magnet to eliminate the rotation.

The optical sensor consists of an infrared emitter and an infrared detector positioned near the end of the bar with the magnet. The circuitry of the apparatus is shown in the illustration below. The emitter is a simple diode that radiates a small amount of heat as current from a constant-voltage source flows through it. The detector is a transistor with electrical characteristics that change when it is exposed to heat. (The combined emitter and detector can be bought as an optical interrupter unit, one such being the General Electric H13A1.)

Between the emitter and the detector a vane is attached to the bar. Initially the vane shields half of the detector from the radiation. As the bar begins to rotate in one direction the detector is more exposed and the servomechanism acts to restore the bar to its initial orientation. Rotation in the other direction shields more of the detector and again brings the servomechanism into play.

The output of the detector passes through a network of two operational amplifiers and other devices. The system constantly compares the detector's output to a reference voltage of two volts. If the bar is in its proper position and the vane shields half of the detector from the infrared emitter, the output matches the reference voltage. If the bar rotates so as to expose the infrared detector, the match is disrupted and the system sends current through the solenoid, generating a magnetic field that pulls on the magnet attached to the bar. The amount of current required for the correction is read indirectly from the change in the output voltage (V_o in the illustration). This output appears on a voltmeter or a strip-chart recorder.

The first step in calibrating the apparatus is to calculate the torsion constant of the wire from which the bar is suspended. When the bar is rotated through an angle, the twisted wire supplies a countering torque equal to its torsion constant multiplied by the angle of rotation. To compute the torsion constant Crandall determines two numbers. With the servomechanism turned off he makes the bar swing around its equilibrium position and measures the period of swinging. This amount of time is called the free period of the bar. Next the bar's moment of inertia (with respect to the point of suspension) is computed. The torsion constant of the wire is then calculated by the formula $k = I(2\pi/P)^2$, where k is the torsion constant, I is



The circuitry of the experiment with gravity

the moment of inertia and P is the free period.

The next step is to relate the output voltage from the servomechanism to a torque on the bar. With the servomechanism turned on the support for the suspension wire is moved through a measured angle in order to twist the wire. The bar remains in place because of the magnetic pull from the solenoid. Thus the output voltage from the circuit shifts. The voltage shift is divided by the angle of twist and the torsion constant to yield what Crandall calls the torsional gain of the apparatus. This factor (volts per unit of torque) allows him to convert any output voltage measured during an experiment into the value of the torque on the bar because of the gravitational pull on one end of the bar.

For example, when a 15-kilogram mass almost touches one of the mounted masses, the output voltage changes by about 25 millivolts. The strength of the force on the mounted mass is computed by multiplying the shift in the output voltage by half the length of the bar and dividing by the torsional gain. The gravitational constant is calculated by inserting this force into Newton's equation for gravitation, along with the distance between the center of the 15-kilogram mass and the center of the mounted mass.

To ease the introduction of an additional mass Crandall sometimes employs a rotatable plank with a 15-kilogram mass mounted at each end. It functions something like the revolving tray known as a lazy Susan. By rotating the plank in one direction one of the masses is brought to a side of the mass mounted on the bar. The bar is pulled toward the additional mass but barely moves because of the servomechanism. When the plank is rotated the other way, the other 15-kilogram mass is brought to the opposite side of the mounted mass.

The settling time for the apparatus is set by the compensation capacitor, which has a value of about one microfarad. When Crandall brings a 15-kilogram mass near one end of the bar, the shift in the output voltage appears after only a few minutes. This speed makes Crandall's version of the Cavendish experiment ideal for a classroom demonstration. An accurate measurement of the gravitational constant calls for only 20 minutes or so of observation.

For his work on the gravitational constant Crandall won the 1981 Apparatus Competition of the American Association of Physics Teachers. This experiment and the one measuring Planck's constant will appear soon in *American Journal of Physics*. The experiment with the Doppler shift of light has already been published there. Crandall's colleague Bruce Eaton contributed ideas to some of these projects.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

METAMAGICAL THEMAS

MIND, BRAIN, AND HUMANIST VALUES. Roger Sperry in *New Views on the Nature of Man,* edited by John R. Platt. University of Chicago Press, 1965.

THE SELFISH GENE. Richard Dawkins. Oxford University Press, 1976.

THE SCHEME OF THINGS. Allen Wheelis. Harcourt Brace Jovanovich, Inc., 1980.

CHANCE AND NECESSITY: AN ESSAY ON THE NATURAL PHILOSOPHY OF MODERN BIOLOGY. Jacques Monod. Alfred A. Knopf, Inc., 1981.

THETAMAGICAL MEMAS: SENTENCE-LIKE VIRUSES AND SELF-STRUCTURING REPLICATORS. Egbert B. Gebstadter in *Literary Australian,* Vol. 1, No. 248, pages 22–14; January, 1983.

THE LARGE-SCALE CULTIVATION OF MAMMALIAN CELLS

TISSUE CULTURE: METHODS AND APPLICATIONS. Edited by Paul F. Kruse, Jr., and M. K. Patterson, Jr. Academic Press, 1973.

CELL CULTURE. Edited by William B. Jacoby and Ira H. Pastan. Academic Press, 1979.

OPTIMIZING CULTURE CONDITIONS FOR THE PRODUCTION OF ANIMAL CELLS IN MICROCARRIER CULTURE. J. M. Clark and M. D. Hirtenstein in *Annals of the New York Academy of Sciences,* Vol. 369, pages 33–46; 1981.

MACHINES THAT WALK

ANIMALS IN MOTION. Eadweard Muybridge. Dover Publications, Inc., 1957.

USING BODY SIZE TO UNDERSTAND THE STRUCTURAL DESIGN OF ANIMALS: QUADRUPED LOCOMOTION. T. A. McMahon in *Journal of Applied Physiology,* Vol. 39, No. 4, pages 619–627; 1975.

ADAPTIVE LOCOMOTION OF A MULTI-LEGGED ROBOT OVER ROUGH TERRAIN. R. B. McGhee and G. I. Iswandhi in *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. SMC-9, No. 4, pages 176–182; 1979.

DYNAMIC STABILITY AND RESONANCE IN A ONE-LEGGED HOPPING MACHINE. M. H. Raibert in *Proceedings of the 4th Symposium on Theory and Practice of Robots and Manipulators,* edited by A. Morecki, G. Bianchi and K. Kedzior. Elsevier Scientific Publishing Co., 1981.

A WALKING ROBOT. Ivan E. Sutherland. The Marcian Chronicles, Inc., Pittsburgh, 1982.

THE HAIR CELLS OF THE INNER EAR

THE AUDITORY PERIPHERY: BIOPHYSICS AND PHYSIOLOGY. Peter Dallos. Academic Press, 1973.

SENSITIVITY, POLARITY, AND CONDUCTANCE CHANGE IN THE RESPONSE OF VERTEBRATE HAIR CELLS TO CONTROLLED MECHANICAL STIMULI. A. J. Hudspeth and D. P. Corey in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 74, No. 6, pages 2407–2411; June, 1977.

THE FOOTPRINTS OF EXTINCT ANIMALS

ICHNIA TETRAPODORUM. O. Kuhn in *Fossilium Catalogus I: Animalia,* edited by F. Westphal. W. Junk, The Hague, 1963.

DIE FOSSILIEN CAURIERFÄHRTEN. Hartmut Haubold. Die Neue Brehm-Bücherei, Ziemsen, Wittenberg Lutherstadt, 1974.

FOSSIL TRACKS AND IMPRESSIONS OF VERTEBRATES. William A. S. Sarjeant in *The Study of Trace Fossils: A Synthesis of Principles, Problems, and Procedures in Ichnology,* edited by Robert W. Frey. Springer-Verlag, 1975.

NMR SPECTROSCOPY OF LIVING CELLS

CELLULAR APPLICATIONS OF ^{31}P AND ^{13}C NUCLEAR MAGNETIC RESONANCE. R. G. Shulman, T. R. Brown, K. Ugurbil, S. Ogawa, S. M. Cohen and J. A. den Hollander in *Science,* Vol. 205, No. 4402, pages 160–166; July 13, 1979.

MAPPING OF METABOLITES IN WHOLE ANIMALS BY ^{31}P NMR USING SURFACE COILS. Joseph J. H. Ackerman, Thomas H. Grove, Gordon G. Wong, David G. Gadian and George K. Radda in *Nature,* Vol. 283, No. 5743, pages 167–170; January 10, 1980.

EXAMINATION OF A CASE OF SUSPECTED MCARDLE'S SYNDROME BY ^{31}P NUCLEAR MAGNETIC RESONANCE. Brian D. Ross, George K. Radda, David G. Gadian, Graeme Rocker, Margaret Esiri and James Falconer-Smith in *The New England Journal of Medicine,* Vol. 304, No. 22, pages 1338–1342; May 28, 1981.

IN VIVO CARBON-13 NUCLEAR MAGNETIC RESONANCE STUDIES OF MAMMALS. J. R. Alger, L. O. Sillerud, K. L. Behar, R. J. Gillies, R. G. Shulman, R. E. Gordon, D. Shaw and P. E. Hanley in *Science,* Vol. 214, No. 4521, pages 660–662; November 6, 1981.

NUCLEAR MAGNETIC RESONANCE AND ITS APPLICATIONS TO LIVING SYSTEMS. David G. Gadian. Oxford University Press, 1982.

THE PHYSICS OF ORGAN PIPES

SOUNDING MECHANISM OF THE FLUTE AND ORGAN PIPE. John W. Coltman in *The Journal of the Acoustical Society of America,* Vol. 44, No. 4, pages 983–992; October, 1968.

ON THE MECHANISM OF SOUND PRODUCTION IN ORGAN PIPES. Samuel A. Elder in *The Journal of the Acoustical Society of America,* Vol. 54, No. 6, pages 1554–1564; December, 1973.

SOUND PRODUCTION BY ORGAN FLUE PIPES. N. H. Fletcher in *The Journal of the Acoustical Society of America,* Vol. 60, No. 4, pages 926–936; October, 1976.

WAVE PROPAGATION ON TURBULENT JETS. S. Thwaites and N. H. Fletcher in *Acustica,* Vol. 45, No. 3, pages 175–179; July, 1980.

HARMONIC GENERATION IN ORGAN PIPES, RECORDERS AND FLUTES. N. H. Fletcher and Lorna M. Douglas in *The Journal of the Acoustical Society of America,* Vol. 68, No. 3, pages 767–771; September, 1980.

WAVE PROPAGATION ON TURBULENT JETS, II: GROWTH. S. Thwaites and N. H. Fletcher in *Acustica,* Vol. 51, No. 1, pages 44–49; June, 1982.

THE ORIGIN OF THE COSMIC X-RAY BACKGROUND

THE LUMINOSITY OF SERENDIPITOUS X-RAY QSOS. Bruce Margon, Gary A. Chanan and Ronald A. Downer in *The Astrophysical Journal,* Vol. 253, No. 1, Part 2, pages L7-L11; February 1, 1982.

THE MASS PRODUCTION OF IRON CASTINGS IN ANCIENT CHINA

THE DEVELOPMENT OF IRON AND STEEL TECHNOLOGY IN CHINA. Joseph Needham. The Newcomen Society, 1958.

T'IEN KUNG K'AI-WU (CHINESE TECHNOLOGY IN THE 17TH CENTURY). Sung Ying-Hsing, translated by E-Tu Zen Sun and Shiou-Chuan Sun. The Pennsylvania State University, 1966.

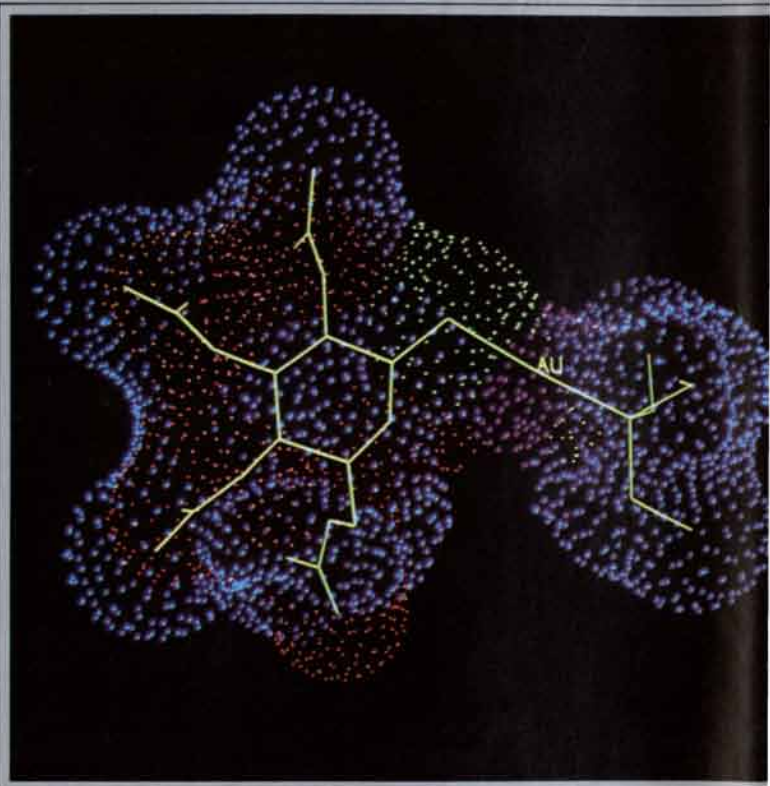
A HISTORY OF METALLURGY. R. F. Tylecote. The Metals Society, 1976.

THE AMATEUR SCIENTIST

LISTENING TO THE DOPPLER SHIFT OF VISIBLE LIGHT. R. E. Crandall and E. H. Wishnow in *American Journal of Physics,* Vol. 49, No. 5, pages 477–478; May, 1981.

SmithKline

SmithKline Beckman Corporation/World Headquarters



Before SmithKline united with Beckman Instruments to become SmithKline Beckman Corporation, those two world leaders in technology had made independent advances for diagnosis and treatment of rheumatoid arthritis. We now continue our complementary efforts.

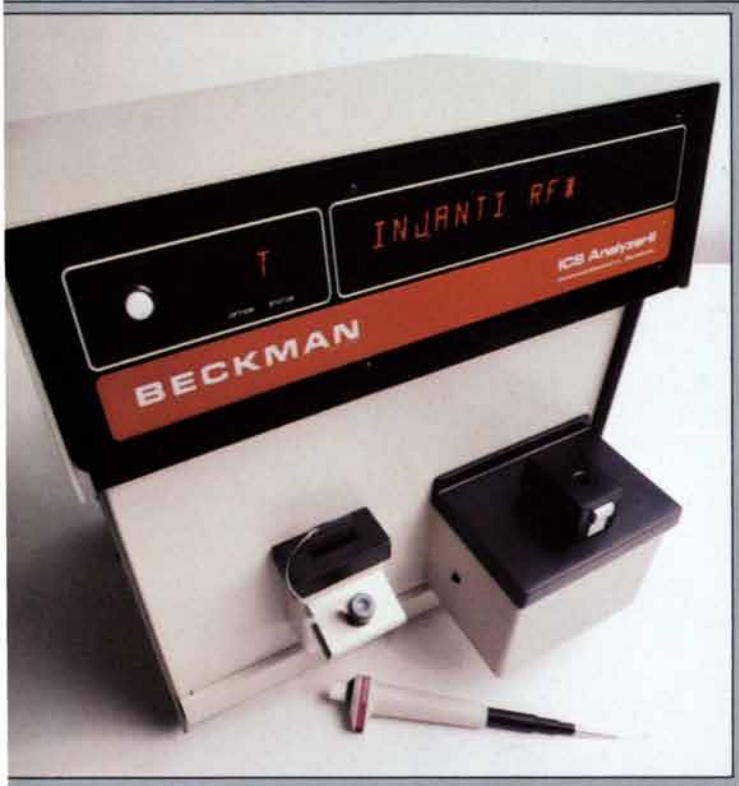
Immunology and Therapeutics. The SmithKline research program that yielded our promising oral gold agent now under investigation (above, a computer model of the auranofin molecule) has been extended to study on the molecular level how the body's immune system operates and how changes in immunity may affect autoimmune diseases and alter susceptibility to infection.

Expanding our research programs in

Converging

Beckman

One Franklin Plaza/Philadelphia, PA 19101



cellular and molecular biology, we are also investigating new aspects of gene regulation and function to develop new therapeutic agents targeted to key cellular macromolecules.

Immunochemistry and Diagnostics. The capabilities of the Beckman rate-nephelometric Immunochemistry System (above) now include precise analysis of the RF protein marker correlative with rheumatoid arthritis.

Advances in our understanding of similar biochemical markers in other diseases promise new opportunities for development of sensitive assay methods and sophisticated instrumentation.

Look into SmithKline Beckman. And look into the future.

Technologies

Panasonic Omni Series™ Component Video. Television may be the last reason you buy it.

Today, television takes a quantum leap into tomorrow. A leap into the worlds of video recreation, computer information and prerecorded programming. And with Panasonic Omni Series Component Video, you'll be ready for this video revolution!

Ready, with a separate video monitor, with the computer-designed Panasonic CompuFocus optical system. A state-of-the-art comb filter, ultra-wide lens, and tinted screen combine to give you a picture that's 20% sharper than any Panasonic before.

Ready, with a separate master control that lets you add a computer, video game and video recorder. And switch easily from one to the other, or even to cable, turning your television into a complete home entertainment center.

Ready, with separate stereo speakers that can deliver the sound the Omni Series picture demands.

You'll find the quality of Omni Series technology in every Panasonic color TV, from the world's smallest, 2.6" to a giant screen 45" projection TV (both meas diag).

So don't buy just another TV set. Buy Panasonic Omni Series Component Video. It's ready for the future. Right now.

Panasonic.

just slightly ahead of our time.