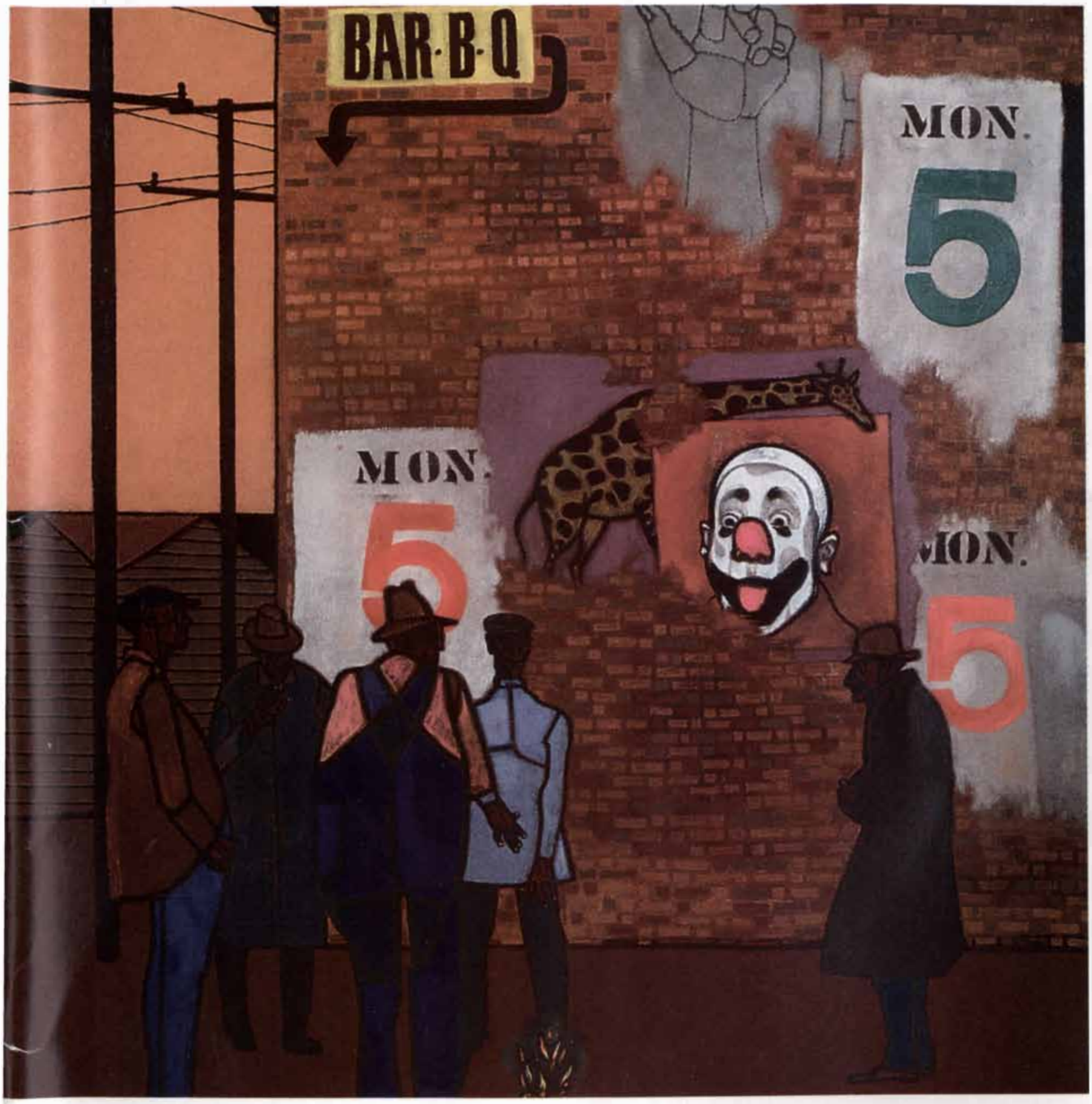
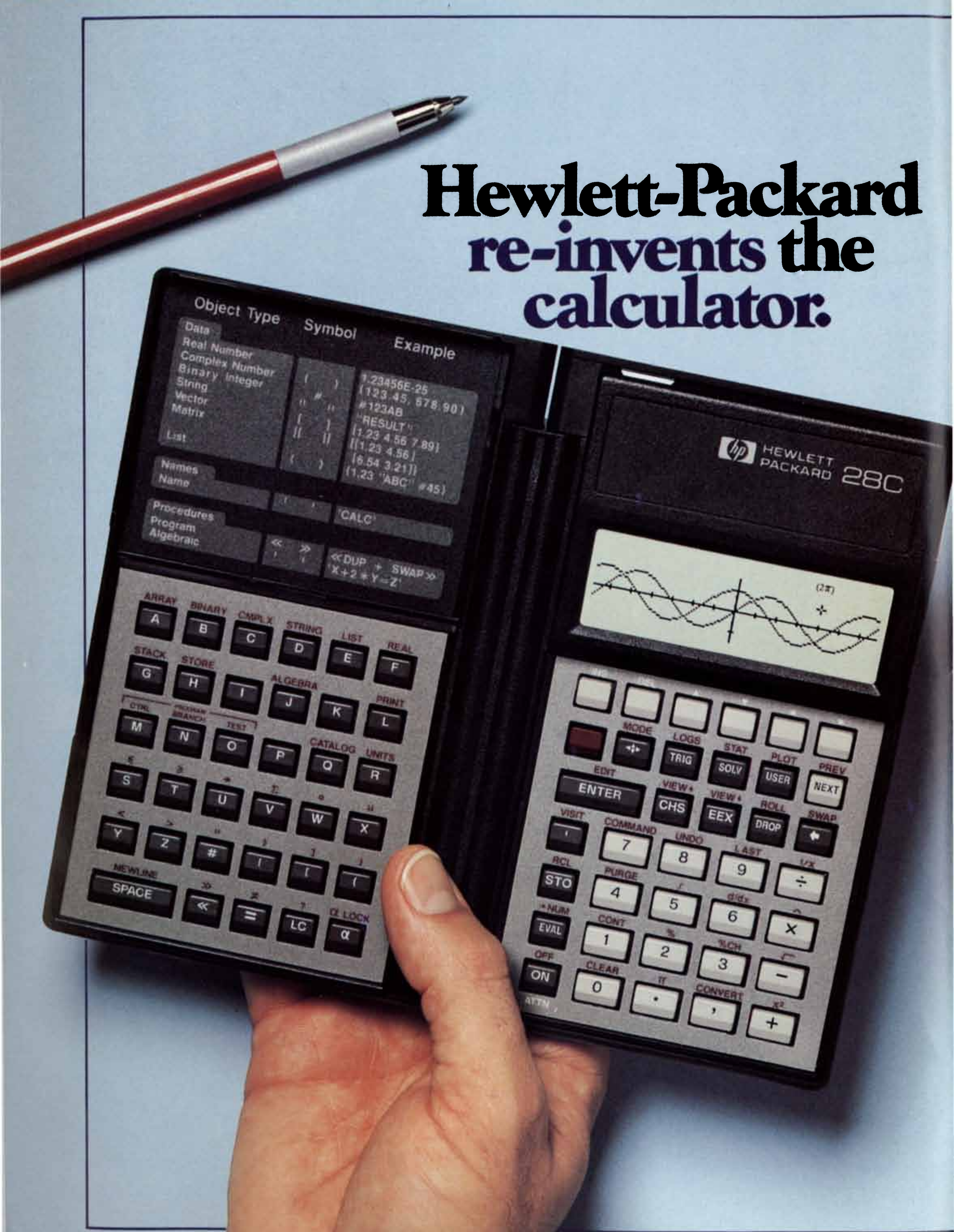
SCIENTIFIC AMERICAN



HUNGER IN THE U.S.

$2.50

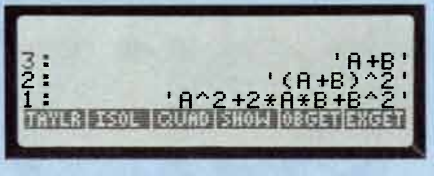
February 1987


Hewlett-Packard
re-invents the
calculator.
HEWLETT PACKARD 28C
Object Type
Symbol
Example
Data
Real Number
Complex Number
Binary Integer
String
Vector
Matrix
List
Names
Name
Procedures
Program
Algebraic

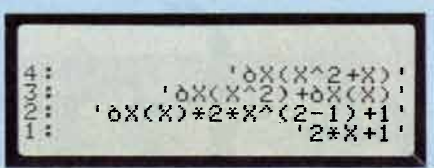
ENGINEERING RESPONSIBILITY ☒	SEPIA ☒	D-00028-40001-1		
SYM	REVISIONS		APPROVED	DATE
A	AS ISSUED		DRE	9/8/86

THE NEW HP-28C DOES THINGS NO OTHER CALCULATOR CAN. AND IT DOES MORE THINGS THAN ANY OTHER CALCULATOR CAN. MORE THAN THE UNIQUE FUNCTIONS DISPLAYED ON THE LEFT, THE HP-28C ALSO BRINGS TOGETHER SUCH FEATURES AS KEYSTROKE PROGRAMMABILITY; RPN LOGIC WITH ALGEBRAIC EXPRESSION ENTRY; 120 UNIT CONVERSION FACTORS; HEX, OCTAL AND BINARY MATH--AND CONVERSIONS; ADVANCED STATISTICS; AND THE ABILITY TO USE AN OPTIONAL PRINTER VIA INFRARED BEAM. FOR A DEMONSTRATION, CALL 1-800-367-4772, EXT. 142 A AND ASK FOR THE NAME OF YOUR NEAREST HP DEALER. THEN SEE QUOD ERAT DEMONSTRANDUM.

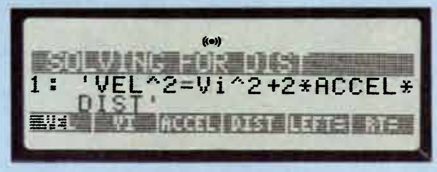
THE FIRST CALCULATOR THAT DOES SYMBOLIC ALGEBRA



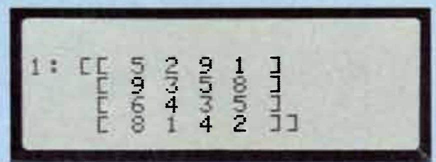
THE FIRST CALCULATOR THAT DOES SYMBOLIC CALCULUS



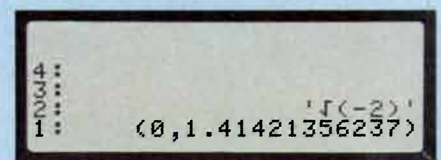
THE FIRST SCIENTIFIC CALCULATOR THAT ACCEPTS YOUR OWN FORMULAS, THEN SOLVES FOR ANY UNKNOWN



THE FIRST CALCULATOR THAT PERFORMS MATRIX OPERATIONS AS EASILY AS FOUR-FUNCTION MATH



THE FIRST CALCULATOR THAT DOES COMPLEX NUMBER ARITHMETIC AS EASILY AS FOUR-FUNCTION MATH



Q.E.D.

HEWLETT PACKARD

DO NOT SCALE THIS DRAWING			© 1987 Hewlett-Packard Company	PG12611		SEE NOTE
UNLESS OTHERWISE SPECIFIED, DIMENSIONS ARE IN INCHES.	ITEM	QTY.	PART/MATERIAL-DESCRIPTION	MAT'L-PART NO.	MAT'L-DWG. NO.	MAT'L-SPEC.
TOLERANCES .XX±.02 .XXX±.005	JERRY STEIGER DRAWN BY	4/20/86 DATE	TOPCASE - NUMERIC		00028-40001 PART NUMBER	
	JERRY STEIGER ENGINEER/CHECKER	4/20/86				
	RELEASE TO PROD.		TITLE			
SEE CORP. STD. 608	SUPERSEDES DWG.		2X SCALE	SHEET 1 OF 1	D-00028-40001-1	

Established 1845 SCIENTIFIC AMERICAN February 1987 Volume 256 Number 2

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: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

SCIENTIFIC AMERICAN
In Other Languages

LE SCIENZE
L. 3,500/copy L. 35,000/year L. 45,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Via del Lauro, 14,
20121 Milano, Italy
Advertising correspondence:
Publietas, S.p.A., Via Cino de Duca, 5,
20122 Milano, Italy

サイエンス
Y880/copy Y9600/year Y13,000/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
450 Ptas/copy 4950Ptas/year $35/(abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

POUR LA SCIENCE
25FF/copy 245FF/year 290FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

Spektrum DER WISSENSCHAFT
9.80 DM/copy 99 DM/year 112.20 DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft GmbH & Co.
Moenchhofstrasse, 15
D-6900 Heidelberg,
Federal Republic of Germany
Advertising correspondence:
Gesellschaft Für Wirtschaftspublizistik
Kasernenstrasse 67
D-4000 Duesseldorf,
Federal Republic of Germany

科学
1.40RMB/copy 16RMB/year $24/(abroad)
Editorial, subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

В МИРЕ НАУКИ
2R/copy 24R/year $70/(abroad)
Editorial correspondence:
MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc.
12224 Parklawn Drive,
Rockville, MD 20852, USA

TUDOMÁNY
98Ft/copy 1,176Ft/year 2,100Ft/(abroad)
Editorial correspondence:
TUDOMÁNY
H-1536 Budapest, Pf 338
Hungary
Subscription correspondence:
"KULTURA"
H-3891 Budapest, Pf. 149
Hungary

العلوم
1KD/copy 10KD/year $40/(abroad)
Editorial, subscription, advertising correspondence:
MAJALLAT AL-OLOOM
P.O. BOX 20856 Safat,
13069 – Kuwait

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017
Telephone: (212) 754-0550 Telex: 236115

THE COVER

The painting on the cover captures a social milieu particularly vulnerable to the hunger that afflicts some 20 million Americans of all races and regions (see "Hunger in the U.S.," by J. Larry Brown, page 36). It is a detail from *Bar B-Q,* painted by the American artist Robert G. Gwathmey in 1978. Gwathmey's stark paintings are commonly peopled with sharecroppers and the urban poor, groups in which poverty, hunger and malnutrition are often found. When hunger is defined as a chronic shortage of the nutrients required for growth and good health, the number of people in the U.S. now found to be hungry approximates 9 percent of the nation's population.

THE ILLUSTRATIONS

Cover painting by Robert G. Gwathmey,
courtesy of Terry Dintenfass, Inc.

Page	Source
17–22	Carter Bays and Philip Moore, University of South Carolina
36	Jim Kelly
39–41	Johnny Johnson
43	Carol Donner
44	U.S. Geological Survey
45	Jeffrey L. Travis, Vassar College
46	Dieter Seitz-Tutter and Dieter G. Weiss, Technical University of Munich
47	Alan D. Iselin (*top*), Dieter Seitz-Tutter and Dieter G. Weiss (*bottom*)
48	Dieter Seitz-Tutter and Dieter G. Weiss
49	Alan D. Iselin (*top*); George M. Langford, University of North Carolina School of Medicine (*bottom*)
51–56	George Retseck
90	Sheldon Wolff, University of California at San Francisco
92	Andrew Christie
93	Andrew Christie (*top*); Charles M. Radding, Yale University School of Medicine (*bottom left*); Andrew Taylor, Fred Hutchinson Cancer Research Center (*bottom right*)
95–100	Andrew Christie
102	Michael Rosenfield, Carmello Aliotta and Gordon Wilson, IBM Thomas J. Watson Research Center
104–110	George V. Kelvin, Science Graphics
112–113	Peter Ott
114	Andrew Tomko (*top*), Hank Iken (*bottom*)
115	William K. Sacco, Yale University
116	Larry Martin, LGL Ecological Associates, Inc. (*top*); C. Hans Nelson (*bottom*)
116A	William K. Sacco
116B	Andrew Tomko
120–125	Jerome Kuhl
127	David Rentas
128–130	Gabor Kiss
132	James Kilkelly
133	Gabor Kiss
135	Quesada/Burke
136–138	Michael Goodman

A COMPUTER MATERIALS CENTER LIKE NO PLACE ON EARTH.

Because they're faster and more powerful than silicon chips, microchips made of gallium arsenide are the future of information processing.

But when gallium arsenide is "grown" on Earth, imperfections occur that can render it useless. In gravity-free space, gallium arsenide can be made pure enough for all its applications: Small computers, fiber optics, satellite Earth stations, electronic defense systems.

That's one reason why McDonnell Douglas is working with team members RCA, IBM, Lockheed and Honeywell to put a manned space station in orbit in the 1990s.

For three decades, we've worked toward making a manned station operational. We've built space laboratories, flown space factories and tested space station frameworks.

And we've developed the first gallium arsenide computer chip to be used as a microprocessor.

Gallium arsenide microchips could revolutionize the power and economy of information processing. And we think that's a very good reason to build a computer materials center like no place on Earth.

For more information, write: McDonnell Douglas, Space Station, P.O. Box 14526, St. Louis, MO 63178.

LETTERS

To the Editors:

Your report "Untrashing Margaret Mead" ["Science and the Citizen," SCIENTIFIC AMERICAN, November, 1986] is biased in the extreme and markedly unscientific.

Following nine months' fieldwork in 1925–26 Margaret Mead, in *Coming of Age in Samoa,* published her conclusion (fully in accord with what was then the ruling ideology of American cultural anthropology) that biological variables are of no significance in the etiology of adolescent behavior; "we cannot," she said on page 197, "make any explanations" in terms of the biological process of adolescence itself.

Today, as Stephen Jay Gould has put it, "every scientist...knows that human social behavior is a complex and indivisible mix of biological and social influences." It is hence evident, in the light of modern scientific knowledge, that Mead's conclusion could not possibly have been true in 1928 or for that matter at any other time, and, being contrary to nature, reason and common sense, is, conspicuously, a preposterous conclusion.

Yet from the late 1920's onward this conclusion, having been endorsed by Franz Boas, became a principal dogma of American cultural anthropology. Thus Mead's erroneous conclusion was repeated as though it were scientific fact in an unbroken succession of anthropological textbooks, as it was in *American Anthropologist,* the journal of the American Anthropological Association, and, not having been seriously questioned by any American anthropologist, was taught in American universities and colleges right up until my refutation of 1983.

As W. W. Bartley III noted in his essay "A Popperian Harvest," "there is a basic asymmetry between the verification and the falsification of a theory," for while "no amount of observation" can ever "verify a theory logically, one single observation may falsify it, serving logically as a counter-example to it." For instance, when Mead asserts, as part of her romantic depiction of Samoa, that "the idea of forceful rape or of any sexual act to which both participants do not give themselves freely is completely foreign to the Samoan mind," all that is needed to refute this absolute generalization is a test statement in the form of an authenticated instance of forcible rape that occurred in Samoa during the time to which Mead's findings refer. As I have documented, there are numerous instances of forcible rape by male Samoans of Samoan females recorded in the archives of the high courts of both American and Western Samoa for the years 1920–29. It is in these Popperian terms that my refutation proceeds.

Mead's erroneous conclusion can, moreover, be definitively falsified by the evidence she herself provides. For example, in Mead's own account it is evident that at least four individuals in her sample of 25 adolescent girls were delinquents. This means, as the calculation of a rate from these cases demonstrates, that in the 1920's delinquency, with its attendant storm and stress (as described by Mead herself), was present among Samoan adolescent girls at as high a rate as has been established for Samoa in later decades and for the other 20th-century societies, such as the U.S. and Australia, for which delinquency rates are available.

All the evidence in my book *Margaret Mead and Samoa: The Making and Unmaking of an Anthropological Myth* has been subjected to rigorous critical testing, including testing by Samoan academics. Much of it is drawn from court archives and from the proceedings of congressional and royal commissions referring specifically to the 1920's. For Roy A. Rappaport, in the article on which your story was based, to dismiss this range of empirical evidence as "no more" than "half-truths" is both a failure to comprehend the significance of "test statements" and a scientifically unwarranted evasion of their cogency.

Such irrational defense of Mead's erroneous conclusion has been commonplace among those who have come fervently to believe in it, and for whom prophecy has failed. For example, at the annual meeting of the American Anthropological Association in 1983, my attempt to rescue the members of that association from "the philosophic indignity of persisting unnecessarily in error" was denounced *by vote,* in terms reminiscent of the days of Calvin, as though scientific issues could be settled politically! And one young professor, expert in things Samoan, who at the outset had averred that "people who work in Samoa know Margaret Mead was wrong, and Freeman's book shows that beyond doubt," not wishing to be mobbed, dramatically changed his stance. It was these and comparable happenings that in 1984 caused Richard Goodman (who has published a wholly independent refutation of *Coming of Age in Samoa*) to observe in the *Newsletter of the American Anthropological Association* that "something is rotten in the state of American anthropology." Now Rappaport, the president-elect of this association, lacking any research experience in Samoa, has attempted, in a massive essay in anthropological mystification in the *American Scholar,* in spite of its demonstrable errors and false conclusion, to reinstate *Coming of Age in Samoa* as a "humane and liberating" text!

Further, in defense of Mead's cherished if erroneous conclusion, Rappaport has resorted to decidedly reprehensible methods. In my writings on Samoa I have repeatedly noted that my researches involved detailed observations of and inquiries addressed to Samoan girls. Yet Rappaport, in spite of the ready availability of this evidence, makes the quite false assertion that my concern, in contrast to Mead's, was "with an ideology of sex held publicly by elite adults." Again, while I have always been an outspoken opponent of biological determinism, racism and Galtonian eugenics, and while in my book I specifically note the "terrible culmination [to which eugenics led] in National Socialist Germany," Rappaport depicts me as condoning these repugnant ideologies. These false assertions and others, which are obviously designed to discredit, are so deployed by Rappaport as to amount to a campaign of disinformation. Such debased political techniques have no rightful place in scientific controversy.

I am now working on a second book in which, in the course of a comparative analysis of the role of controversy in science, I shall deal in detail with the radically irrational reaction of members of the American Anthropological Association to the cogency of my refutation and with Professor Rappaport's studied obscurantism. The title I have in mind for this study is *When Prophecy Failed.* A fitting epigraph would be the chastening words of Franz Boas: "All that man can do for humanity is to further the *truth,* whether it be sweet or bitter."

DEREK FREEMAN

Emeritus Professor of Anthropology
Research School of Pacific Studies
The Australian National University
Canberra

To the Editors:

I shall try to put aside my resentment at being accused of resorting to "decidedly reprehensible methods" in "defense of Mead's cherished if erroneous conclusion" to observe that Professor Freeman's representation of Mead's conclusion as a categorical dismissal of biological factors in the explanation of human behavior does not accord with what she said. Here is the relevant passage, on page 197 of *Coming of Age in Samoa*:

"A further question presents itself. If it is proved that adolescence is not necessarily a specially difficult period in a girl's life—and proved it is if we can find any society in which that is so—then what accounts for the presence of storm and stress in American adolescents? First, we may say quite simply, that there must be something in the two civilisations to account for the difference. If the same process takes a different form in two different environments, we cannot make any explanations in terms of the process, for that is the same in both cases. But the social environment is very different and it is to that we must look for an explanation. What is there in Samoa which is absent in America, what is there in America which is absent in Samoa, which will account for the difference?"

The "same process" that takes "different form" in "different environments" is identified on page 196 as "the process of growth by which the girl baby becomes a grown woman." Mead assumes the "developing girl is a constant factor in America and in Samoa" and asks if "the sudden and conspicuous bodily changes which take place at puberty [everywhere]" are necessarily accompanied by the kinds of emotional and cognitive upheavals common in American girls. Are conflict and stress inevitable concomitants of "change in the girl's body"?

Mead, it seems clear, recognized the biological character of puberty, never claimed that biological factors have nothing to do with behavior and simply stated that differences in the emotional and cognitive correlates of "the same [biological] process" in "different environments" are to be accounted for by differences in environment. Hardly preposterous. Furthermore, neither Mead's mentor, Franz Boas (himself a physical anthropologist as well as an ethnographer), nor *any* of his students *ever* propounded what Freeman called in his book "the Boasian Paradigm"—the notion "that human behavior can be explained in purely cultural terms." They did not argue against the importance of biological components in the behavior of what is, after all, an organism. Boas and his students did, however, reject the notion that differences between *cultures* are to be accounted for racially or genetically. Taking mental capacities to be biological in nature, Boas put it this way in 1911:

"We are not inclined to consider the mental organization of different races of men as differing in fundamental points.... The average faculty of the white race is found to the same degree in a large proportion of individuals of all other races."

An overwhelming majority of contemporary anthropologists would, I think, still endorse this statement.

Which brings up the matter of racism. I did *not* depict Freeman as condoning the "repugnant ideologies" associated with eugenics. He himself noted that eugenic thought reached its "terrible culmination" in Nazi Germany. It was not *in spite of* this observation but because he had made it and yet had taken the eugenicists seriously that I said he seemed to view racism as a flaw in an otherwise serious, although ultimately erroneous, scientific enterprise. I take such an evaluation of eugenics as an error in judgment, but not evidence of racism. *I wish to make it very clear that I believe Freeman's record to be exemplary in this regard.*

Freeman characterizes my statement that he was largely concerned with "an ideology of sex held publicly by elite adults," in contrast to Mead, who was concerned with the "sexual behavior of adolescent girls," as part of a "campaign of disinformation" which I have "obviously designed to discredit" him. As in the case of his representation of Mead's conclusions, he lifts matters out of context. The statement appears in a discussion expanding on a quotation from a review of Freeman's book in which the writer, Robert Levy, addresses the general problem of disagreements among the ethnographies of particular societies. Levy points out that "Freeman's areas of clearer vision are different from Mead's. The problem...is one of context *and the nature of the observer as part of that context.*" (The italics are mine.) The personal characteristics of the ethnographer Freeman, a grown man holding a chiefly title, would give him easy access to categories of informants different from those among whom the 23-year-old Mead could casually mingle. I did not argue that Freeman did not do any work among Samoan girls. I will say that what adolescent girls are likely to say about sex to chiefly adults is at least as likely to reflect elite ideology as it is to report their own behavior. I might note, furthermore, that my characterization of his research as concerned with an elite ideology of sex was qualified.

We should not be surprised at differences between the accounts of ethnographers differing in personal characteristics and working in different parts of an archipelago in 1925, on the one hand, and in 1940 and later years, on the other. We have long known that no ethnographer can present the "whole truth" about any society, that ethnographic knowledge is accumulated from the accounts of many ethnographers differing in personal characteristics, theoretical concerns and working context and that differences among their accounts are likely to include contradictions. Such apparent contradictions do not necessarily indicate error but may indicate the need for deeper, more synthetic understandings.

In recent years it has become clear to anthropologists that cultures are not as internally consistent as Mead and Ruth Benedict thought when they were young. That is to say, contradictions between accounts may reflect contradictions internal to cultures themselves. For this reason, among others, a Popperian program seeking to falsify an ethnography by adducing evidence contradicting some of its observations and conclusions cannot help but produce an account as incomplete or oversimplified, and thus as wrong, as the account it criticizes; hence my characterization of Freeman's book as "a heap of...half-truths." For a fuller discussion of this point I refer readers to the original article in the *American Scholar*.

ROY A. RAPPAPORT

Professor of Anthropology
University of Michigan

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 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 __________

Hitachi's advances in transportation include the joint development with Japan National Railways of a linear motor train with a potential speed of 500km/h, a smoother, quieter linear motor train, large-capacity thyristor and monorail car.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

FEBRUARY, 1937: "The route mileage of the world's airlines now approximates 300,000 miles. This enormous network has been spun in less than 20 years. In the first year of true commercial air transport, 1919, only about 3,000 miles of routes were in operation. The past year saw immense new additions to the airway map in the Far East, in Russia, in China and, most spectacular of all, the opening of the wide Pacific as a road of the air that alone added some 9,000 miles to the existing air-transport routes. The U.S. has 28,000 airway miles."

"Days of laborious computation by trained mathematicians are no longer necessary at the Massachusetts Institute of Technology. A new one-ton machine with many complicated levers and gears will in a single action solve nine simultaneous equations with nine unknowns."

"There is a growing interest in the importance of studying tree rings. A surprisingly large amount of valuable information can be obtained in this way, which is of importance in meteorology, history and anthropology. A study of the growth rings of trees reveals information about wet and dry years, floods and forest fires for thousands of years in the past. One of the interesting results of studying tree rings is the dating of woods found in ancient ruins."

"Just 40 years ago an inventor was elated when his invention blew up; to him the explosion proved the soundness of the principle he had so carefully developed. The inventor was Dr. Rudolf Diesel. Today the Diesel engine powers vehicles of the land, sea and air and supplies the driving force in practically every field where power is used. Diesels have a future so bright that no one can predict the ultimate extent of their use."

"The disastrous floods that took place during the spring of 1936 have focused the nation's attention on the problem of devising methods to control the runoff of water and prevent recurrences of seasonal catastrophes of this sort. The last Congress authorized a gigantic flood-control program, and this action has probably been most effective in acquainting the public with the fact that devastating floods are often the result of mismanagement of a great resource."

"The air is full of tiny insects and other living matter that have been swept high above the earth by air currents. Numerous flights of airplanes carrying traps for collecting insects in Louisiana revealed that a column of air one mile square extending from 50 feet to 14,000 feet above the ground contains on an average for all seasons about 25 million insects, the range being from 12 million in January to 36 million in May."

"The succulent tender steaks of the skilled chef may soon become commonplace home fare. Methods of tenderizing tough meat are being investigated, and eventual success seems indicated. Already one product said to be an extract of papaya is being offered for sale to restaurants and hotels as producing tender steaks from cheap cuts at a cost of a cent or two. The liquid is brushed on the meat five or 10 minutes before cooking. There is every prospect that in due time even cheap meat will be tender."

"The ears get larger, the nose gets longer and broader and the mouth gets wider as one grows older. The curious age changes in the dimensions of these facial organs, sufficient to change materially the appearance of the individual, have been established by Dr. Ales Hrdlicka, Curator of Physical Anthropology of the Smithsonian Institution, by measurements of thousands of men and women."

SCIENTIFIC AMERICAN

FEBRUARY, 1887: "A glance at any map of the Eastern and Middle States will show the need of a bridge over the Hudson River at a point midway between New York and Albany. All railroad traffic between the New England States and the West and South is subject to more or less delay, caused by crossing the Hudson. The Poughkeepsie bridge, together with about 12 miles of road to be built between Poughkeepsie and Gardiner, will obviate this difficulty. The advantages to be derived by the transportation of coal over this route, and by the passenger and freight traffic between New England and the West and South, are apparent."

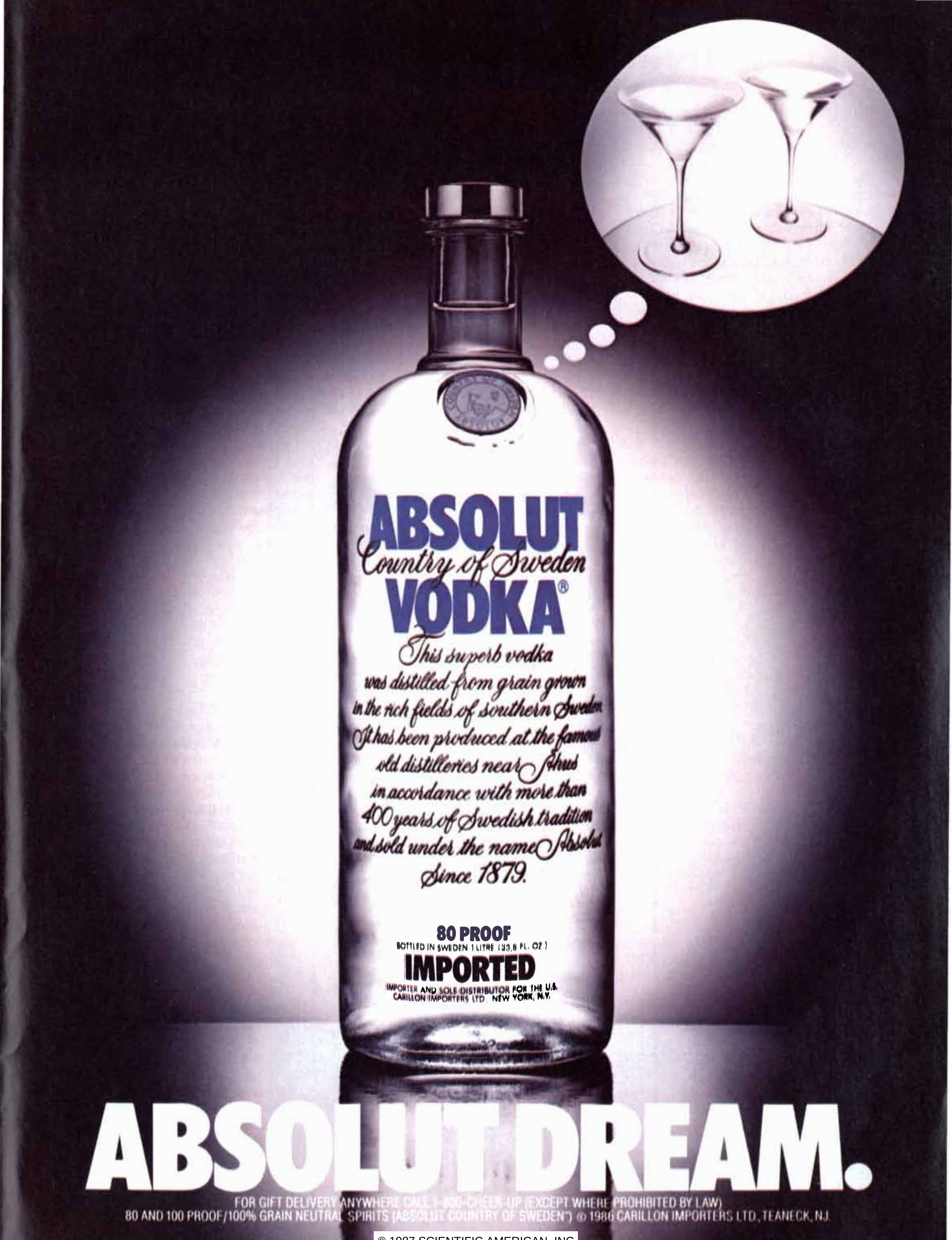
"At the forthcoming American Exhibition in London, we are promised, among other novelties, a house of straw, which is now being made in Philadelphia. This house is to represent an American suburban villa. It is constructed entirely of materials made from straw. It will demonstrate how far the inventive Yankee has succeeded, not in showing us how to make bricks without straw but in how to produce timber from straw."

"Salt is now the means for clearing away snow in Paris. A regular service for the removal of snow on its first appearance has been organized, as it is important to clear away the snow before it has been compressed into ice by the passage of vehicles, when it is far more difficult to remove. As falls of snow rarely occur at Paris with a temperature much below the freezing point, salt may be sprinkled on the snow, producing a liquid, of which the temperature may descend to 5 deg. Fahrenheit without its freezing. This cold mixture does no harm to paved roads, asphalt and wood pavements. The small cost of the system, and the advantages to traffic, are sufficient reasons for an early and wide extension of this use of salt. In this work the street crews spread salt that is not suitable for ordinary purposes."

"After an experiment of several months the ferry boats plying between San Francisco and Oakland, which had been fitted up for burning petroleum, have now gone back to coal. The economy, so far as the consumption of fuel is concerned, is said to be decidedly in favor of petroleum, but the trouble in its use came from the intense heat produced, by which the iron of the furnaces and boilers began to indicate rapid deterioration."

"A number of cases of confirmed cocaine habit have recently been reported. If the cases continue to multiply, there may be room for questioning the utility to man of the discovery of this anaesthetic. It is doubtful whether all the services in local anaesthesia rendered by it can compensate for the ill it has already done."

"Dr. Giles de la Tourette has recently published a monograph on normal locomotion and the variations in the gait caused by diseases of the nervous system. He found that the average length of pace for men is 25 inches; for women, 20 inches. The step with the right foot is somewhat longer than that with the left. The feet are separated laterally about 4½ inches in men and about five inches in women."


ABSOLUT
Country of Sweden
VODKA®
This superb vodka was distilled from grain grown in the rich fields of southern Sweden. It has been produced at the famous old distilleries near Åhus in accordance with more than 400 years of Swedish tradition and sold under the name Absolut Since 1879.
80 PROOF
BOTTLED IN SWEDEN 1 LITRE (33,8 FL. OZ.)
IMPORTED
IMPORTER AND SOLE DISTRIBUTOR FOR THE U.S. CARILLON IMPORTERS LTD. NEW YORK, N.Y.
ABSOLUT DREAM.
FOR GIFT DELIVERY ANYWHERE CALL 1-800-CHEER-UP (EXCEPT WHERE PROHIBITED BY LAW)
80 AND 100 PROOF/100% GRAIN NEUTRAL SPIRITS (ABSOLUT COUNTRY OF SWEDEN™) © 1986 CARILLON IMPORTERS LTD., TEANECK, N.J.

THE AUTHORS

J. LARRY BROWN ("Hunger in the U.S.") teaches at the Harvard School of Public Health and is executive director of the university's Community Health Program as well as chairman of the Harvard-based Physician Task Force on Hunger in America. His interest in hunger in this country dates to 1966, when a Peace Corps assignment in India alerted him to the problem on an international scale. After his return in 1968, he earned a Ph.D. in social policy from Brandeis University, and in 1971 he joined the Massachusetts Department of Health. In 1976 Brown took a position with the Federal Government administering VISTA and Peace Corps projects; he moved to Harvard in 1978.

ROBERT DAY ALLEN ("The Microtubule as an Intracellular Engine") was Ira Allen Eastman Professor of Biology at Dartmouth College and a long-time associate of the Marine Biological Laboratory in Woods Hole, Mass. After receiving a Ph.D. from the University of Pennsylvania in 1953, he did postdoctoral research on a Guggenheim fellowship in Sweden, then began his teaching career at the University of Michigan in 1956. Four years later he went to Princeton University; in 1966 he became chairman of the biology department at the State University of New York at Albany, and in 1974 he went to Dartmouth. Allen died of cancer in March, 1986. Dieter G. Weiss of the Technical University of Munich assisted in preparing his article for publication.

JACOB SHAHAM ("The Oldest Pulsars in the Universe") came to Columbia University from Israel as an Ernest Kempton Adams fellow in 1982; in 1984 he was appointed professor of physics and joined the astrophysics laboratory. He got his B.Sc. (1963), M.Sc. (1965) and Ph.D. (1971) from the Hebrew University of Jerusalem. After three years in a postdoctoral assignment at the University of Illinois at Urbana-Champaign, he returned to Jerusalem and worked for 12 years at the Racah Institute of Physics. Shaham's main interest is the physics of neutron stars and black holes. He is himself something of a "star": between 1980 and 1983 he hosted an Israeli television series on science called "Tatzpit," or "Observation."

FRANKLIN W. STAHL ("Genetic Recombination") is professor of biology at the Institute of Molecular Biology at the University of Oregon. He began his career with an A.B. from Harvard College in 1951 and earned a Ph.D. from the University of Rochester in 1956. He worked for several years at the California Institute of Technology before joining the faculty at the University of Missouri in 1958. Although Stahl has been at the University of Oregon since 1959, sabbaticals and visiting professorships have taken him to Medical Research Council laboratories in Britain, the Hebrew University of Jerusalem and the International Laboratory of Genetics and Biophysics in Naples, Italy.

MORDEHAI HEIBLUM and LESTER F. EASTMAN ("Ballistic Electrons in Semiconductors") met in 1985 when Eastman joined the visiting technical staff at the IBM Thomas J. Watson Research Center, where Heiblum had worked since 1978. Born in Israel, Heiblum earned a B.Sc. at the Israel Institute of Technology in 1973 and an M.Sc. at Carnegie-Mellon University in 1974. He received his Ph.D. in 1978 from the University of California at Berkeley and then went to IBM. Heiblum says much of his time is devoted to his research, now mainly in electronic ballistic transport, but that he also remains devoted to his wife and four children. Eastman has taught in the electrical engineering department at Cornell University since the university awarded him his Ph.D. in 1957. He also received his B.S. (1953) and M.S. (1955) at Cornell. In addition to his experience with IBM, he has held visiting positions at the RCA Laboratories in Princeton and at the Massachusetts Institute of Technology's Lincoln Laboratory. Eastman also directs Cornell's Joint Services Electronics Program, which he founded in 1980.

C. HANS NELSON and KIRK R. JOHNSON ("Whales and Walruses as Tillers of the Sea Floor") study mammal feeding on the sea floor as an adjunct to their primary research interests. Nelson, a marine geologist with the U.S. Geological Survey since 1967, is currently involved in an interdisciplinary study of the Atlantic and Mediterranean continental margins off Spain, sponsored by the Department of State. He holds a Ph.D. in geological oceanography from Oregon State University (1968), an M.A. from the University of Minnesota (1962) and a B.A. from Carleton College (1959). Nelson has supplemented his work for the government with visiting professorships at the University of Barcelona, the University of Utrecht and Stanford University. Johnson, who met Nelson when he went to work for the Geological Survey in 1982, is pursuing a Ph.D. in geology and paleobotany at Yale University. Before joining the Geological Survey Johnson earned an A.B. at Amherst College and in 1985 he received his master's at the University of Pennsylvania.

JONATHAN R. COLE and HARRIET ZUCKERMAN ("Marriage, Motherhood and Research Performance in Science") are visiting scholars at the Russell Sage Foundation, on leave from the sociology department at Columbia University. Their current study of gender in science grew out of an earlier collaboration on the reward system in science. Cole got both his B.A. (1964) and his Ph.D. (1969) from Columbia. He has worked in the sociology department since then, except for a three-year professorship at Columbia's Barnard College that began in 1973. He also directs Columbia's Center for the Social Sciences. Zuckerman has for many years been interested in the connection between social organization and the development of knowledge. Her work has led to several books and many papers, as well as a Guggenheim fellowship in 1980. She earned an A.B. at Vassar College in 1958 and a Ph.D. from Columbia in 1965 and has been on the university's faculty since then.

MAX V. MATHEWS and JOHN R. PIERCE ("The Computer as a Musical Instrument") began working together in the 1960's at the Bell Telephone Laboratories when Mathews, then director of their Acoustical and Behavioral Research Center, developed a program that enabled computers to produce complex musical sounds. Mathews went to Bell Laboratories in 1955 after getting an M.S. in 1952 and an Sc.D. in 1954 at the Massachusetts Institute of Technology. His tenure as director of the research center ended in 1985, but he continues to do research on electronic violins, real-time computer systems and the perception of musical sounds at what is now called AT&T Bell Laboratories. Pierce, who earned a B.S. (1933), an M.S. (1934) and a Ph.D. (1936) from the California Institute of Technology, joined Bell Laboratories in 1936 and stayed until 1971, when he returned to Caltech as professor of engineering. Pierce was chief technologist at the Jet Propulsion Laboratory between 1979 and 1982 and now enlivens his retirement with a visiting professorship at Stanford University.

The problem facing a growth industry is that it may quickly outgrow its workforce.

An uneducated workforce can soon cripple a company's growth. Many decision makers have learned this lesson the hard way.

Others plan for growth by looking for sites in places like Maryland. Where one out of every five people has spent four or more years in college. Where there are more than 70 colleges and universities, including Johns Hopkins and the University of Maryland. And where these institutions provide strong R&D support and extensive continuing education for local businesses.

If you're looking for a workforce with real growth potential, contact Michael Lofton, Department of Economic and Community Development, Dept. 152, 45 Calvert St., Annapolis, Maryland 21401. Or call (301) 974-3514.

Maryland
More than you can imagine.

COMPUTER RECREATIONS

The game Life acquires some successors in three dimensions

by A. K. Dewdney

Life, the popular cellular-automaton game played on a two-dimensional grid, has now inspired some analogues in three dimensions. Carter Bays, a computer scientist at the University of South Carolina, has explored a wide variety of three-dimensional versions and found two to be most promising. He calls them Life 4555 and Life 5766. Both versions reproduce many of the features of the original Life (such as its blinkers and gliders); one of them will without doubt emerge as a worthy partner of the game invented in 1968 by the mathematician John Horton Conway of the University of Cambridge.

Conway's game, as many readers will recall from several of the "Mathematical Games" columns by Martin Gardner, is played on an infinite two-dimensional grid of square cells. Each cell has eight neighbors (four at the corners and four at the sides) and may exist in one of two states, alive or dead. Somewhere a great clock ticks away. At each tick certain cells may come alive and others may die. The fate of a cell is determined by how many of its neighbors are alive. If, for example, at one tick a living cell has fewer than two living neighbors or more than three, it will be dead at the next tick, the rationale being that a living cell can be undernourished or overcrowded. A dead cell, on the other hand, will be reborn at the next tick if it has exactly three living neighbors: a birth requires three "parents."

Conway called his game Life because the cells can be either alive or dead. It quickly became apparent, however, that the name was more appropriate than suspected. Various configurations of living cells show surprisingly complex and almost lifelike behavior [*see upper illustration on page 20*]. The behavior is cyclic: at each tick of the clock the configurations change, but after a finite number of ticks the original patterns reappear. Some patterns remain stationary and others travel through the grid by shifting one cell at a time in a horizontal, vertical or diagonal direction. Both types of pattern carry fanciful names. Examples of stationary configurations include beacons, beehives, blinkers and blocks; examples of "travelers" are gliders and spaceships. Conway's Life goes far beyond mimicking natural phenomena, however. As I pointed out in this department in May, 1985, it is even possible to build a computer within Life's cellular plane.

It should not come as a surprise that in the commodious cellular space of three dimensions analogous versions of Conway's game might give rise to even more fantastic phenomena. Such versions are Carter Bays's Life 4555 and Life 5766. Here each cell is a cube instead of a square and has 26 neighbors instead of eight.

The names Life 4555 and Life 5766 are drawn from a lean lexicon developed by Bays. The first two numbers dictate the fate of the living cells. The first number indicates the fewest living neighbors a cell must have to keep from being undernourished; the second indicates the most it can have before it will be overcrowded. The third and fourth numbers govern the fate of the dead cells. The third indicates the fewest living neighbors a dead cell must have to come alive; the fourth indicates the most it can have to come alive. (In each version of Life I shall discuss here the third and fourth numbers are identical with each other; in general, however, it is not so.) According to Bays's notation, then, Conway's Life becomes Life 2333.

Life 4555 operates just as simply as Life 2333. A living cell dies if it has fewer than four or more than five living neighbors. A dead cell comes to life if it has exactly five living neighbors. In a routine investigation of rules in this range of values, Bays was first drawn to Life 4555 when he noticed an odd configuration of cubical cells wriggling out of the depths of his Macintosh display screen [*see lower illustration on page 20*]. It was a three-dimensional glider, which then cycled through four distinct patterns before repeating itself. Each pattern consisted of 10 cubes in oblong formation, strangely blunt, moving through space like a sofa in free fall.

Intrigued, Bays decided to probe the rules for 4555 a little further by setting up a number of "primordial soup" experiments. Seeding an initial space with cubes randomly brought to life, he set his cellular universe in motion. In each generation some cubes died and others came to life. The number of living cubes dwindled with each generation, but not before Bays noticed some curiously stable ensembles that did not change from one generation to the next. Some of them reminded him of pedestals, crosses, steps, balls and barbells [*see illustration on page 21*]. Subsequent seedings yielded yet other stable configurations, as well as many cyclic ones to which Bays has given such whimsical names as rotor and bucking bronco.

Like a nuclear physicist with new particles to play with, Bays set up collisions among gliders and other small configurations at every conceivable angle. "Among the most surprising collisions," he notes, "is a certain glider collision with a ball, where the resulting confused mass swells to a population of 29, in itself not particularly remarkable. But suddenly the glider reappears, lagging by several generations and shifted somewhat."

The yields of primordial soups and even the results of glider collisions are called nature by Bays. Some configurations are produced quite easily. Other and more exotic patterns require more work. For example, one can hook together arch-shaped configurations to make a new stable pattern Bays calls an arcade. An entire architecture of fences, stairs, walls and chains emerges. Walls can be bent into helixes and a great variety of stable exotic forms that are limited only by the imagination of a Life 4555 enthusiast [*see illustration on opposite page*].

Further phenomena beg description in these pages. There is a lonely seven-cube form called a greeter that dies unless it is in the presence of another greeter. A glider may pass a greeter only to be gripped, or suspended in its travels. A second glider may by chance rescue its kin by colliding with the greeter and exploding it.

Of the two fruitful three-dimensional versions discovered by Bays, Life 4555 seems to be his favorite. Curiously enough, the digits of Life 4555 can be obtained by adding 2 to the digits of the code for Conway's Life, namely 2333. Perhaps the coincidence foreshadows the eventual emergence of Life 4555 as a worthy partner to Conway's Life.

In a strict sense, however, Life 5766 mimics Conway's game more closely than Life 4555 does. In particular, under certain special conditions Life 5766 will simulate Conway's Life in the plane. The conditions are specified by a theorem hit on by Bays.

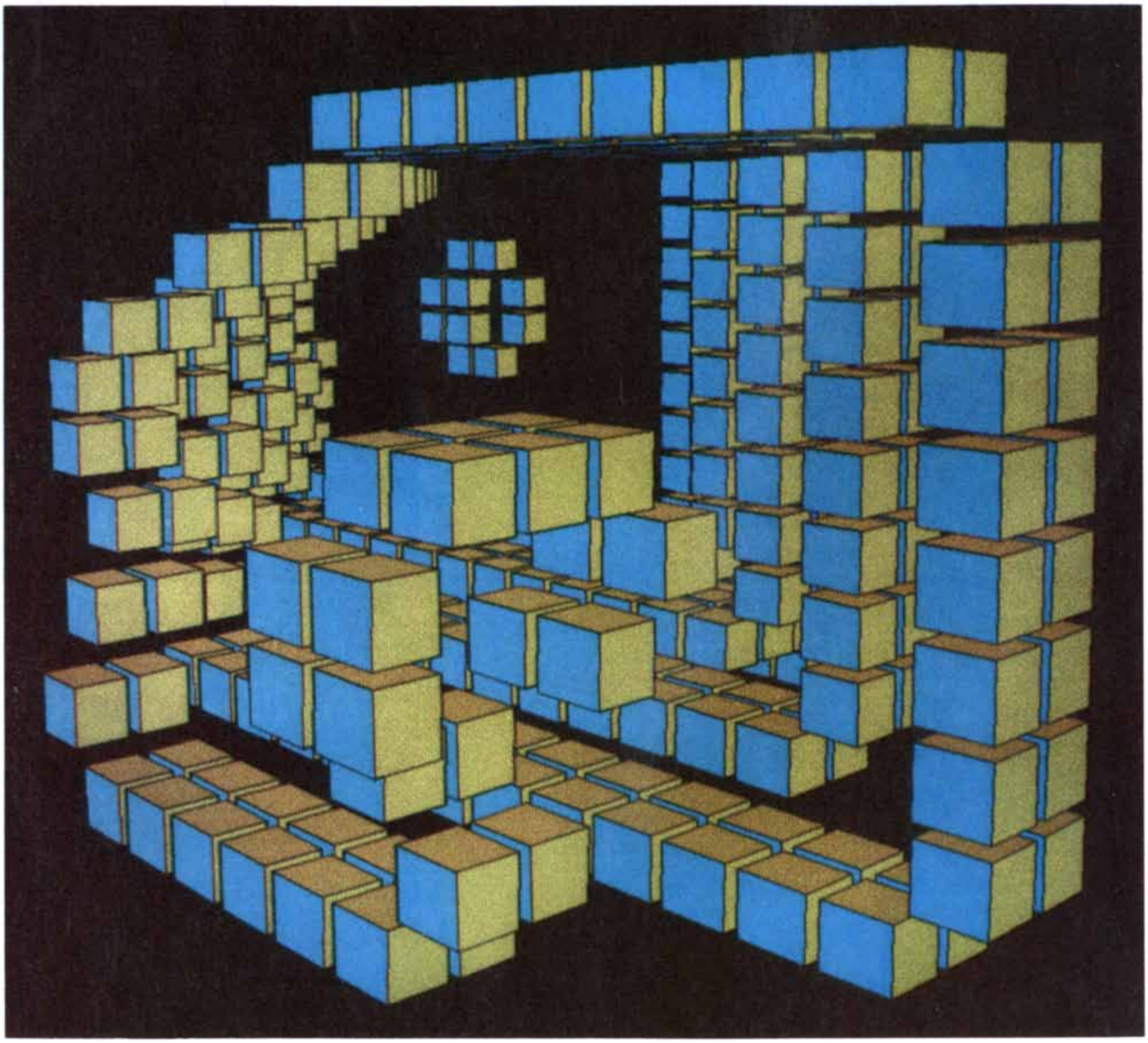
Imagine looking down on the plane grid of Conway's Life and there seeing a particular configuration of living (square) cells. Now place a living cube directly on top of each square in the configuration and another directly under it. Following the rules for Life 5766, the cubes will perfectly mimic forever the behavior of the sandwiched Conway cells if (and only if) two conditions are satisfied:

1. No living square cell on the plane ever has five living neighbors.
2. No dead square cell on the plane ever has six living neighbors.

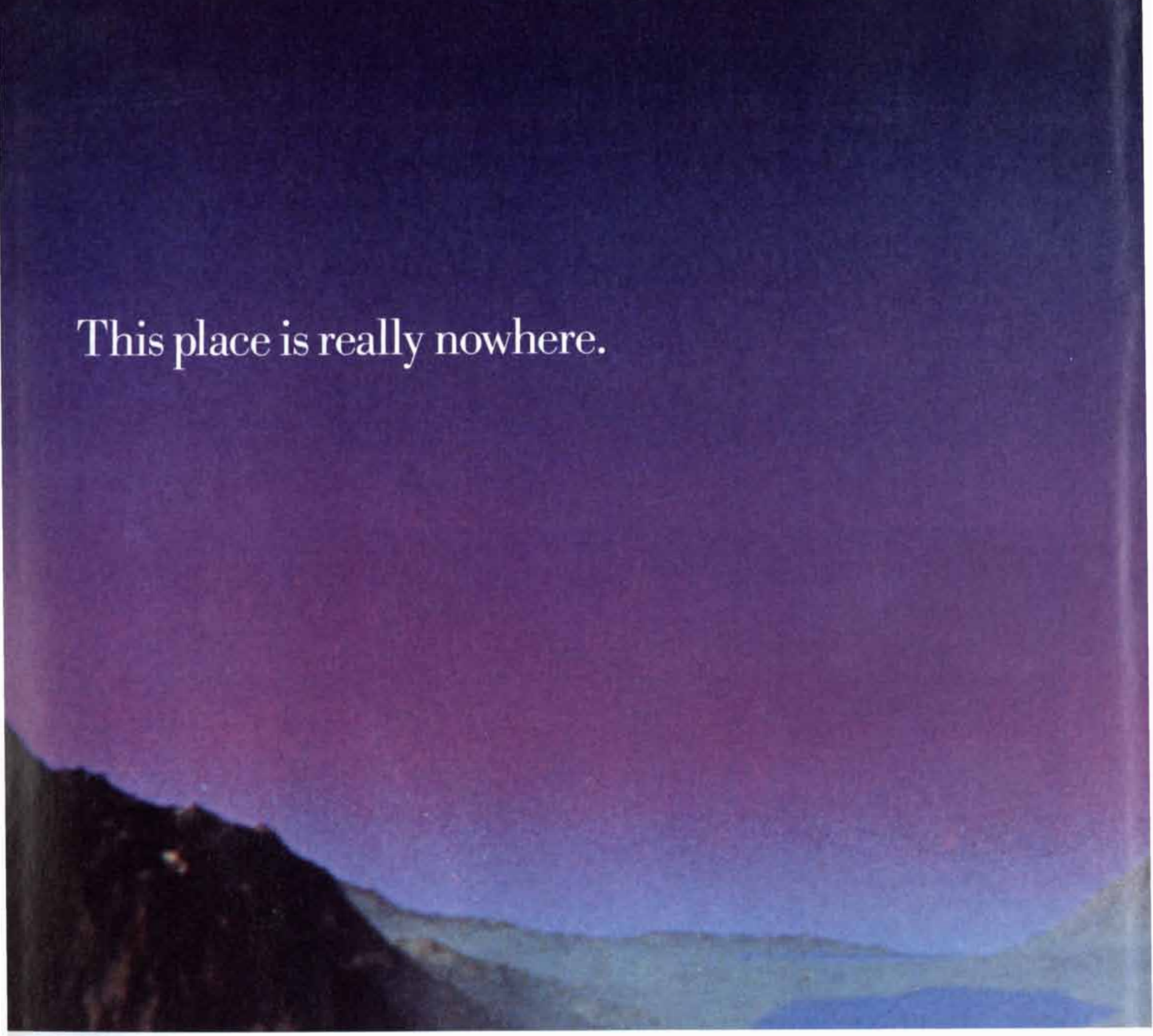
Many Life configurations, including Conway's glider, satisfy the conditions. A Life 5766 glider can be created by converting a two-dimensional glider into cubes. It occupies two adjoining layers of cells in three-dimensional space and is confined to move eternally therein [*see top illustration on page 22*].

Many other forms of Conway's Life such as the beacon, beehive, blinker and block as well as more exotic patterns such as the boat, clock and barber pole, also satisfy the conditions of Bays's theorem. They exist in two-layer form as Life 5766 entities that behave precisely like their two-dimensional counterparts. Unfortunately not all the configurations in Conway's Life satisfy the criteria. Among the black sheep of the family is the famous glider gun discovered by R. William Gosper, Jr., while he was a student at the Massachusetts Institute of Technology in 1970. If one inspects the glider gun at any stage of its life cycle, one finds living cells that have five living neighbors. Any violation of either of the two conditions of Bays's theorem results in a departure from the simulation. "When this happens," reports Bays, "the object, theretofore confined to two planes, almost always forms a roundish, three-dimensional mass that usually dies rather quickly."

Life 5766 can be made to simulate



A glider makes its way through a stable helix in Carter Bays's three-dimensional Life 4555



This mountain exists inside a computer, and can be seen on a computer screen. It was built only with numbers. And it exists only because of one man's idea.

The man is Benoit Mandelbrot, a mathematician and an IBM Fellow. And his idea is known by the name he gave it: fractal geometry.

This geometry is a new, non-traditional area of mathematics. Now scientists and artists can create computer images–like this mountain–that have all the quirks and irregularities of natural objects. That wasn't possible before fractals, and it's changing the way we look at the world.

Conway's Life more completely with the erection of what Bays calls a time-space barrier. This consists of a single sheet of living cells with holes punched in it [*see bottom illustration on page 22*]. The holes are arranged in such a way that each cube in the barrier has exactly seven neighbors. Cubes in the barrier barely avoid death by overcrowding, whereas any cube immediately adjacent to the barrier will have more than six living neighbors and will never come to life. The resulting dead zone on each side of a time-space barrier can be exploited to reproduce Conway's Life exactly. One constructs a sandwich of two parallel time-space barriers a distance of four cubes apart. Of the four intervening planes the central two may support Life in any form: the two conditions of Bays's theorem are now obeyed by any population of living cubes occupying the center of the sandwich.

In spite of its ability to mimic Conway's Life in one form or another, Life 5766 lacks what might be called cellular pizzazz. According to Bays, random primordial soups in Life 5766 always seem to "settle down" faster than those in Life 4555. The unsettled condition of Life 4555 hints at a wider range of computational possibilities. Indeed, Life 4555 has an abundance of stable and oscillating forms that are symmetrical. Perhaps the final decision as to which version will make the worthier partner to Conway's Life will hinge on whether Life 4555 can be made to simulate the two-dimensional game. If it can, it would be superior to Life 5766 in all ways.

Bays has a 40-page document titled "The Game of Three-dimensional Life" that readers may order from him at the Department of Computer Science, University of South Carolina, Columbia, S.C. 29208. The sum of $3 should be sent to cover the costs of production and postage. The monograph describes everything here and much more, including advice on programming the games efficiently. Bays also has made available a Macintosh program that runs both forms of three-dimensional Life.

A program that computes and displays successive generations of either form of three-dimensional Life is simple to write—at least in principle. The program can even be modified to carry out a search of rules in the hope of discovering a third form of three-dimensional Life that Bays may have missed.

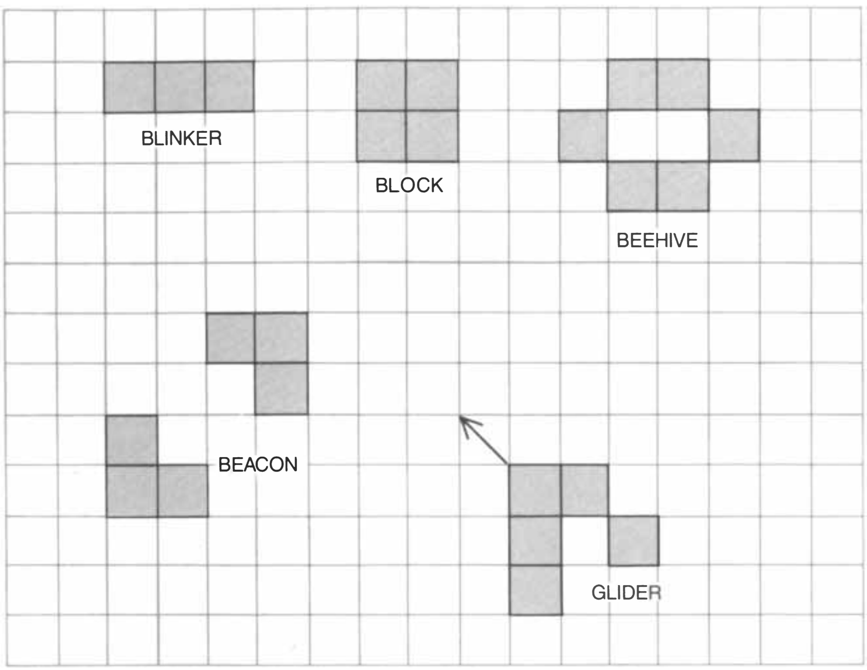
Two large, three-dimensional arrays called *cells* and *newcells* are assigned three indices, *i, j* and *k,* which correspond to the three coordinates of the cellular space. The content of each array element indicates whether the corresponding cell is alive or dead. Let 1 signify life and 0 signify death.

Three nested loops, not surprisingly, are needed to compute the status of each cell in each generation. The outer loop uses the *i* index to compute a succession of planes that sweeps through the space. Within this loop are two others using *j* and *k* respectively. The *j* loop computes successive rows within each plane and the *k* loop computes successive cells in a row. One can use the following generic form as a guide:

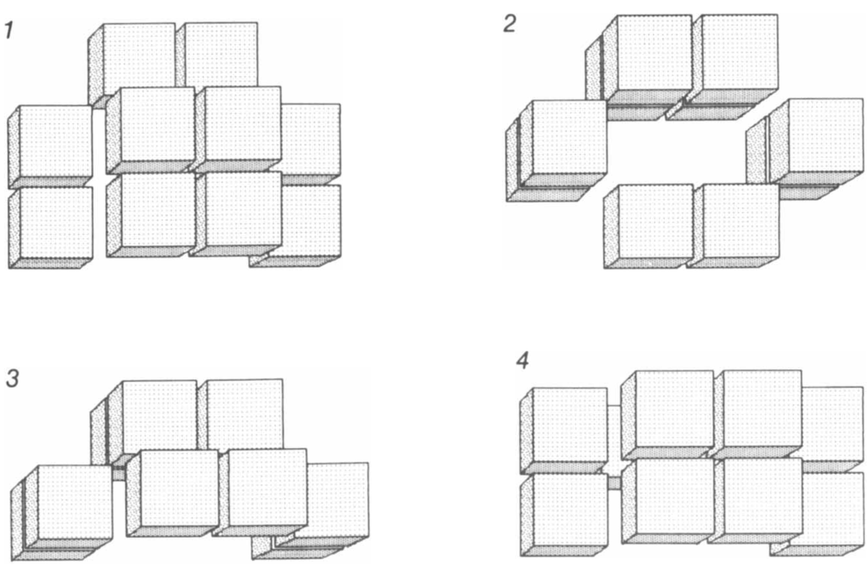
```
for i = 1 to 30
  for j = 1 to 30
    for k = 1 to 30
      compute neighbors
      decide status
      display live cubes
```

The number 30 is arbitary, of course. Only readers with infinite patience are advised to try numbers greater than 30, however, as the calculation time becomes exceedingly long.

Inside the innermost loop there are three basic tasks for the program to do. The task *compute neighbors* requires the program to examine the 26 neighbors of each cell and total the number currently alive. This can be done with three miniature loops or by listing all possible coordinates of the



Some sample forms in John Horton Conway's two-dimensional Life



A glider in Life 4555

26 cells. In loop form the procedure might use the following algorithm:

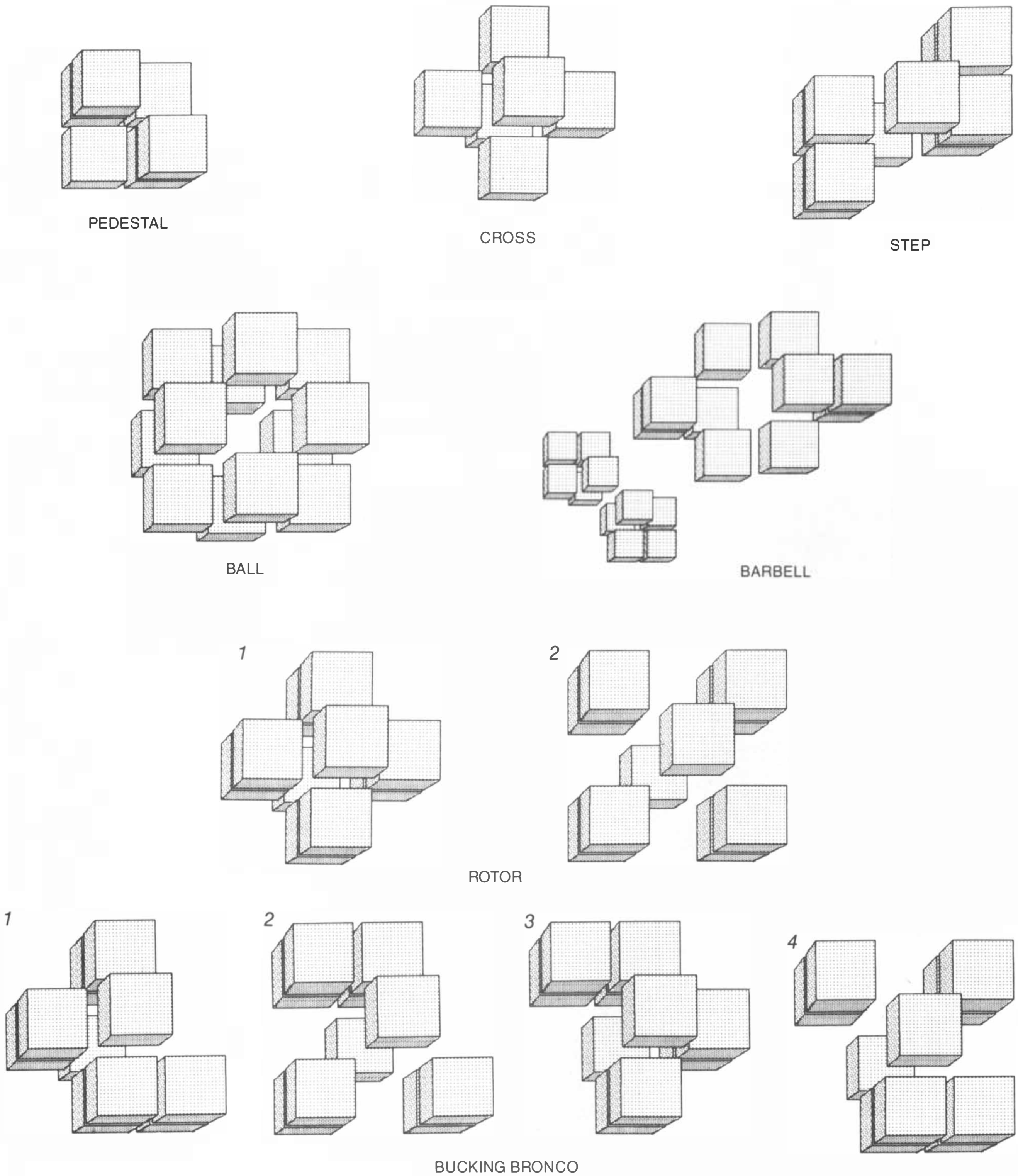
```
tot ← 0
for l = i − 1 to i + 1
  for m = j − 1 to j + 1
    for n = k − 1 to k + 1
      if cells(l,m,n) = 1
        then tot ← tot + 1
tot ← tot − cells(i,j,k)
```

The last line of the procedure ensures that the status of *cells*(i,j,k) does not contribute to the total.

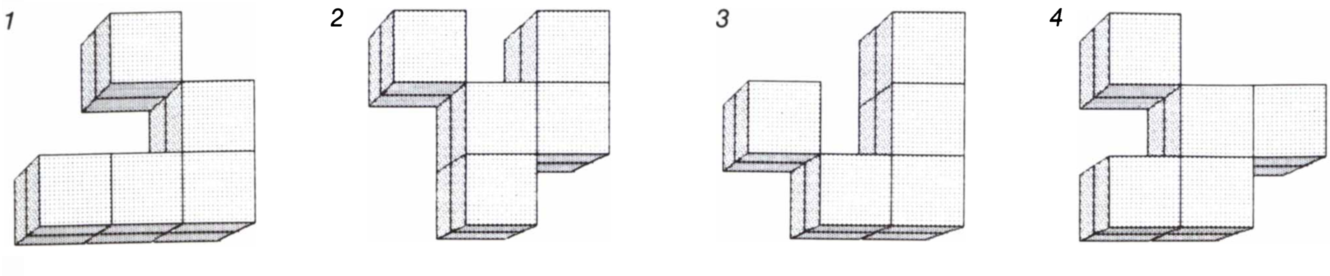
Having decided on the total *tot* of living neighbors, the program must next decide the new status of the current cell, *cells*(i,j,k). The task *decide status* is merely a matter of checking the size of *tot* in relation to the status of *cells*(i,j,k):

```
if cells(i,j,k) = 0
  then if tot = 5
         then newcells(i,j,k) ← 1
         else newcells(i,j,k) ← 0
if cells(i,j,k) = 1
  then if tot < 4 or tot > 5
         then newcells(i,j,k) ← 0
         else newcells(i,j,k) ← 1
```

Here I have assumed that the reader



Stable and cyclic forms in Life 4555



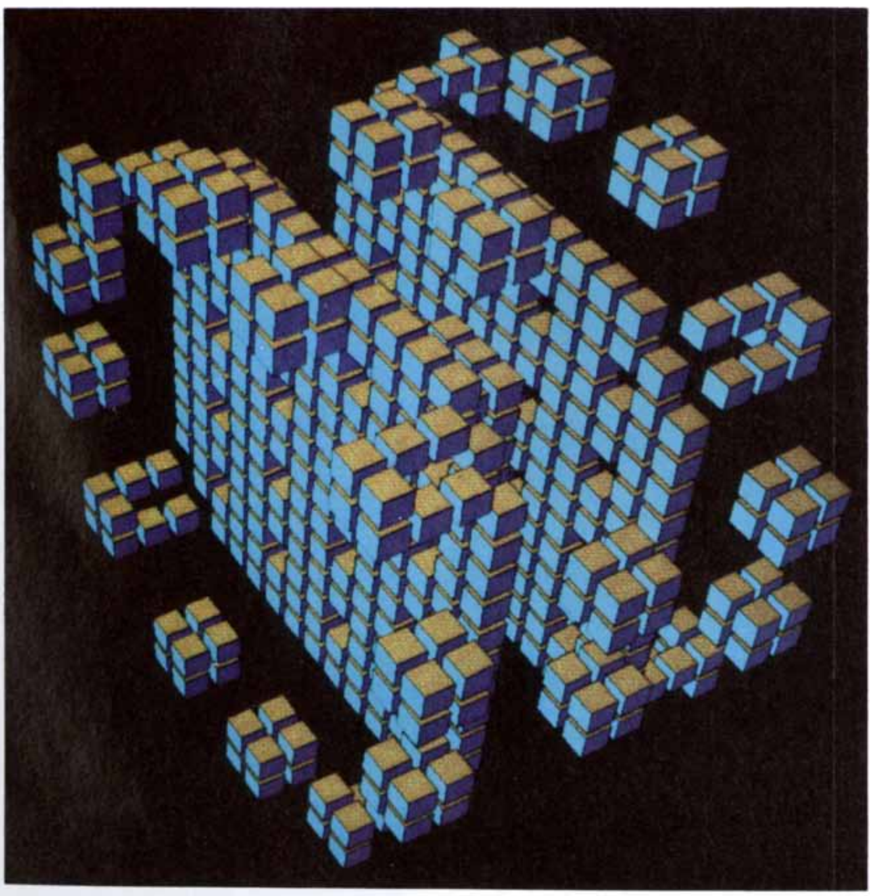
A glider in Bays's three-dimensional Life 5766

is programming Life 4555. One can change the algorithm to fit Life 5766 or make it general enough to manage any three-dimensional rule whatever. It seems worthwhile to digress on this point for a moment.

A general version of the foregoing status computation might use four variables Bays calls *el, eu, fl* and *fu*. The letters *e* and *f* stand for environment and fertility and *l* and *u* for lower and upper. Thus *el* and *eu* are the lower and upper bounds for the continued life of a cell in its environment; the cell will stay alive if the number of living cubes surrounding it is greater than or equal to *el* but less than or equal to *eu*. By the same token, *fl* and *fu* are the conditions of fertility for a dead cell. Its rebirth is guaranteed if the number of living cubes surrounding it is greater than or equal to *fl* but less than or equal to *fu*. The general algorithm is therefore

```
if cells(i,j,k) = 0
  then if tot < fl or tot > fu
        then newcells(i,j,k) ← 0
        else newcells(i,j,k) ← 1
if cells(i,j,k) = 1
  then if tot < el or tot > eu
        then newcells(i,j,k) ← 0
        else newcells(i,j,k) ← 1
```

At this point in either version of a three-dimensional Life program the contents of *newcells* can be moved into *cells* by means of the appropriate triple loop. This frees up *newcells* for the next generation of living cubes.



Two time-space barriers in Life 5766

In the final stage of the computation process, *display live cubes,* the program displays the particular cube—if it is living. When the actual cubes are drawn, it is advisable to fill in the visible surfaces. If just the bare skeleton frames are used, a very cluttered, well-nigh indecipherable scene results. The simplest way to ensure that cubes in front properly obscure those behind is to make sure that *i,* the outer index, sweeps from the back of the cellular space toward the front in relation to the viewer. The sad fact that some cubes must be obscured in this way points up the only disadvantage inherent in three-dimensional Life in any form: we cannot command the sweeping view of all that goes on as we can in Conway's two-dimensional version. At the same time, the disadvantage of any three-dimensional game is shared by us in our real, three-dimensional world. We cannot see everything that is going on, fortunately enough.

When the final stage of the computational process is embodied in a program, it will tend to be rather slow. One simplification that may speed matters up somewhat is to replace the cubes by spheres (actually filled-in disks whose size varies with their "depth" in the screen). For the rest, Bays's monograph referred to earlier has many orders-of-magnitude improvements in speed to suggest.

As usual I shall be happy to report unusual or interesting phenomena uncovered by readers. First, for those intrepid explorers who demand the utmost in generality, are there any forms of three-dimensional Life (worthy of the title) that Bays may have missed? Next, readers may also enjoy setting up their own primordial soups. For the rest, one wonders what undiscovered gliders, spaceships, glider guns and other configurations there are. To avoid duplicating what Bays already knows, access to "The Game of Three-dimensional Life" is mandatory.

The game I identified as Star Trek in the November column existed, in spirit at least, long before the popular television series by the same name. Robert Leonard Nelson, Jr., of Austin,

Tex., remembers that a game called Spacewar!—the grandfather of all the shoot-em-up spaceship games—was written in the early 1960's for the PDP-1 computer at the Massachusetts Institute of Technology. The authors of the game were M.I.T. fellows and graduate students who were proud to call themselves hackers, according to an article in the August 1981 issue of *Creative Computing* [see "Bibliography," page 140].

The PDP-1 was inconvenient to program, of course, and input parameters were set from console switches. J. Martin Graetz, a member of the group, describes the atmosphere motivating the rapid developments that nonetheless ensued: "When hackers are aroused, anything that can happen will.... Spacewar! worked perfectly well from the test word switches on the console, except that the CRT was off to one side, so that one player had a visual advantage. More to the point, with two excitable space warriors jammed into a space meant for one reasonably calm operator, damage to the equipment was a constant threat."

Kenneth L. VanEseltine of Schoolcraft, Mich., has honed the Spacewar! program I described in November. At first he called it STAR TRUCK because of its slowness. It only became STAR TREK when he found ways to move most of the floating-point arithmetic out of the main loop. He has some tips to share with fellow readers. To avoid computing square roots, he precomputed a two-dimensional array of distances for each possible point (x,y) from the origin (0,0). The distance between two arbitrary points (x_1,y_1) and (x_2,y_2) is then calculated by looking up the distance for $(x_2 - x_1, y_2 - y_1)$ in the table. VanEseltine also tabulated sines and cosines in two arrays of 36 entries each, one for each 10 degrees. He also found that missiles are easier to manage if one always has 20. At any given time, of course, many of them may be virtual, in which case they simply are not plotted. Moreover, the velocity arrays should be not two-dimensional but rather paired one-dimensional arrays. On many computers this simple redefinition alone can cut overhead time considerably. Determined readers will undoubtedly rediscover many of the other improvements in speed that VanEseltine found.

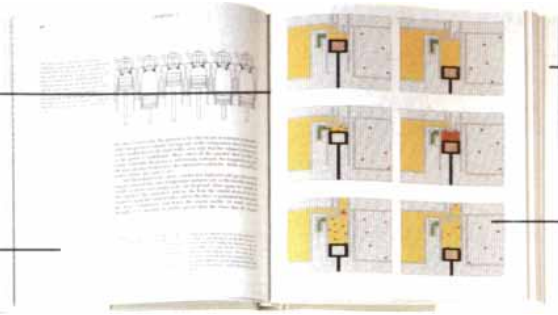
Finally, for those readers who simply want to sample the space-warrior experience, Bill Seiler of Scotts Valley, Calif., has a program called SPACEWAR that runs on IBM PC's and a number of other machines. Seiler is willing to make his program disk available; write to him at 317 Lockewood Lane, Scotts Valley, Calif. 95066 for details.

Now in its third year, SCIENTIFIC AMERICAN LIBRARY has won recognition as an exciting new genre in the literature of science. Created out of the collaboration of scientists and editors that distinguishes SCIENTIFIC AMERICAN itself, each volume in the LIBRARY:

☐ Covers a topic of major interest with immediate relevance to ongoing research in its own and related fields.
☐ Carries the authority of an author who has made a significant personal contribution to its subject.
☐ Proves there is no topic in any field of science that cannot be explained to readers who want to understand.

Impeccably printed on high quality paper. Heavy cloth covers and binding to insure years of trouble-free reference.

Ample margins for visual relief and notes.



Large, inviting format (8½" x 9¼") opens flat for convenient reading.

Handsomely illustrated with many full-color charts, diagrams, maps and photographs.

JOIN OUR LIBRARY. THERE IS NO RISK INVOLVED.

A Two-Month Cycle.
As a library member you will receive a new volume approximately every two months. (Your first volume, EINSTEIN'S LEGACY, will be shipped UPS for fastest delivery.)

15-Day Free Examination.
Examine each new volume at your leisure. If you decide within 15 days not to own a particular volume, simply return it, and your account will be promptly credited.

No Prepayment Necessary.
Pay only for the volumes you decide to keep.

No Minimum Purchase.
You may choose to keep as many or as few volumes as you like. There is no yearly minimum.

Low Members' Price.
Members may purchase any volume at the exclusive low price of $24.95—well below the bookstore price.

Your Guarantee.
You may cancel your membership at any time without notice, for any reason, without penalty.

MEMBERSHIP RESERVATION

Free with your paid order:
TELESCOPING POCKET TRAVEL ALARM.

Its sleek black case slides open to reveal a 7-function alarm and calendar clock with a dependable quartz movement. Comes with mini-support leg for your viewing convenience.

Please speed my premier volume—EINSTEIN'S LEGACY—on its way via UPS and enroll me as a member of the Scientific American Library. Thereafter, send me a volume approximately every two months for 15 days' free examination.

I need send no payment now. For each volume I keep, I will be billed $24.95 plus postage and handling. This price is guaranteed for at least one full year. I understand that I may cancel my membership at any time, for any reason, and without penalty.

Yes, I'm looking forward to receiving my LCD Travel alarm after you receive payment for my first volume.

Signature ______________________

Name ______________________

Address ______________________

City ______________ State ________ Zip ________

287PAP

SCIENTIFIC AMERICAN LIBRARY

Mail this form to: SCIENTIFIC AMERICAN LIBRARY, P.O. BOX 646 HOLMES, PA 19043.
Or, to join right away call toll-free 1-800-345-8112 (1-800-662-2444 in Pennsylvania).

BOOKS

Murals of the Maya, cancer's time and place, glass pleasure domes and the diversity of life

by Philip Morrison

THE MURALS OF BONAMPAK, by Mary Ellen Miller. Princeton University Press ($67.50). In May of 1946, while he was working in the field, the filmmaker Giles Healy was given a precious gift by his Lacandon Indian hosts. They took him to a nameless and all but inaccessible ruin—they called it by their word for a pile of stones—deep in the steaming rain forest along Mexico's border with Guatemala, south of the dry, scrubby plains of Yucatán. Healy was the first outsider to view in astonishment the richly painted walls that are the only sustained passage of Classic Maya mural art we have.

A third of the way up the stepped platform that is more than 100 feet high there is a masonry structure divided into three adjoining rooms of modest size. On the interior walls and roof vaults of those rooms splendid artists painted a wealth of coherent and colorful images. There are hundreds of costumed human figures, about two-thirds life size; incomplete inscriptions suggest a probable date within a few years of A.D. 790.

Heavy rains have long infiltrated cracks to dissolve away a little of the vaulting stone, for the structure was poorly built. The picture surfaces below have been calcified by the slow drip; that mineral growth, as in a cave, has preserved the paintings, and now it veils them from view. Today they can be made out only by soaking the whitened surfaces with water or kerosene. Casual visitors expecting large bright pictures are sure of disappointment. Nevertheless, the provisions made over the years for tourists, the tin roofing and clearing of the forest nearby, have set the paintings suddenly into a dry, sunlit climate after 1,000 years of solitude in the cool, damp jungle. Fungus is overspreading the surfaces and "the paintings are slowly flaking from the walls."

"The path to Bonampak was full of snares." Even its name means something closer to dyed walls than to painted ones, for the senior archaeologist who chose the name was no master of Mayan. The circumstances of the discovery and the epoch in which it happened plunged the site into notoriety and conflict; both were heightened by the personalities and fates of the discoverer and his companions, as well as the polarization that arose between national pride and the active presence of well-heeled *norteamericano* archaeologists. Healy was making his film for the United Fruit Company; his traveling companion, Carlos Frey, who months earlier had missed finding the ruin himself by only a few feet, quickly gained lively support from the eminent muralists Diego Rivera and David Siquieros. Frey was drowned during the expedition he soon organized to revisit the site.

The furor has smoldered for decades, leading to tabloid hyperbole, scandals, romances, even a ballet. In 1981 a new tale of Bonampak came out: Frey and his Lacandon mistress were novelized in "a spicy tale of sex and passion in the jungle." Intense public interest has led to diffuse responsibility for preservation. By the time the paintings are ruined, our author concludes wryly, it will be no one's fault.

Professor Miller is a young Mayanist, now an art historian at Yale. Her volume is a model of the meticulous confrontation of evidence. It is not easy for the general reader to follow in all detail her learned figure-by-figure and glyph-by-glyph reading, but it is fascinating to inch our way around the rooms at Bonampak, helped by her keen eye.

Healy, the cinematographer who first reported the site, made extensive photographs early on; instead of systematically trying to record it all, he worked mainly in one room. The first and only full photographic record was made in 1964 under the auspices of Mexico's famous National Museum of Anthropology. The images included both large black-and-white negatives and color transparencies. That color film has now disappeared from the archives. Many photographers have since worked at Bonampak, but they understandably record the best-preserved areas; no other comprehensive coverage exists.

It is the archaeological copyists, painting with care and insight, who provide most of our evidence. They have made four copies of the murals. The first two copyists worked at one-quarter scale during two rushed seasons, using a grid they penciled on the walls; the third traced her large outlines on acetate panels placed directly over the painted surfaces; the fourth and fifth, working as a team, reproduced on a reduced scale only the first decorated room. They worked under commission from the Florida State Museum in Gainesville, where their brilliant copy is now displayed on the walls of a matching room. They were required to reconstruct all damaged areas; that entails some comparisons before reliance on detail. Those two most recent copyists, Felipe Dávalos and Kees Grootenboerg, "certainly show greater skill than any other archaeological draftsmen who have worked on this project."

The reader is offered as data an indispensable labeled outline sketch of all the painted walls (the captions omit the designations of direction given in the text) along with maps, plans and scores of images of relevant details from this site and others. The color plates reproduce in full the quarter-scale copy of 1946, perforce reduced by another order of magnitude, so that the human figures we see on the page are only an inch high. Finally we have good color photographs, again small in scale, of the copyists' painted room in Gainesville.

The colors may have a symbolic meaning, but they are broadly natural: red-brown flesh tones, green plants, red blood, yellow jaguar skins with black spots. The paintings display ordered and vivid processions, musicians, combat, ceremony and figures—armed, costumed and masked. The details are rich, complex, enigmatic. Glyphs stud the paintings: ancient labels and captions. There are even a few inconspicuous but well-drawn line images out of context, interpreted as signatures or graffiti!

The outcome of the study is marvelously revisionist. The most distinguished Mayanists of a generation ago—who could not read glyphs—saw the ordered murals as presentations beyond history, ceremonies of a timeless theocracy. The combat and the bloodletting from naked captives, rather clearly visible, were viewed as carefully staged ritual, even symbolic. "As the Maya are not by nature sadists," wrote one of the authors of the definitive Carnegie Institution of Washington report on Bonampak in 1955, "I do not think torture of the captives is depicted."

The glyphs are not without ambiguity, but the present readings seem to identify the events of the paintings as the real actions of the ruler of Bonampak, by name Chaan-muan. The first scenes bear an important astrological date, the coincident arrival of the sun overhead at Bonampak and a conjunction of the planet Venus. On the auspiciously omened day the ruler presented an infant in arms as heir to rulership in a ceremonial pageant, resplendent in costume and alive with music.

The second room sounds a terrifyingly distinct note. The presentation would be followed by the need for ritual victory. The paintings in this room begin with armed combat and move on to public torture of despoiled captives, abject before the victors. Many details support this harsher interpretation, although the reader cannot be sure of all that is pointed out. That a prisoner, Human Figure 102, who is certainly pitiable, "opens his mouth in a howl of pain" is not evident from the photograph.

The third room is the least well preserved; the surfaces have gaps and there are no helpful glyphs. Captions are incomplete; a background was prepared but glyphs were not drawn. Human sacrifice is under way at a large pyramid. A wonderful scene of dance or mime includes figures bearing light wings for costume. Monstrous headdresses are worn; there are bearers of parasols and of standards.

On an adjoining wall the ruling family is presented on a large throne. The ruler himself seems to be identifiable. His figure was interpreted decades ago as an "old lady eating with great delicacy and refinement." Miller reads the same act as penitential bloodletting, done by scarifying the tongue with sharp spines. The other guest lords are drawn here as well, each with a hand similarly poised at the mouth. Cartouches of a deity to whom blood sacrifice was dear surmount the entire painting. Counterpart scenes at other sites and later data from other cultures support the grim interpretation.

"These are powerful scenes of violence and beauty. The sensuous quality of line...has evoked the greatest admiration from the twentieth century at the same time that the torture...has evoked the greatest horror." The finest painters of the land had been gathered to record the solemn events, in a conservative style. The Bonampak murals were never finished. The dynasty came to an end. It is unlikely that the little heir ever reached the throne. "The site was abandoned; the artists were dispersed." Mesoamerican painting never again achieved such quality.

Is Bonampak a chance record demonstrating the end of the Classic Maya by aggressive war among the cities? There is a puzzle still, even if the elaborate unriddling of painted conflict is as fully accurate as it is plausible. The cities of the Classic Maya came to an end, but they show no marks of battle. Combat was not in the city at all but in the countryside. Mayan cities were not fortified. As far back as John Lloyd Stephens the explorers, with Europe in mind, had felt the pacific nature of unwalled cities whose great sculptured figures bore no arms.

In the time of the Toltec influence over post-Classic Yucatán, war became genuine. Earlier in Bonampak neither the battle nor the cruel muster of prisoners is painted with many people engaged; 100 fighters at most are represented. Is this a painterly convention or is it merely a hint that the battles were scheduled as part of a ritualized taking of prisoners in a ceremonial war?

The view today, shared by Miller, is that destructive battling had already occurred among the Maya of the forest centuries before anyone arrived from distant Mexico to teach grim war. But the cities of stone were not so much centers of economic life or population as they were sacramental places, and therefore those imposing pyramids, courts and galleries rarely became targets for military action; they died by abandonment rather than by destruction.

GLOBAL GEOCANCEROLOGY: A WORLD GEOGRAPHY OF HUMAN CANCERS, edited by G. Melvyn Howe. Churchill Livingstone ($127.50). The count of the quick and the dead is the first datum of vital statistics. By now a great many countries do it well; the census taker and the official death certificate are widespread. Age too is reasonably sampled; the epidemiologists have even boldly prepared a model age structure for the world population, tabulated for 1982 on the last page of this expert and cosmopolitan compilation of grim but tantalizing riddles. By demographers' consensus half of us human beings are over the age of 26, although only one out of 25 lives to a full three score and 10.

The causes of death? Those are far less securely known. With one exception: for a generation a special effort has been made to establish a critical yet far-flung tally of illness and death from cancer, uniformly reported year after year to international bodies and

Depiction of a musical procession: a copy of a mural in Room 1 at Bonampak

ELDOR

The Driving Sp

The quality is readily apparent. The comfort and convenience are pure Cadillac. Yet the 1987 Eldorado delivers an added dimension, the exhilaration of commanding a luxury car that loves the road. As you settle comfortably behind the wheel, you note that everything is close at hand, ready for you to take control. Cadillac convenience features let you know you're driving a special personal coupe. A coupe backed by a new longer limited warranty.

This Cadillac takes your driving pleasure very seriously.

The power of a transverse-mounted, fuel-injected V8 with the traction of front-wheel drive. The controlled ride of four-wheel independent suspension with Electronic Level Control. The assurance of four-wheel disc brakes. And the response of power-assisted rack-and-pinion steering. All of these make Eldorado a true driver's luxury car.

A personal environment designed around you.

Adjust the leather-wrapped tilt-and-telescope steering wheel. Position the six-way power driver seat. Set the power rearview mirrors. This Cadillac is designed to accommodate your individuality while it encourages and rewards driver involvement.

Cadillac is a proud sponsor in the quest to bring back the America's Cup.

'ADO
t of Cadillac.

Technology that enhances your control.

Eldorado acknowledges that information translates into driver control and confidence. Within its available American walnut instrument panel, an advanced multi-function Driver Information Center monitors everything from engine rpm to the outside temperature.

The confidence you feel behind the wheel is also aided by Eldorado's sleek aerodynamic styling, which contributes to remarkable road stability.

Technology that enhances your convenience.

Of course, Eldorado has all the comfort and conveniences that make a Cadillac a Cadillac. But this Eldorado uses leading-edge technology to further enhance these Cadillac hallmarks. Witness the sun-load sensor that measures the amount of sunlight entering the passenger compartment and then automatically instructs the Electronic Climate Control to adjust itself.

Eldorado's new 5-year/50,000-mile limited warranty.

Another word for a company's confidence in its product's quality is warranty. And another word for quality is Cadillac. Eldorado is backed by a new longer 5-year/50,000-mile limited warranty.*

Your Cadillac dealer invites you to experience the joy of the open road in Eldorado, the driving spirit of Cadillac.

*In some cases, a deductible applies. See your dealer for details.

LET'S GET IT TOGETHER...BUCKLE UP. GM

by now published in four or five big volumes of tables. At present there are 84 cancer registries, from El Paso to Kampala and Singapore, "unevenly distributed throughout the world" in 32 countries over all continents. China is the grand near-omission.

At each registry legally mandated reports of cancer diagnosis, treatment and death are received and painstakingly sifted—on the average the particulars of hundreds of cancer deaths a week. Was there at least nonmedical testimony of death? Is there the report of any physician? What do entries mean when they are outside the standard vocabulary or are unsupported by pathological inspection?

In this book we learn about the current yield of those siftings, country by country, in thoughtful, even reflective summaries by about 30 scholars, supported by thematic maps, figures and tables of trends and correlations. Expert testimony rather than basic data, the long string of articles is circumstantial enough. Certainly this insulation from the data themselves is essential for a general reader, although there echoes still a steady note of uncertainty, put persuasively by one contributor. "After spending many years with the Estonian Cancer Registry the writer now thinks more intensively about data quality than about the application of refined statistical and cartographic methods to data analysis." One comes away pretty sure that these age-standardized registry reports are broadly comparable from place to place and to a degree from time to time. It is the standardization for age that makes cancer counts comparable; the disease is so often slow to appear that for a long time it seemed somehow lacking among the youthful populations of developing countries.

The results are certainly strange. Consider cancers by site of the tumor. The standardized incidence of respiratory-tract cancer is about 100 times as high among black men in New Orleans and Maori males in New Zealand as it is among the men of Dakar (also black) and a dozen times as high as it is among American Indian males in New Mexico. About 100 recognizable groups rank between those extremes. Genetic factors may have a speaking part but hardly a star role: one overall estimate puts the genetic component at about 20 percent. New Orleans whites rank very high on the list and New Mexican Hispanics quite low. Urban and rural differences seem real, the poor suffer double the toll of the rich, and the "causal relationship between cigarette smoking and lung cancer is now reasonably well established." Yet the Yukon, which is certainly pretty rural, ranks high here and teeming Bombay ranks very low.

Prostate cancers claim black males in California at a rate again two orders of magnitude greater than that prevailing in Shanghai. Elderly Swedes and Swiss suffer considerably; East Asians in several nations, Senegalese and Czechs are better off by a factor of five. Men of Japanese stock in Hawaii rank high. Such anomalies are general. There must be many causes, much entangled, the manifold permutations of various features of the physical and cultural environment.

Consider trends with time. Japan and Chile have always reported an extremely high incidence of gastric cancers. It is the commonest form in both countries. Yet in both of them the rate has been decreasing steadily for the past three decades, an experience confirmed in many countries. The decrease in Chile is confined to the big cities; it is plausible that in rural places where the cause of death is not medically determined other abdominal sites load the apparently unchanged rate. In Japan a mighty effort has gone into finding the correlates of the disorder. Smoking and its synergies with occupation are surely important; so is diet. Salty food is overwhelmingly indicated as a promoter of the disease; the decrease over time may reflect the steady rise of the preservation of foods by electric refrigeration.

In Chile a regional relation between the natural nitrates of the desert environment and cancer has long been suspected, but at present that hypothesis remains moot. Geographic clustering of gastric tumors in Estonia led to the notion that oil-shale deposits were important environmental contributors, a proposal that proved wrong.

In Iran there is a well-studied concentration of cancers of the esophagus along the Caspian Sea, part of a vast zone of high incidence that extends across Asia from the Middle East to China. Opium seems to be exonerated: many opium-using regions did not share the effect. There is a high correlation between esophageal cancer and the consumption of homemade bread—and carcinogenic silica crystals have been recovered in the stone-ground flour that is baked into the staple bread diet of the region. Physicians recall also the chronic inflammation of the esophagus usual in this place of drinkers of fiercely hot tea, whose diet is vitamin-poor. So far this is only a promisingly intricate hypothesis.

One column of graphics in the report of the United Kingdom is striking. It is a set of five pairs of small diagrams; one of each pair pertains to men, the other to women. The five display data decade by decade from 1931 through 1981 (with 1941 omitted). In each square a set of bold contours divides the overall death rate by percentage into eight or nine causes of death, plotted by age groups. The result is an eloquent visual history of medicine. First of all the black shadow of infection goes away. A mountain in 1931, it is by 1981 a minor rise among young children. Maternal deaths among women and digestive deaths in general became unplottably small before 1960. But the cancer slope that in 1931 first came into view at age 40 is now a great double peak, high for children and important again after the age of 30. Accidents and violence kill young adults more than any other cause, where once infection was dominant. A wartime year would have been interesting to see.

HOUSES OF GLASS: A NINETEENTH CENTURY BUILDING TYPE, by Georg Kohlmeier and Barna von Sartory, translated by John C. Harvey. The MIT Press ($65). Those faceted and shimmering pleasure domes in many cities display still to our eyes the power of 19th-century Europe, the first society able to lavish glass, iron and steam-raising coal on the cultivation of tropical plants from the far colonial world. A winter garden had magically become real for northern lands; the tall palms and giant water lilies romantically bespoke and then outdid flowering nature, which otherwise notoriously receded farther and farther from city life. The experience became as available as a daydream to a public able to pay the price of admission to 100 filigreed palaces of glass, catalogued here from Belfast to Lisbon to Leningrad.

It was the princes who first enjoyed the company of trees growing under glass. The noble Congress of Vienna once dined, albeit without room to dance, on tables set along the 600-foot arched gallery of the orangery at the Palace of Schönbrunn. "One saw nothing but the flowering trees and plants, between them statues and fourfold waterfall streaming over rocks, lit by 3,136 lights." The royal orangery was a preindustrial precursor; amply glazed with expensive windows along its south wall, its climate was hardly tropical. The fabric was still of masonry and wood, the roof was opaque, and the scale, like the use of the structure, was not yet a public one.

A typical public winter garden in one of the bigger cities soon held under lightly framed glass a sunny area larger than any orangery in Europe by a factor of 10, and its ceiling was proportionately higher. The structures were built with modular iron space frames in novel forms; they had transparent roofs and walls and were heated by coal-fueled stoves tended from outside, or later by steam boilers that fed radiators, all bright after dark under generous gaslight.

The attraction of the building type reflects important perceptions of the time, touching on both aesthetics and politics—not only the rise of Romanticism but also the growth of social utopias promising release from the thrall of the dark Satanic mills. Charles Fourier, Robert Owen and Sir Ebenezer Howard, visionaries all, saw the civic virtue of the future enclosed in one way or another in walls of clear glass. A more immediate impetus was the lucrative market for entertainment in those wildly growing, gloomy, wintry cities whose inhabitants enjoyed plenty of cash but little central heating. A London designer of just such warm, green, bright spaces wrote in 1840 of a Berlin winter garden: "In the early morning one finds there gentlemen reading the newspapers, drinking chocolate, and discussing politics. After three o'clock one sees...people of every sort sitting among the trees, conversing, smoking, with punch, grog, coffee, beer, or wine." The why of this development seems all too easy to understand in our time of trendy atrium and enclosed mall.

How it came about is most interesting. The gifted Joseph Paxton, whose extraordinary modular Crystal Palace of 1851 (some 21 acres under glass) was the largest of them all, designed the "basic structural and spatial concept...in a few hours—during a committee meeting." Sir Joseph drew on decades of experience that began with simple forcing houses, which he had built of glass on a light wood frame when he first came to work as a young gardener for the Duke of Devonshire, one of the wealthiest peers in England. His masterpiece there was a greenhouse whose span rivaled the famous railway-station concourses of the day.

Paxton's Great Conservatory at Chatsworth was "a singularity in the history of building," a "crystalline mountain" everywhere glazed with four-foot panes of glass, so modular in its design that no panes needed to be cut for installation. Its ridged curvilinear form followed an aesthetic that owed nothing to tradition and everything to the intrinsic logic of its structure. It lasted for 80 years, or until the cost of maintenance (eight hungry boilers and seven miles of steam pipes) was held to be too high. Five attempts with dynamite were needed to bring down the wonderful structure with its cast-iron columns and wood ribs.

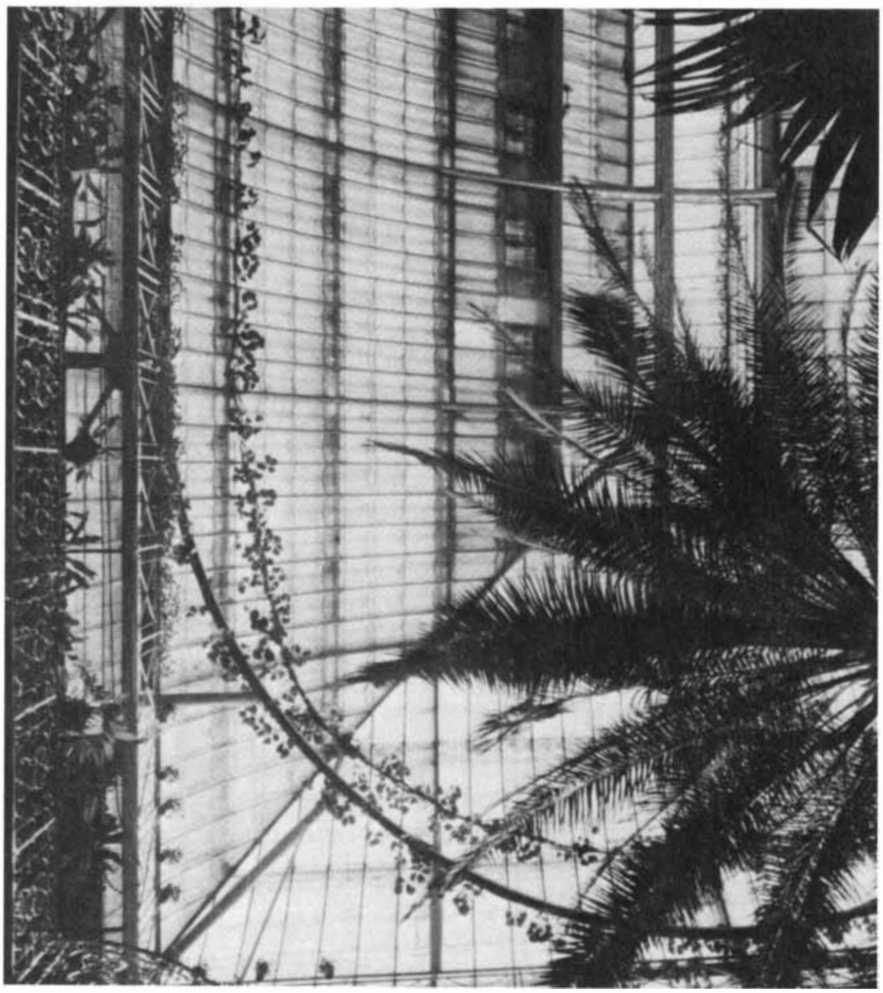
Such buildings cried for mass production of parts, and even for prefabrication. Paxton himself made steam-driven saws and routers to turn out standardized wood sash bars by the tens of miles for his big conservatory. Cast iron and then rolled wrought iron components became standard in due course. The firm of Robert Lucas Chance made 300,000 identical panes of glass to Paxton's order for his Crystal Palace, a third of Britain's annual glass production at the time.

A digression by these delighted authors tabulates the ratio of thickness to span for a variety of famous enclosures. The dome of St. Peter's is measured at one to 13; the finest examples of Victorian houses of glass improve on such classic masonry performance by twentyfold. In our day the best shell roof has outdone the glass houses by another factor of five. (For comparison, an eggshell has a thickness-to-span ratio of one to 100.)

This attractive volume gives over three-fourths of its big pages to the gleaming structures themselves. The alphabetized catalogue, with its histories, architectural drawings and period illustrations, is a learned and yet easily read delight. There follows a 200-page album of photographs old and new. The Great Palm House at Berlin-Dahlem is a late-Edwardian example with remarkable structural novelties; it was rebuilt 20 years ago with large plastic sheets. The refit was economical, but the delightful texture of the old sash-bar filigree unfortunately vanished.

Look then for the winter garden built in the 1840's by an Englishman at Liechtenstein Castle, using the pioneering curvilinear system of John Claudius Loudon, great forerunner of Paxton's larger works. Its elegance and fine texture are appealing still, even to the iron columns cast in the form of jointed bamboo. It is also easy to envy the citizens of Glasgow, who have so long enjoyed their Kibble Palace, a half-gift, half-investment by its thrifty designer. The transparency and purity of its large main dome should easily draw the traveler's eye away from the pensive white sculptures and curly cast-iron ornaments with which it was overendowed as an Art Palace in 1872.

Not even the Gothic cathedral can



The central pavilion of the Great Palm House in Park Schönbrunn, Vienna

Streamline your DP operations

with a Tandy® 3000 HD XENIX® system.

Break away from your mainframe with the powerful Tandy 3000 HD. The 3000 HD supports the versatile XENIX System V (based on UNIX® System V, the standard of the UNIX world). XENIX offers extras like a "C-shell" programming environment, a menu-driven help system and support for Tandy peripherals. The high-performance Tandy 3000 HD makes systems development more efficient and cost effective.

The Tandy 3000 HD (25-4011) is compatible with the IBM PC/AT® and offers greater hard disk storage (40 megabytes vs. the PC/AT'S 30). Yet the 3000 HD is priced at only $4299 (vs. $5295 for the PC/AT*).

Based on the 16-bit Intel 80286 microprocessor, the 3000 HD operates at 8 megahertz. And since it's a multiuser system, people throughout your office can simultaneously access it from inexpensive data terminals—for program development, as well as user applications. The Tandy 3000 HD is your key to total office automation.

Stop by your local Radio Shack Computer Center today . . . we're ready to talk business.

Send me a Tandy 3000 brochure.

Mail To: Radio Shack, Dept. 87-A-501
300 One Tandy Center, Fort Worth, TX 76102

Name ______________________
Company ______________________
Address ______________________
City ______________ State ________
ZIP __________ Phone __________

Available at over 1200 Radio Shack Computer Centers and at participating Radio Shack stores and dealers.

TANDY COMPUTERS: In Business . . . for Business™

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

*Based on IBM price list as of April 2, 1986. Tandy 3000 HD price applies at Radio Shack Computer Centers and participating stores and dealers. Monitor, data terminals and XENIX sold separately. XENIX/Reg. TM Microsoft Corp. UNIX/Reg. TM AT&T. IBM PC/AT/Reg. TM International Business Machines Corp.

satisfy every architectural need, and the houses of glass are even more limited in what they offer. Yet those hard-surfaced gems, so free in form, so expressive in function, declare the industrial age, to which as builders we still belong in spite of easy catchwords. Le Corbusier caught a single glint from the Crystal Palace and Chartres: "The physical history of architecture shows ...an untiring fight on the side of light versus the obstacle imposed by the law of gravity."

ILLUMINATIONS: A BESTIARY, by Rosamund Wolff Purcell and Stephen Jay Gould. W. W. Norton & Company ($35). "The only overarching theme of a bestiary is diversity." That high theme is doubly pursued here: we follow it the first time through eloquent still-life images by Rosamund Purcell, who fills out a most generous alphabet, some 70 photographs of specimens from Albatross to Zozymus, the diversity of 600 million years of evolution expressed in organic form. The second diversity we recognize is that of human culture: this artist's glowing work has been complemented picture by picture by a pointed essay-caption out of science, the work of paleontologist Stephen Gould, some 40 half-page pieces nicely freed of jargon, pointed, personal and as illuminating as the joint title promises.

No forest or field appears; not one of these photographs is taken from life. The lighting is natural, like all the animals here, but it is conceded that nature is here only in part. These are animal fragments, forms of emphasis, deletion or wear, all drawn from the storehouses of specimens that crowd the private research wings of the big old museums: the dry bones of Minerva's owl posed against the skeletal pages of a ruined book, the viper bottled fangless in alcohol, the nail-pierced cluster of huge Goliath beetles, the striped pelt of a Grevy's zebra, the eyespots of pinned moths, the molars of a rhinoceros in the shape of the Greek letter pi, the vascular labyrinths of prepared squirrel monkeys. Such is the rich taxonomic ore long mined and refined into our evolutionary biology.

For a museum specimen, death makes not an end but a beginning. Life is only the first layer to be dissected away among many that may then be removed. That perception justified this discerning artist errant as she intruded with her camera and tripod into the hushed aisles reserved for cabinets, calipers and learned curators.

The unity of the disciplines crystallizes out slowly until, with the pattern of a fossil shrimp in the fine-grained Solnhofen limestone, its branched intricacy retained ever since the Jurassic, "art and science meet...in the most palpable way." That smooth stone was once an ooze so fine that it could harden without damage to the most delicate tissues. The waters there were at the same time so deep and still that they lacked the oxygen that promotes organic decay, for decay is an active process, not merely the spontaneous work of time. Scientists found in that stone the most prized of fossil evidence, while for artists the same rock became the medium of "the world's great lithographs, including...the best illustrations of organisms."

Any enduring collection, however well ordered, itself comes to evoke the chanciness of fossilization. One superbly lighted and harmonious composition shows the deliberate traces of the photographer's finger across the dusty glass cover that protects four curiously banded eggs of the glossy ibis. The scientist reflects: "I once opened a drawer in an old part of our collections. The contents had been dumped and sheepishly piled back in disarray.... I found a note, also encrusted by the universal patina. It was dated 1861, and contained...an apology ...penned by a terrified student lest the intense and temperamental boss of the museum, Louis Agassiz himself, discover such a calamity without appropriate documentation. Agassiz, obviously, never opened the drawer. The student, Nathaniel Southgate Shaler, went on to become a famous scientist. The dust still rains and reigns."

A pair of photographs show details of the jaws of two whales. "Yet, for color and form alone, we might be viewing America's...desert at 30,000 feet over Arizona." That ambiguity is one of scale. Yet all life is rigorously modular, truly mason's work, and the ubiquitous atomic modules have a fixed size, never changed under the hammer of evolutionary time. "When I understood science far less well than I do today," Gould recalls, "I used to get annoyed at...professional talks. The speaker would show a slide, often quite beautiful...and someone would interrupt by shouting: 'What's the scale?' 'How crass and narrow-minded,' I used to say to myself.... Now I understand the deep sense behind such a question."

This bestiary's artist could indeed wish to exploit such a profound ambiguity. A frank scale bar next to the image, even a numerical statement in the caption, might be held crass. But her scientist partner nodded in neglecting to offer us even a tabular summary of object sizes, placed diffidently in the end matter. What's the scale, Steve?

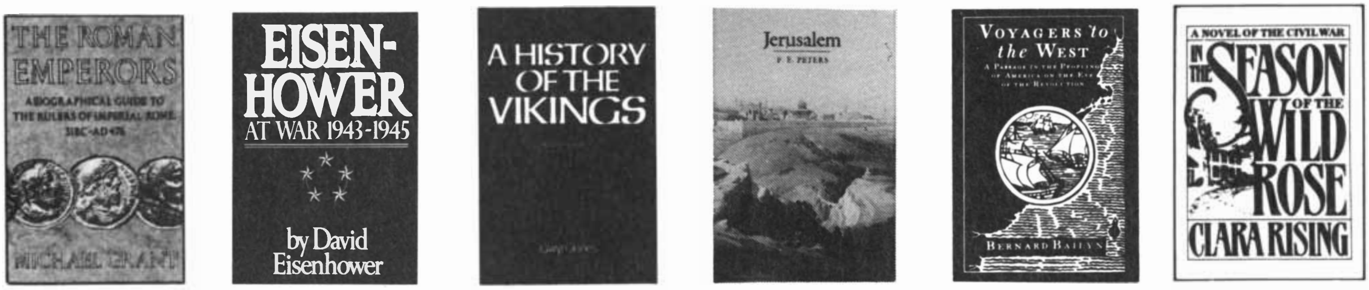
A deeply felt celebration of diversity in word and image, this intimate collaboration is dedicated to Agassiz' Museum of Comparative Zoology at Harvard. Specimens were sought not only there but in Washington, Dublin, Leiden and Copenhagen, all famous and dusty treasuries of this kind of probing into life after death. Shadowy among the bright pages of this entirely modern volume a reader glimpses a rapt and ancient scene, a shaman dancing entranced, shaking her rattle of bear claws. We perforce share life and death with every other animal species, but the symbolic sharing of insight is ours virtually alone.

The pi-shaped molars of the rhinoceros

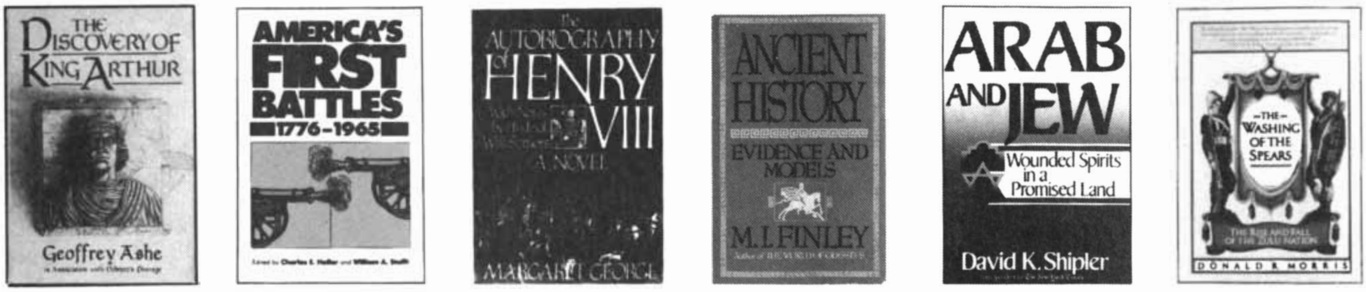
TAKE ANY 3 FOR $1.00 EACH

(First price is publishers' list. **Boldface** shows Members' Price.)

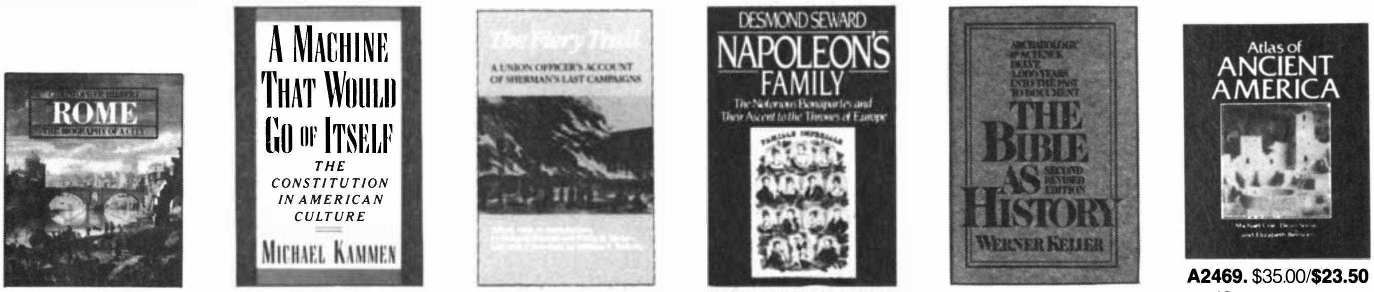
With Trial Membership.



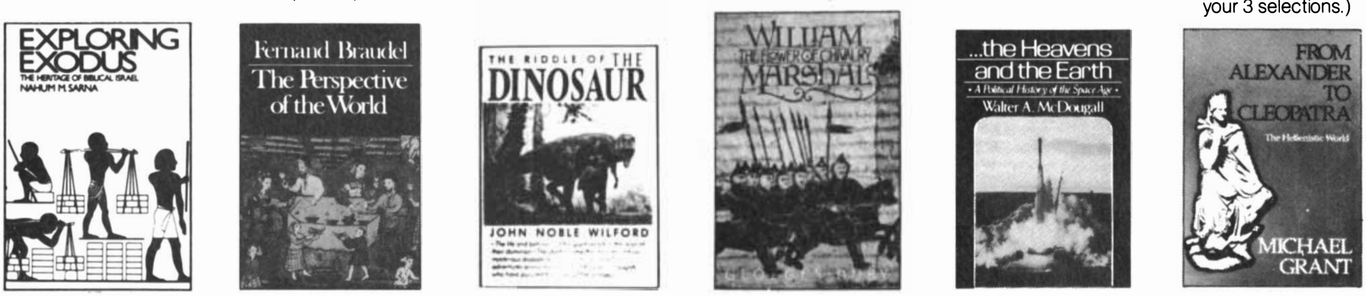
1065. $25.00/**$17.50** | 1917. $29.95/**$22.50** | 1172. $25.00/**$16.50** | 4861. $35.00/**$23.50** | 2279. $30.00/**$18.50** | 4036. $19.95/**$15.95**



1339. $18.95/**$14.95** | 3657. $29.95/**$19.95** | 3509. $19.95/**$15.95** | 7047. $17.95/**$14.50** | 2402. $22.50/**$17.50** | 2436. $21.95/**$15.50**



2071. $25.00/**$17.50** | 2485. $29.95/**$19.95** | 5942. $22.50/**$16.50** | 1222. $16.95/**$13.95** | 3012. $14.95/**$10.95** | A2469. $35.00/**$23.50** (Counts as 2 of your 3 selections.)



2246. $17.95/**$15.50** | 1453. $35.00/**$18.50** | 4390. $22.95/**$16.50** | 5629. $15.95/**$12.95** | 2865. $25.95/**$17.50** | 5736. $19.95/**$13.95**

No matter what area of history you enjoy reading about most, you'll find that The History Book Club offers some of the finest selections being published today. And no book club we know of offers greater savings – an average of 30% off publishers' list prices.

You can save even more by taking advantage of our Introductory Offer. Select any three books on this page for $1.00 each when you take a fourth book at the low Members' Price. Your total savings as a Trial Member, including the Introductory Offer, can be more than 50%.

How the Club works: As a Club member, you'll be able to choose from the 150 to 200 books featured each month. Our books are always equal to the publishers' editions, never "economy" reprints, and are available at savings of $3, $4, $5, or more, off bookstore prices. You need take only four more books in the next two years, after which you'll save even more through Bonus Books that you may choose from our entire list.

Approximately every four weeks (14 times a year), you will receive our *Review* and a dated reply card. If you want the Editors' Choice, do nothing – the book will come automatically. If you want another book, or no book at all, return the card by the date specified. (Book shipments will be charged to your account at low Members' Prices, plus postage and packing.)

If you should receive an unwanted book because you had less than 10 days to decide, you may return it and pay nothing. We guarantee postage.

The History Book Club, Dept. N,
40 Guernsey Street, Box 790 Stamford, CT 06904-0790

Please enroll me as a Trial Member and send me the four books whose numbers I have entered below. Bill those on the left at $1.00 each and the fourth at the low Members' Price, plus postage and packing.

[ ] [ ] [ ] [ ]

I may return the books within three weeks at your expense and owe nothing. Or, I will buy four more selections within the next two years, *not including my choices above.* Either the Club or I may cancel my membership anytime thereafter. All Club purchases are at low Members' Prices, and a postage-and-packing fee is added to all shipments. SB0018SP

Print Name________

Address________ Apt.#____

City______ State______ Zip______

In Canada: Mail to U.S. Address

ST87G

FOOD PROGRAM at St. Bartholomew's Episcopal Church in New York typifies the effort that private agencies throughout the U.S. have made to cope with the rise in hunger resulting from cutbacks in Federal programs. For five years this Saturday Morning Feeding Program has provided as many as 275 people per week with packages of food to supplement their meager rations.

Established 1845 SCIENTIFIC AMERICAN February 1987 Volume 256 Number 2

Hunger in the U.S.

Virtually eliminated in the 1970's, this blight has returned because of Federal cutbacks. By a common definition of hunger some 12 million children and eight million adults are hungry

by J. Larry Brown

There is no universal definition of hunger, but a concept generally accepted in the medical community is that a hungry person is chronically short of the nutrients necessary for growth and good health. Asking how many Americans are hungry by that definition, one encounters an astounding statistic: 12 million children and eight million adults, or about 9 percent of the population. A few of those people might be going hungry because of ignorance or indifference, but for most of the group the reasons are an economy that leaves many families below the poverty level and a social-welfare system that gives them insufficient help.

It is particularly ironic that extensive hunger was identified three decades ago and that the Government's responses during the 1960's and 1970's virtually eliminated the problem. The food-stamp program was expanded to serve 20 million people. Programs to reach and feed the isolated elderly were instituted. School lunch and breakfast programs were increased so that poor children would have the nutritional basis for doing well in school. A supplemental feeding program was set up to supply food and nutritional advice to poor pregnant women and mothers, and food to their infants.

In 1981 the Reagan Administration began reducing and modifying those programs. By 1982 signs of hunger were widespread. Increasing numbers of people coming to churches and social-service agencies said they did not have enough food. Physicians in some parts of the country began to hear from an increasing number of patients that they were hungry and to observe patients who manifested illnesses associated with hunger—among them anemia, tuberculosis, poor growth in children and osteoporosis in adults.

Later that year came the first of a number of reports documenting the extent and the spread of hunger in the U.S. Altogether 20 national studies on hunger were made between October, 1982, and March, 1986. The reporting groups included the U.S. Conference of Mayors, the National Council of Churches, the U.S. Department of Agriculture and the organization of which I am chairman: the Physician Task Force on Hunger in America, based at the Harvard School of Public Health. Each of the studies reached the conclusion that hunger had reappeared as a serious national problem.

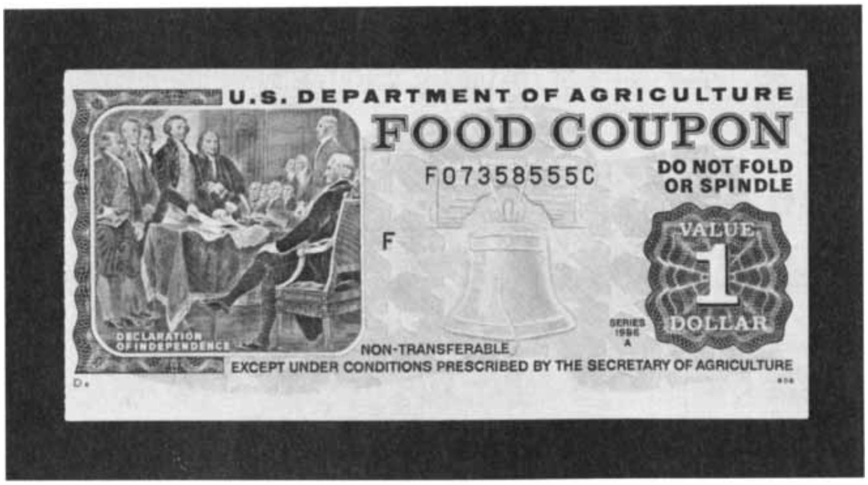
Typical findings were turned up by the mayors' conference, which late in 1984 surveyed cities and towns to determine whether and how the reported economic recovery was affecting the poor. Nearly 75 percent of the municipalities reported that the need for emergency food assistance had increased in 1984. The number of emergency food centers (including soup kitchens, which serve meals, and food pantries, which give out groceries) increased by 15 percent during the year. The total number of meals provided rose by 50 percent over the figure for 1983. Several cities reported even larger increases: 200 percent in Boston, 182 percent in Chicago and 100 percent in Dallas.

In November, 1984, the Food Research and Action Center surveyed some 300 emergency food programs and found that 65 percent of them reported an increase during the year in the number of hungry families seeking food. The increase in the number of families fed was more than 20 percent. A majority of the emergency centers reported that more than half of the hungry people they served were families with children.

In 1985 the Physician Task Force calculated that hunger afflicts some 20 million Americans. Although more recent data are not available, later evidence suggests that the problem of hunger in America has grown worse.

People in certain population groups are particularly susceptible to the ravages of hunger. Pregnant women, infants and children on the one hand and elderly people on the other are likely to suffer the most harm when food is inadequate. At any age malnutrition can cause weakness and lethargy. It can impair the functioning of the immune system, making the person more vulnerable to infectious diseases. The absence of specific nutrients can cause a range of deficiency diseases affecting all the body's organ systems. Severe and protracted malnutrition can lead directly or indirectly to death.

For a pregnant woman the quality and quantity of food available is a critical factor in the healthy development of her child. When nutrition is inadequate during pregnancy, a number of dangers arise for the mother. One of them is anemia, which reflects an inadequate supply of iron for the increased needs of pregnancy. Toxemia, a complex of symptoms involving an increase in blood pressure and a swelling of tissues due to excessive retention of fluid, is a common and serious hazard of pregnancy; recent findings suggest it



FOOD STAMP issued by the U.S. Department of Agriculture is intended to help needy families obtain food. The stamps can be redeemed for food at grocery stores. Because of recent changes in regulations only about 19 million people receive food stamps even though some 33.4 million are living in poverty. The average benefit per meal is 49 cents.

may reflect inadequate nutrition (particularly a lack of protein).

The growing fetus requires a diet rich in protein, vitamins and minerals as its tissues and organs develop. The infant's health may therefore be compromised by poor maternal nutrition. The risks include prematurity (defined as birth at or before 37 weeks of gestation) and low birth weight (defined as less than 2,500 grams, or about 5.5 pounds). Premature infants are at risk of respiratory distress and a weak immune response. Malnutrition of the mother exposes even a full-term baby to such health problems as hypoglycemia (a deficiency of sugar in the blood), hypocalcemia (a deficiency of calcium in the blood) and polycythemia (an excess of red blood cells) as well as longer-term deficiencies in growth and development.

The hazards facing the frail newborn increase its risk of dying in infancy. Even when complex and costly technology is brought to bear, a low-birth-weight baby faces 40 times the normal likelihood of dying before it reaches its first birthday. Low birth weight is a factor in more than half of the infant deaths in the U.S. and in 75 percent of the deaths of babies less than one month old.

The potential hazards of malnutrition for a pregnant woman and her fetus can be overcome by nutritional supplements supplied to the mother during her pregnancy. When food was tightly rationed during World War II, the United Kingdom made sure that pregnant women got adequate diets, and infant mortality actually declined. In places such as Leningrad and the Netherlands, where pregnant women did not get special supplementary rations, birth weights declined and infant mortality increased.

More recent studies demonstrate the positive impact of nutritional supplementation for pregnant women. The studies examined the effect of the Special Supplemental Food Program for Women, Infants and Children (WIC), which was established by the Government in 1972. Eight such studies were made between 1976 and 1986, and all reached the conclusion that the program had resulted in better maternal health and better outcomes of pregnancy. Among the findings were a decrease of more than 20 percent in low-birth-weight babies, a reduction in late fetal deaths, fewer premature infants and fewer infant deaths.

Whatever the circumstances of an infant's gestation and birth, the child once born is at considerable risk if adequate nutrition is lacking. The human brain develops most rapidly from conception until about the age of three. During those years brain function can be impaired by nutritional deprivation. Other risks are stunting (defined as height below the fifth percentile for a given sex and age) and wasting (weight below the fifth percentile for age and height). In addition malnourished children are particularly vulnerable to lead and other environmental toxins, which can affect the brain and compound the direct effects of malnutrition on the child's intellectual development.

As at all ages, malnutrition in childhood can weaken resistance to infection. Poorly nourished youngsters are at risk of more frequent colds, ear infections and other infectious diseases. In this way the developmental hazards inherent in poor nutrition may be further compounded by absence from school and even the curtailment of some childhood activities.

In fact, recent research indicates that functional impairment may result from poor childhood nutrition even in the absence of overt physical harm. The studies suggest that before easily measured changes in growth appear and before any nutrition-related disease is evident, a child's body may adapt to inadequate food by reduced metabolism. The implication is that the child may show no overt signs of impairment and yet be deprived of social and cognitive experiences that advance development.

Loss of function also characterizes the effect of mild malnutrition on young and middle-aged adults. Both reduced productivity at work and impaired social function are potential outcomes. Adults of every age are vulnerable to infections and to various deficiency diseases associated with malnutrition.

In old age the risks of malnutrition are heightened further. During these years the impact of food on the maintenance of health and the prevention of disease is particularly crucial. The elderly often need special diets because of common ailments such as hypertension and diabetes. Deficiency diseases increase the requirement for certain nutrients; an increased intake of calcium to prevent osteoporosis is an example. Some conditions of old age may impair digestion or the absorption of nutrients, making the choice of nutrient-dense foods critical. At the same time old age often entails a number of obstacles to an adequate diet. Difficulties in such activities as shopping and cooking, trouble in chewing some foods and a lack of appetite due to social isolation and the resulting depression can all interfere with good nutrition in old age.

In this catalogue of the effects of hunger the situation of infants merits a further look. A nation's infant mortality rate (the number of babies per 1,000 live births who die before their first birthday) is widely accepted as a meaningful indicator of the society's general health status. By this measure the U.S. does not do well compared with other industrialized nations. Its infant mortality rate of 11.2 places it 18th in the world. Japan, Sweden and the U.K. are among the countries that do better; Spain and Hong Kong also outdo the U.S.

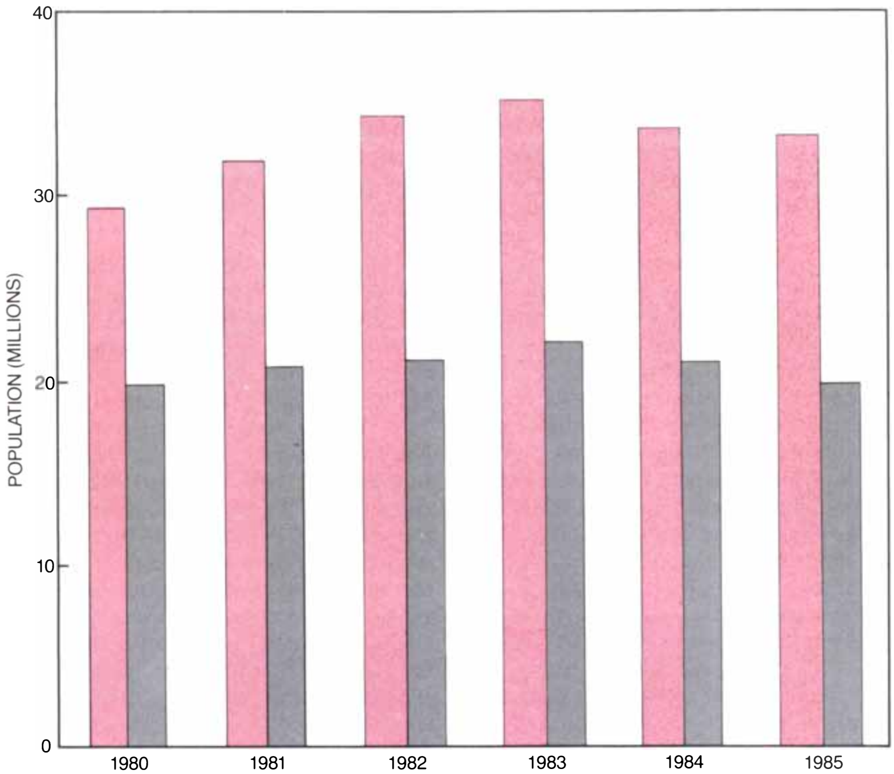
This record masks discrepancies between subpopulations in the U.S. Nonwhite infants are about twice as likely as white ones to die in the first year of life. As the Children's Defense Fund stated in its 1984 report, *American Children in Poverty:* "The nonwhite infant mortality in our nation's capital exceeds that in Cuba and Jamaica, both significantly poorer countries. Statewide data mask more severe infant death problems in some city neighborhoods. For example, two Baltimore census tracts show infant death rates as high as 59.5 per 1,000 live births. This exceeds the 1981 infant death rates in Costa Rica, Panama, Guyana and Trinidad and Tobago, and is more than double rates in the Soviet Union."

A further cause for concern is that the infant mortality rate actually rose in 11 states and several urban areas in the early 1980's. Yet another cause for concern is the possibility that the declining rate of infant mortality for the U.S. as a whole does not necessarily indicate that healthier babies are being born. The U.S. has perhaps the world's most sophisticated technology for saving infants born too early or too small to survive on their own. The technology saves a high enough proportion of low-birth-weight babies to obscure the true impact of inadequate prenatal nutrition and care.

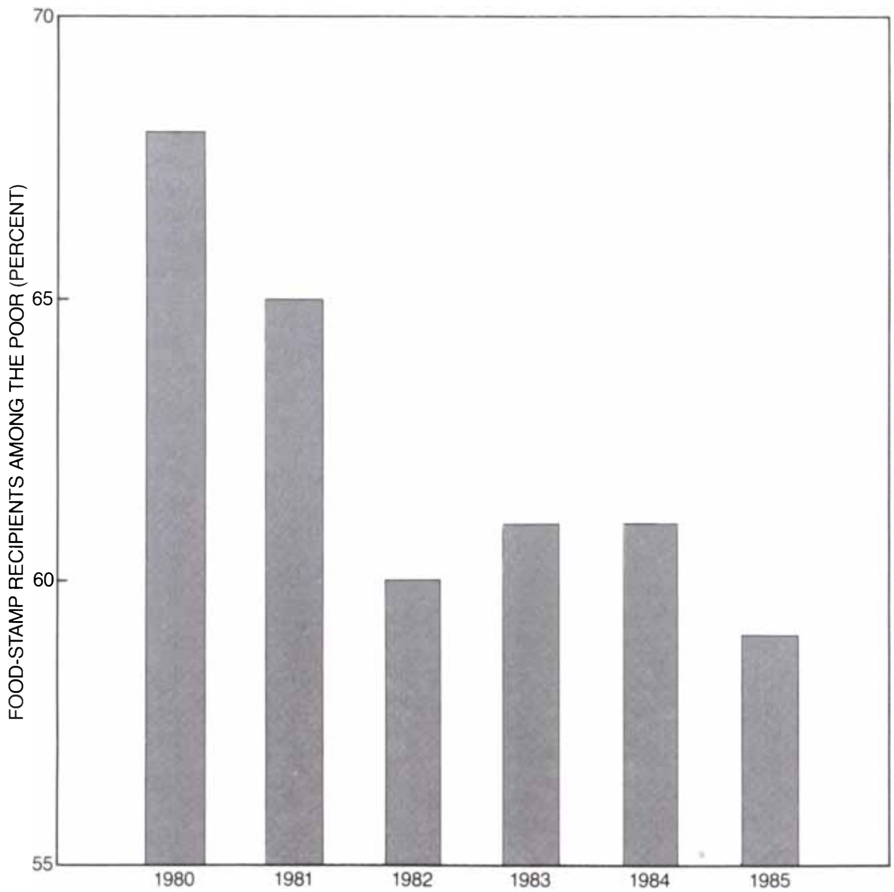
Infant mortality is not the only birth outcome that reveals poor maternal health and nutrition. The records on low birth weight show that it, like infant mortality, is unevenly distributed. The Child Health Outcomes Project of the University of North Carolina found that black babies are more than twice as likely as white babies to be born at a low weight. In New York City, which keeps birth-weight data by community, black central Harlem had the highest percentage (16.3) of low-birth-weight babies in 1982 and the affluent, largely white upper East Side had the lowest (5.4).

The role of poverty as a cause of low birth weight is evident in data on poor black and white patients at Metropolitan Nashville General Hospital. The two groups were quite close (14.4 and 12.9 percent respectively in 1981) in low birth weights. The figures suggest it is not some biological correlate of race but rather income and the correlates of poverty that explain differences in birth outcomes.

Similar socioeconomic variations are reflected in the data on the growth of children. Of the various indicators that characterize the health and nutrition status of American children, anthropometric data (measurements of height, weight and other body dimensions) are the most widely available. Height and weight are routinely measured in medical examinations of chil-



PARTICIPATION in the food-stamp program has declined since 1980 as a proportion of the number of people living at or below the poverty level, defined in 1985 as an annual income of $10,989 for a family of four. The chart shows the number of people living in poverty (*color*) and the number taking part in the Federal food-stamp program (*gray*).



MONTHLY PARTICIPATION in the food-stamp program as a percentage of the number of people living at or below the poverty level is traced. The Government reduced the food-stamp program by $7 billion in the six years and narrowed eligibility requirements.

dren, and so a vast body of material is at hand for analysis.

The analysis depends on the existence of norms with which the findings on any individual child can be compared. When a physician checks the height and weight of a child, he or she records the figures on a chart that graphs growth over time. The chart also allows comparison with other children of the same age and sex. For example, a three-year-old girl who is ranked in the 40th weight percentile is heavier than 40 percent of all other three-year-old girls. Extensive measurements have been done on children to obtain cross-cultural norms. A physician looking at the chart of a three-year-old girl wants to know two things: first, whether she is within the expected range for her age and sex; second, whether her growth is following a fairly smooth curve upward over time.

Generally children who rank above the 95th percentile for weight in relation to height or below the fifth percentile for either weight or height are watched carefully. For some children the pattern may be genetically determined and perfectly normal; for others aberration from the norms may reflect nutritional or other health problems.

Many studies in the U.S. of children in diverse ethnic and racial groups have revealed a close association between poverty and impaired growth. For example, the Massachusetts Nutrition Survey (mandated by the state legislature to assess the health of low-income children) found that 18.1 percent of the children evaluated showed some indicator of chronic nutritional deficiency. The phenomenon was inversely related to income level.

The studies I have cited show that hunger and nutritional deprivation are directly associated with poverty and that many forms of ill health associated with inadequate nutrition are tied to household income. The poor—because they are poor—are likely to have too low an income to support a nutritionally adequate diet.

The poverty level, as defined by the Federal Government, is currently $10,989 per year for a family of four. A number of clinical studies have found malnutrition to be substantially higher among families at or below that level than among the better off. The National Household Food Consumption Survey, published in 1982 by the Agriculture Department, found that more than 80 percent of all households whose expenditure for food is similar to that of the average food-stamp recipient fail to obtain the dietary allowances of nutrients recommended by the National Academy of Science. The Massachusetts Department of Public Health found that children in low-income families are twice as likely to be malnourished as children from middle-income families. Other health data reveal high rates of nutrition-related illness among the poor.

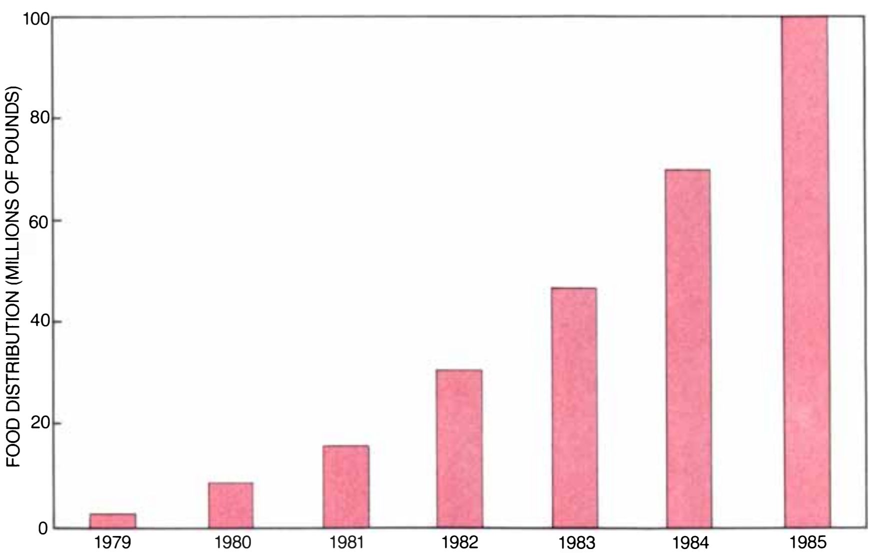
Although income is not the only factor associated with poor health, it appears to be the chief one. What precisely is the connection? One hypothesis often advanced is that it is not income itself that leads to greater malnutrition and illness among the poor but rather lack of knowledge about nutrition, food buying and cooking.

Inadequate information about nutrition is widespread in the population, but little evidence exists to support the notion that it is primarily a problem among the poor. Indeed, some evidence contradicts the notion. The national surveys of food consumption consistently found that poor people actually buy more nutritious food than the rest of the population. Apparently the poor think more about how to make wise choices because they have to stretch their food dollar as far as possible. Both the evidence and common sense demonstrate that the recent surge of hunger in the U.S. cannot be attributed to ignorance among the poor. Those advancing that hypothesis would have to explain why the poor so suddenly became ignorant.

The clear association between poverty and hunger suggests that to understand the phenomenon of increasing domestic hunger one must examine the status of the nation's low-income households. In this sense hunger is like any infectious disease: it has a cause. Three factors appear to account for the "disease" of hunger. One of them is the condition of the nation's safety net: the system of support for people who live in or close to poverty. Another is the rise in economic jeopardy stemming from economic recession. The third factor is recent public policy, in particular the cutbacks in Federal programs relating to nutrition.

The American safety net is weak compared with what is in place in other industrialized nations. For example, the U.S. is one of only two industrialized nations (the other is South Africa) that have no program of national health insurance. In more than half of the states poor families with children are denied welfare assistance if both parents are present; the Federal Aid to Families with Dependent Children (AFDC) program comes into play only if the family breaks up. Many of the same states provide no optional coverage under Medicaid for the special category of families defined as "medically needy."

Nearly half of the Americans in poverty receive no food stamps. Those who do get stamps receive an average benefit of 49 cents per meal. To compound the matter the Agriculture Department ties food-stamp assistance to the "thrifty food plan," which was devised to help people through periods of short-term economic distress. When the plan was developed, the department's nutritionists warned that families living at this level of expenditure for food would fail to receive the recommended daily dietary allowance. Strange as it may seem, the department now uses that plan as the norm. Hence the level of benefits for food stamps is the level at which people are



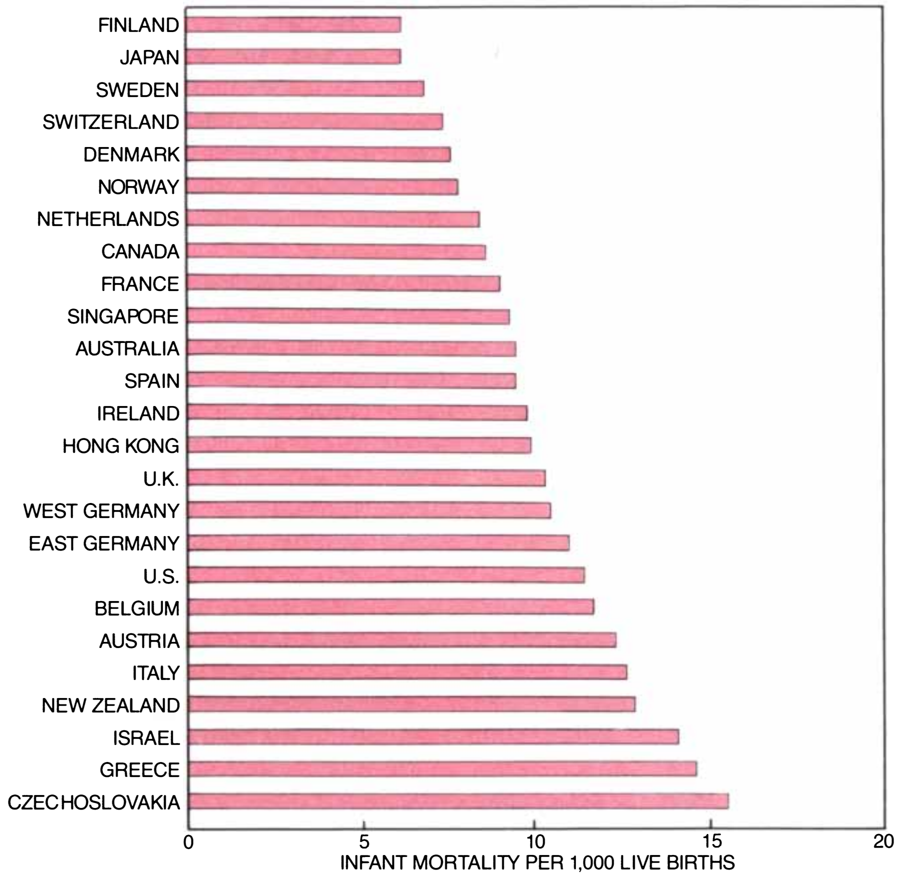
DISTRIBUTION OF FOOD by the Second Harvest National Food Bank Network rose from 2.5 million pounds in 1979 to 100 million in 1985, reflecting the efforts of private organizations to cope with the rise in hunger. Second Harvest services state food banks.

very likely to become malnourished.

Over the past decade the average net income of families in the AFDC program has dropped nearly 40 percent. In other words, well over one-third of their purchasing power was stripped away because benefit levels failed to keep pace with inflation. No other segment of the economy has fallen so far behind inflation in recent years.

Unemployment in the U.S. rose from 6.2 percent in early 1980 to 10.8 percent in 1983. Although it has since fallen to about 7 percent, more Americans are unemployed today than were unemployed in 1980. Moreover, the figures do not depict the severity of unemployment.

A smaller proportion of jobless people receive unemployment benefits today than at any time in recent history. In the recession of 1975, for example, nearly 80 percent of jobless workers received benefits; today the figure is 29 percent, which is an all-time low for the program. The reasons are the prolonged sluggishness of the economy and changes in Federal policy that limit eligibility for unemployment compensation. Today if an unemployed worker living with a spouse and children has exhausted all benefits, the family is unable to get further benefits or AFDC and Medicaid in nearly half the states.



INFANT MORTALITY RATE indicates a nation's health status. The U.S. is 18th among 25 industrialized nations with a population of 2.5 million or more. The rate reflects infant deaths in the first year of life per 1,000 live births. Frequently the babies weigh less than 2,500 grams (5.5 pounds) at birth because of the mother's poor nutrition.

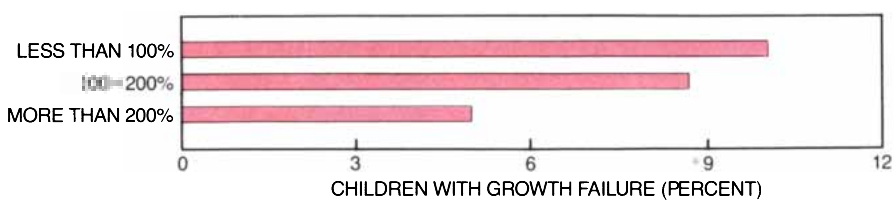
Together these factors and others have pushed economic hardship to a level almost unparalleled in recent history. Increasing numbers of people were forced into the position of needing Federal hunger-relief programs, but the Government's response was to reduce the assistance available to the poor. The cuts in Federal nutrition programs were the severest in the nation's history. Two of the basic food programs (food stamps and school nutrition) were cut substantially. Two others (feeding of the elderly and the WIC program) would have been cut similarly had it not been for congressional intervention.

Among the changes in the food-stamp program, in addition to cuts totaling $7 billion in the fiscal years 1982–85, were the elimination of an adjustment for inflation in 1982 and the delay of certain other inflation adjustments. Furthermore, households with incomes just above the official poverty level were eliminated from eligibility unless they included elderly or disabled members. Several other changes reduced the basic benefits at the same time as inflation, unemployment and poverty were increasing. Consequently today about 19 million Americans receive food stamps at a time when the Government recognizes that 33.4 million are living in poverty.

In 1982 about 23 million children took part in the school lunch program, the largest of the endeavors to improve the nutrition of poor children. The program was cut back substantially in 1982. The income criteria by which eligibility is determined were changed, and the application procedure was altered to make it more difficult for people to qualify. Altogether $5 billion was cut from child-nutrition programs in the fiscal years 1982–85. The immediate impact was that one million children were dropped from the school lunch program. Then, because of the changes in the program, some 2,700 schools stopped offering it. More than 400,000 children and 800 schools were eliminated from the school breakfast program.

One of the notable benefits of the school programs has been that they ensure one and often two nutritious meals a day for low-income children when school is in session. Moreover, large numbers of school officials assert that the programs bring social and educational benefits as well.

The enactment of $12 billion in budget cuts for Federal food-assistance programs came during a time when poverty was reaching the highest level in nearly two decades. The cuts knocked the economic props from under many people—the still poor and the new poor. No safety net was there to catch them.



POVERTY AND INSUFFICIENT GROWTH are correlated. The bars show children found to be of low height for their age in the Massachusetts Nutrition Survey of 1983. The numbers at the left refer to the family's income as a percentage of the poverty level.

The Microtubule as an Intracellular Engine

Microtubules have become familiar elements in the frameworks that organize cell shape and cell division. Now a powerful new technique in light microscopy has revealed their function in two-way transport

by Robert Day Allen

Microtubules are a relatively recent addition to the list of ubiquitous cell components. Thirty years ago one could only guess at the existence of the slender fibers; 25 years ago biologists first discerned their delicate features. Once researchers had learned how to detect them, microtubules turned up in an unexpected variety of roles, from the spindles that orchestrate cell division to the filaments girding blood cells that never divide. They quickly gained a reputation as the stalwart structural units of the cytoskeleton.

Now it seems the versatility of microtubules extends beyond mere structural performance. With the aid of powerful new computer-enhancement techniques for the light microscope, we have been able to watch microtubules as they propel vesicles and organelles across the cytoplasm with surprising speed. Each tenuous thread, many micrometers in length and composed of an abundant protein called tubulin, can move particles in opposite directions simultaneously; moreover, a microtubule can make its way across a glass microscope slide even when it is broken and separated from its cellular surroundings.

My colleagues and I have tried to clarify the mechanism by which microtubules move themselves and their intracellular cargo. Our work and that of many other investigators has linked microtubules to the transport that coordinates traffic between a nerve's cell body and its extremities. Defects in microtubules may be involved in a dozen or more human neurological disorders. Outside the nervous system, phenomena as diverse as color changes in fishes and the movement of viruses within cells depend on microtubule conveyors. Hence our research may yield insights in areas far beyond the simple aquatic protozoan that initially captured my interest.

The particle movement we now call microtubule-dependent transport was studied long before the discovery of the microtubules themselves. The clearest historical account of such transport comes from the 19th-century naturalist and microscopist Joseph Leidy. Leidy found in a pond near his home a curious animal with a remarkable network of filopodia, or "false feet." The organism, which he called *Gromia,* was a foraminifer, a freshwater relative of plankton that is barely visible to the unaided eye. Within *Gromia's* tentaclelike filopodia Leidy observed a dynamic stream of motion, which he described in a report to the U.S. Geological Survey in 1879: "The pseudopodial extensions of *Gromia* consist of pale granular protoplasm with coarser and more defined granules. . .in incessant motion along the course of the threads, flowing in opposite directions in all except those of greatest delicacy."

Later observers echoed Leidy's precise description of *Gromia* filopodia but added little to the understanding of the unusual streaming effects. In human experience, flow within a single enclosure can occur in only one direction, along a pressure gradient. What fascinated me and other modern workers about the foraminifer networks was that, within a single filopodium, particles seemed to be flowing at different speeds and in both directions. Somehow foraminifers were bending the rules.

Disclosing *Gromia's* secrets took a bit of time. In 1957 I visited Zach M. Arnold at the University of California at Berkeley to study the filopodial movement in *Allogromia laticollaris,* a marine foraminifer he had cultured and named. This organism and a related *Allogromia sp.* N.F. would eventually become the raw material for the important experiments linking particle movement to microtubules. They are similar to *Gromia* but, unlike Leidy's foraminifer, grow in sufficient numbers to supply research and so serve as good experimental models.

My visit to the Berkeley laboratory came six years before microtubules were discovered, when the techniques of microscopy were too primitive to depict structures in any more detail than what Leidy himself had seen. Research was slow and tedious. Even so, we could examine some simple hypotheses by predicting patterns of movement and testing the predictions by simple observation. Taking this approach, Theodore L. Jahn and Robert A. Rinaldi of the University of California at Los Angeles suggested in 1959 that the moving particles were caught up on "millipedelike" fibers of gel capable of crawling over one another, each displacing the other by "little legs."

My experience with *Allogromia* and some related foraminifers I had studied at the Roscoff Marine Biological Laboratory in France persuaded me that Jahn and Rinaldi's explanation might be too simple. In the finest of the protozoan "feet," particles travel more quickly toward the tip than away from it. I offered a scenario in which the smallest filopodia contained V-shaped, doubled-over gel fibers that contracted at the bent regions near the tip. Although that model could account for

INSIDE A NERVE CELL microtubules (*yellow*) mediate the transport of vesicles (*blue, crimson*) and organelles (*maroon*) through the corridor that connects the cell body to the terminal extensions of the cell.

faster particle movement in one direction, subsequent research has shown I was no closer to the truth than were the U.C.L.A. investigators.

In the course of my early investigations, however, I also noticed that the filopodial network would shine brightly under a polarization microscope. This phenomenon hinted at the presence of oriented submicroscopic fibers. Whatever the fibers were, they must be very delicate or unstable: the slightest change in external conditions could cause them to break down into rows of cytoplasmic droplets connected by membranous "stockings."

It was this fragility, in fact, that for almost 10 years bedeviled my attempts to show that microtubules were present in large numbers in the filopodial networks of foraminifers. In 1963 David B. Slautterback of the University of Wisconsin Medical School first identified microtubules in the freshwater polyp *Hydra*. That same year Myron C. Ledbetter and Keith R. Porter of Harvard University found them in plant cells. Immediately we suspected that microtubules were the "stiffening elements" in *Allogromia* filopodia. Yet it was not until 1971 that Samuel M. McGee-Russell and I, then working at the State University of New York at Albany, found out how to overcome the poor methods for preparing specimens that had disrupted the microtubules before they reached the electron microscope.

Similarly another 10 years elapsed before we were able to document the direct connection between particle

movements and microtubules. In 1981 two simultaneous advances in my laboratory at Dartmouth College improved visualization of the fragile fibers tremendously. Jeffrey L. Travis, a graduate student working with me on motility in *Allogromia,* devised a method for growing the organism on the glass cover slides used in light microscopy. His slides, treated with positively charged substances such as poly-L-lysine, caused filopodia to stick to the glass and spread into large, flattened sheets that were ideal for the observation of organelle traffic. Travis also prepared whole rather than sectioned samples that for the first time revealed the complexity of the microtubular "wiring diagram" at the level of electron microscopy.

In the meantime my serendipitous discovery of what came to be called Allen video-enhanced contrast (AVEC) microscopy enabled us to visualize structures that were an order of magnitude smaller than had ever been visible with the light microscope. We could see single microtubules even though their diameter was only 25 nanometers, or one-millionth of an inch. By means of AVEC microscopy Travis and I watched the interaction of particles and microtubules for the first time. We saw that mere physical contact with microtubules was enough to spur the acceleration of stationary particles. Travis also observed that the microtubules themselves moved axially and laterally within the filopodial network. We could not, however, determine whether they were actively pushing themselves or being pulled along passively. Neither could we find out whether particles could move on a single microtubule.

We thought our experimental model might be imposing these limitations and so turned our attention from *Allogromia* to other cell systems amenable to AVEC enhancement. We became interested in the transport of organelles along axons, the major processes of nerve cells along which nerve impulses are conducted. Axonal transport in chick neurons was known to move particles in both directions at rates approaching 200 millimeters per day—a fairly fast clip on a microscopic scale. This rapid transport had first been described in 1965 by another graduate student of mine, William Burdwood. When he reported it, the effect was considered to be no less than revolutionary; in the 15 years that followed, many hypotheses took shape to explain it. Because direct evidence remained scant, no single theory had gained favor over the rest.

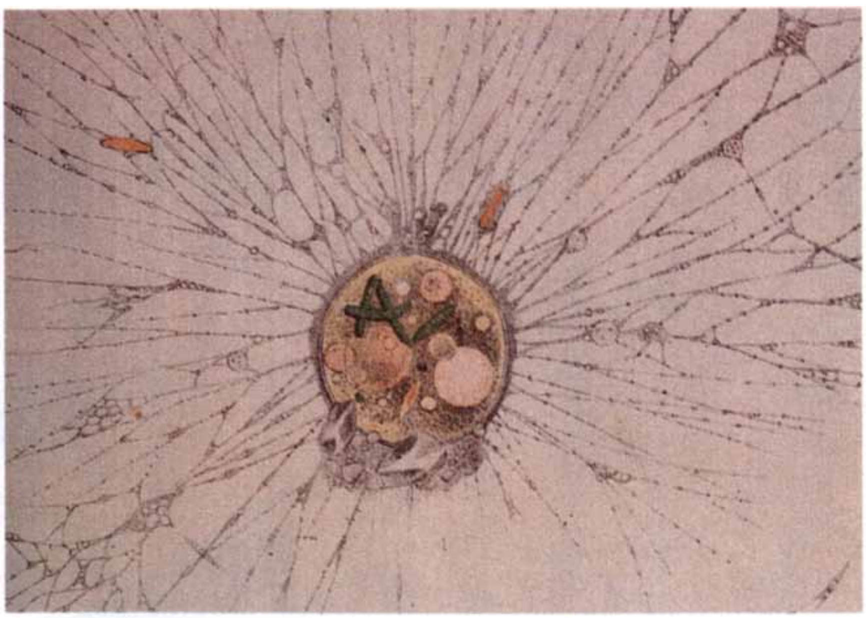
During the summer of 1981, which I spent at the Marine Biological Laboratory in Woods Hole, I decided to see whether AVEC microscopy could provide any insight into axonal transport. I chose the giant axon of the squid as my subject. The axon is part of a large, transparent neuron ideal for my purpose except for one hitch: no one had ever seen any hint of movement in the squid axon. I asked several colleagues who had considerable experience with invertebrate neurons to join me in my laboratory at Woods Hole to see what we could accomplish.

Without AVEC's accoutrements the view from our differential interference contrast microscope was somewhat discouraging. We could see only the hazy outlines of the axolemma (the membrane that envelops the axon) and a thin covering layer of connective-tissue cells. Nothing was moving. Then we engaged AVEC techniques and switched the image to the television monitor. Suddenly the screen was teeming with tiny particles moving in both directions along "filaments" of some kind. The excitement we felt rippled through the rest of the Woods Hole community.

In July we made the first recordings of particle transport in the squid giant axon. We could identify most of the particles by size. The large, elongated mitochondria, harboring the chemical energy of adenosine triphosphate (ATP), moved in both directions in jerky, discontinuous motion. Smaller spherical particles heading toward the cell body were multivesicular bodies carrying "surplus" membrane. The smallest, least visible particles moved continuously in masses, like dense schools of small fish, toward the synaptic terminal—the axon's end. These were precursors of the synaptic vesicles, transporting transmitter substances for release when the nerve cell was stimulated. They had never been seen before with the light microscope.

When the dust settled, we had time to assess our new findings. Our experiments had at least established the directions in which different particles migrate and had shown that larger particles travel more slowly than smaller ones. It seemed we had also provided the first evidence for continuous particle movement in animal cells. Such transport was unexpected because continuous motion had hitherto been observed only in the so-called protoplasmic streaming of plant cells. Particle movement in animal cells had always been characterized by discontinuous motion.

What caused the motion? The particles did not seem to be swept along on some protoplasmic current, which one observes in plant cells. Some investigators had proposed that the electric charge accompanying nerve impulses somehow fueled axonal transport. But when we stimulated axons electrically, transport was not affected. The particles invariably followed the path of the "linear elements" that looked



GROMIA, a freshwater relative of amoebas, provided evidence of two-way transport more than 100 years ago. The microscopist Joseph Leidy, who drew this illustration in 1879, noted "granules" flowing in both directions through the long, slender filopodia that stretch like tentacles from the animal's body. Flow in two directions within a single enclosure was inexplicable, and it sparked the speculation that led to the author's investigations. Outside the microscope's scrutiny, *Gromia* looks no larger than a grain of sand.

a great deal like the microtubules we had seen in *Allogromia*.

Reflecting on our collective observations, we began to suspect that microtubules could be involved in particle transport in both foraminifers and neurons. We had to admit, however, that axons contain other linear elements such as neurofilaments and F-actin, and in 1981 the prevailing view was that, of the three, F-actin was the likeliest to be intimately involved in axonal transport [see "The Transport of Substances in Nerve Cells," by James H. Schwartz; SCIENTIFIC AMERICAN, April, 1980].

With a resolution of about 100 nanometers at best, we could not tell what the linear elements were, how closely they were packed or what their exact relation to the transported particles might be. We thought we could begin to answer these questions if we could take the axon apart, isolate the elements and particles and watch their interaction.

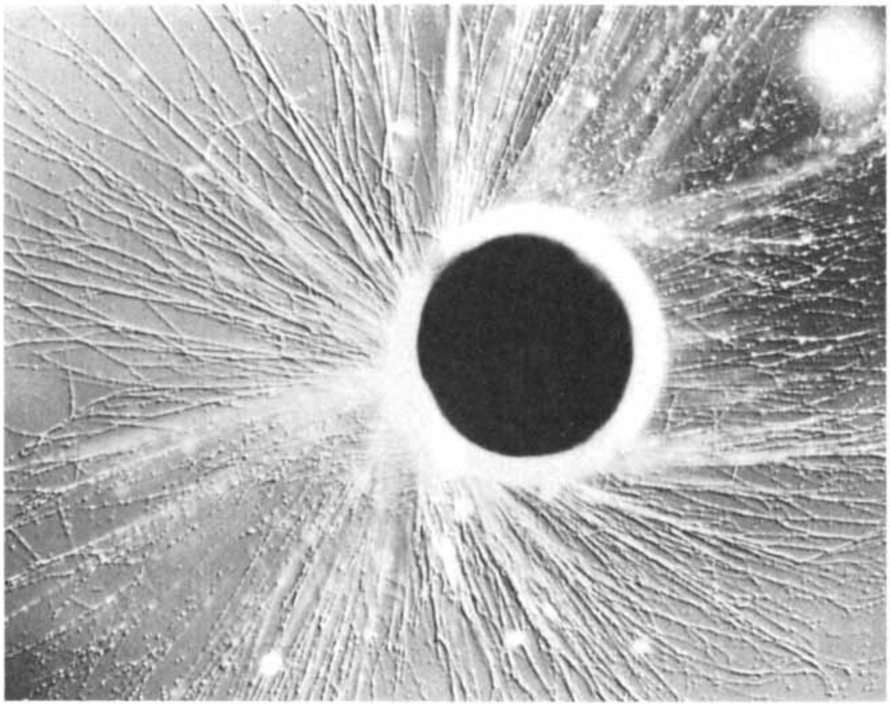
Raymond J. Lasek of the Case Western Reserve University School of Medicine, a leader in the axonal-transport field, and Scott T. Brady, one of his postdoctoral associates, suggested applying a technique developed four decades earlier at Woods Hole by Richard S. Bear, Francis O. Schmitt and John Z. Young. These workers had devised a way to extract the axoplasm, or axonal cytoplasm, from a single axon. The neuron's insides come out as a gel-like cylinder that does not dissolve and so preserves the integrity of its contents.

Brady's adaptation of this technique met with immediate success. Transport continued unabated in axoplasm devoid of any cell membrane and, in the presence of a solution of salts, amino acids and the energy molecule ATP, would remain active for hours, leaving ample time for experimentation. I suggested we "stir" the extruded axoplasm to determine whether the transport mechanism was dependent on the integrity of the axoplasm's architecture. Stirring did not turn off particle movement; it merely disorganized it. And the particles still followed the path laid out by the "linear elements." I grew more confident that isolated microtubules would maintain their activity and vindicate my faith in their role in particle transport.

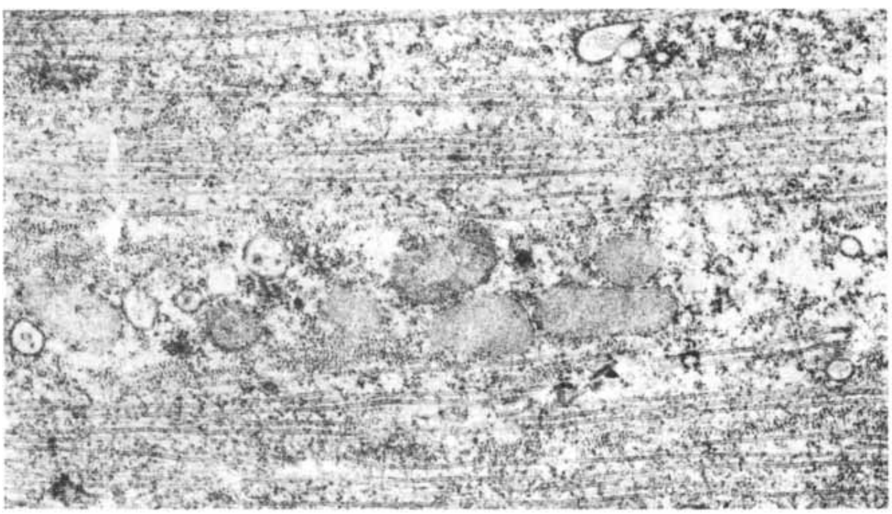
That conviction was considerably substantiated by the work of a postdoctoral trainee in my laboratory at Dartmouth, John H. Hayden. Hayden had prepared from frog eyes a class of specialized connective-tissue cells called keratocytes that had extremely thin peripheries. Within these margins, in living cells, we used AVEC microscopy to see particles moving in both directions along what appeared to be the same "linear elements."

Hayden tested the composition of these elements with antibodies to tubulin provided by Robert D. Goldman of the Northwestern University Medical School. The antibodies bound to the fibers, indicating for the first time that the fibrous tracks the particles were following were made of tubulin and therefore were microtubules.

Could a single microtubule support movements in one direction or in two? In a second set of experiments requiring the utmost skill in specimen preparation, Hayden first recorded movements in both directions along the microtubules with video microscopy, then fixed the same keratocytes for examination with electron microscopy. By comparing the two images, he could see that individual microtubules were shuttling particles in both directions. Here was the first conclusive ev-



INTERLACING NETWORKS of filopodia extend from the dark cell body of *Allogromia*, shown here in a photomicrograph produced by conventional differential interference contrast microscopy. The first direct evidence of particle transport by microtubules was seen in this organism's transparent branches. *Allogromia* measures half a millimeter across.



BUNDLES OF MICROTUBULES in a filopodium of *Allogromia* show a distinct orientation. The stacks align in rows parallel to the long axis of the filopodium. In this electron micrograph loose spirals of tubulin, enlarged 24,000 diameters, look like speckles scattered within the microtubule stacks. Vesicles occupy the spaces between stacks.

idence that a single microtubule could be responsible for the puzzling two-way movement that in *Allogromia* had seemed to defy physical laws. In fact, we thought it highly probable that microtubules could be implicated in all the systems we had studied: foraminifers, tissue cells and neurons alike.

Hayden's advance prompted further questions. Any directed movement in cells needs chemical energy and some kind of "force-generating enzyme" to convert the energy into motion. In cells the commonest source of chemical energy is ATP, which releases stored energy when it reacts with a water molecule in a process called hydrolysis. We assumed the energy for microtubule-dependent transport was supplied by ATP because particle movement grinds to a halt on microtubules deprived of the substance. We wanted to find out more about the FGE's—what they were made of, how they worked and whether they were associated with the microtubules, the particles or both. In order to get the answers we needed to isolate and examine individual microtubules.

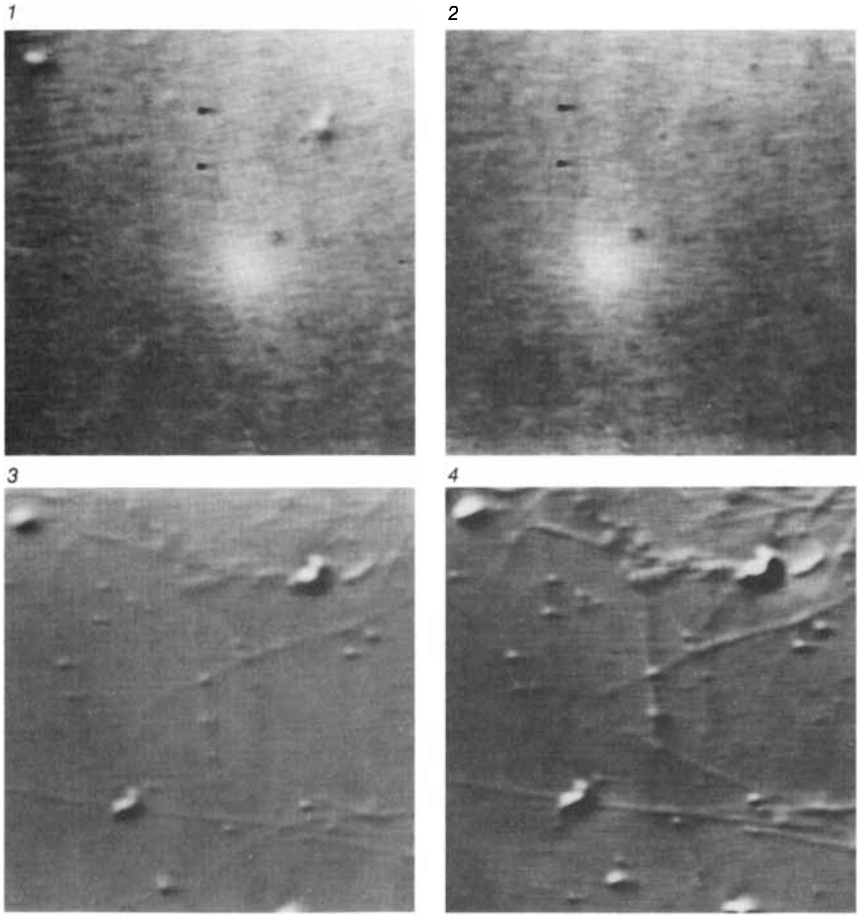
In the summer of 1983, at Woods Hole again, we resumed our investigations of the squid axoplasm. We succeeded in preparing a "diluted" axoplasm in which the long microtubules were separated from one another. In AVEC images of solitary microtubules we could detect the same kind of particle movements Hayden had found inside intact tissue cells. We also noted that the very long microtubules (more than 100 micrometers) could change their shapes slowly while transporting particles, occasionally aggregating in tangles, knots, pretzels and asterisks.

Unfortunately these writhing specimens were so long that they stretched beyond the view of our microscope field, and so we could not tell whether they were colliding with inanimate objects and bending on impact or were merely being deformed by force of interaction with the particles, the slide surface or other microtubules. Along with Lasek, Brady and Dean Martz, also of Case Western Reserve, I had observed that elongated mitochondria in extruded squid axoplasm undergo shape changes that strongly suggest portions of their outer membrane are tugged in different directions by two, three or more attachments to neighboring microtubules. Perhaps a similar tug-of-war was going on among the microtubules themselves.

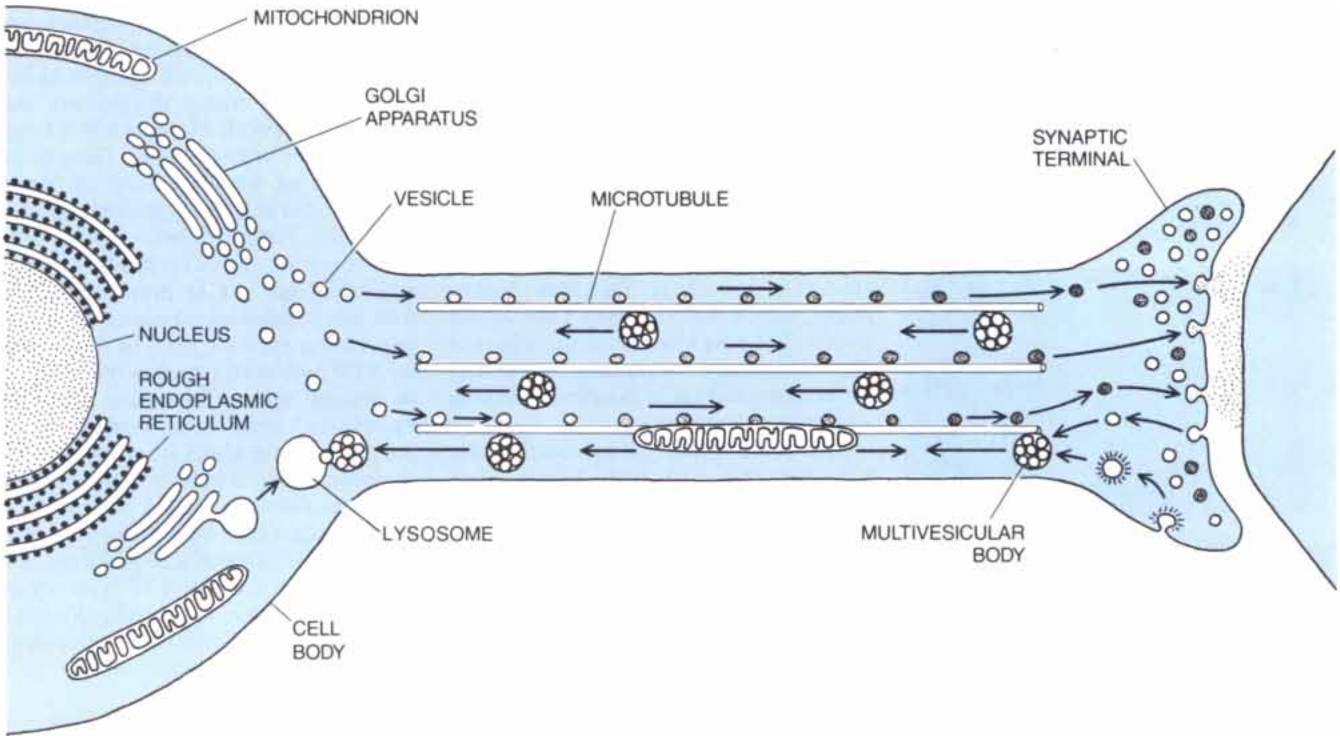
We could not decide among these possibilities until we were able to break the microtubules into shorter segments, a procedure we did not master until the end of the summer. A year later I collaborated with Dieter G. Weiss, who was visiting Woods Hole from the Zoological Institute of the University of Munich, in recording the movements of these fragments. We accumulated some 45 hours of video tapes recorded at normal speed and accelerated 10 times by time-lapse photography. The tapes revealed a tiny world of multifaceted, lively and unexpected movements.

But the most spectacular findings emerged when we succeeded in fragmenting the microtubules, which are usually hundreds of micrometers long, into segments ranging in length from the width of our field of view (25 micrometers) or more down to less than one micrometer. We were astonished to see the short microtubule segments gliding freely across the glass slide. Each piece moved in one direction only, never reversing, and proceeded in a straight line unless it bumped into an obstacle such as an organelle that had stuck to the cover slip. It could bounce off at an angle or, if it hit something head on, it could fishtail in a regular series of serpentine patterns. These gyrations occurred whether or not particles were being transported.

As visually compelling as such unexpected movements were, they were equally important for their contribution to our understanding of FGE's.



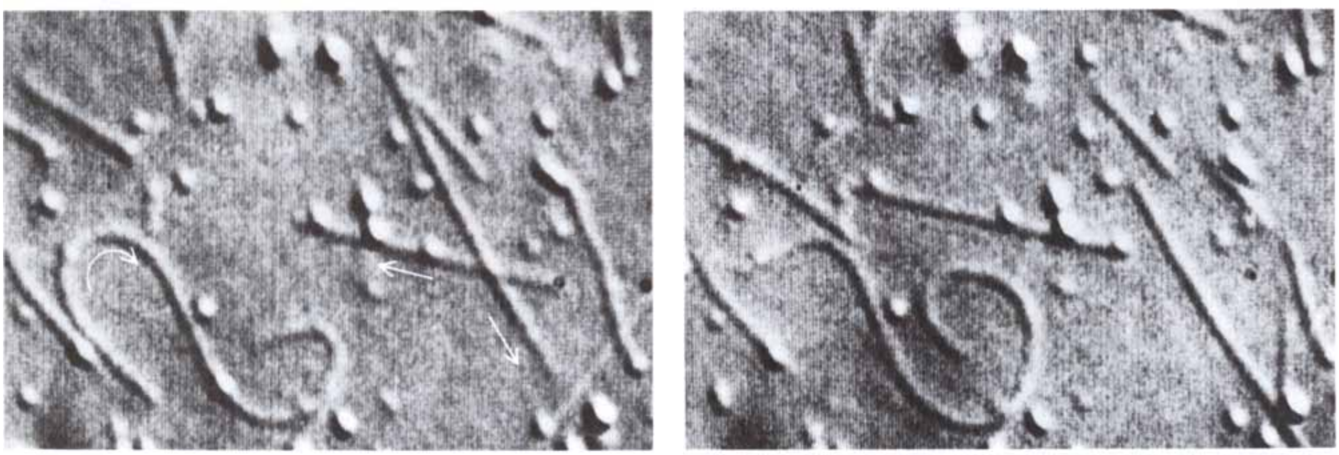
AVEC MICROSCOPY exploits the fact that a television camera can detect weak contrasts better than the human eye. Allen video-enhanced contrast (AVEC) imaging works by amplifying differences in brightness and reducing the average brightness of the final television image. First a light microscope (in this case a differential interference contrast microscope, or DIC) is fitted with a zoom lens to increase magnification. Then the DIC setting is adjusted to admit more light; this step improves the signal-to-noise ratio and optimizes conditions for the television camera, but it makes the image difficult to discern because stray light overwhelms the eye. The image in *1* results when the modulated voltage component of the television signal is amplified, enhancing differences in brightness, and a direct-current voltage, which represents background brightness, is subtracted. Although enhancement makes some elements of the specimen visible (such as the two vesicles in the upper half of the image), it also captures dust particles and lens imperfections that obscure the specimen. Because this mottled pattern remains when the image is defocused (*2*), it can be digitally subtracted from all incoming video frames to yield a clearer view (*3*). Low contrast in the resulting image can be improved by digital enhancement and the signal-to-noise ratio can be raised by averaging the input from a few frames (*4*).



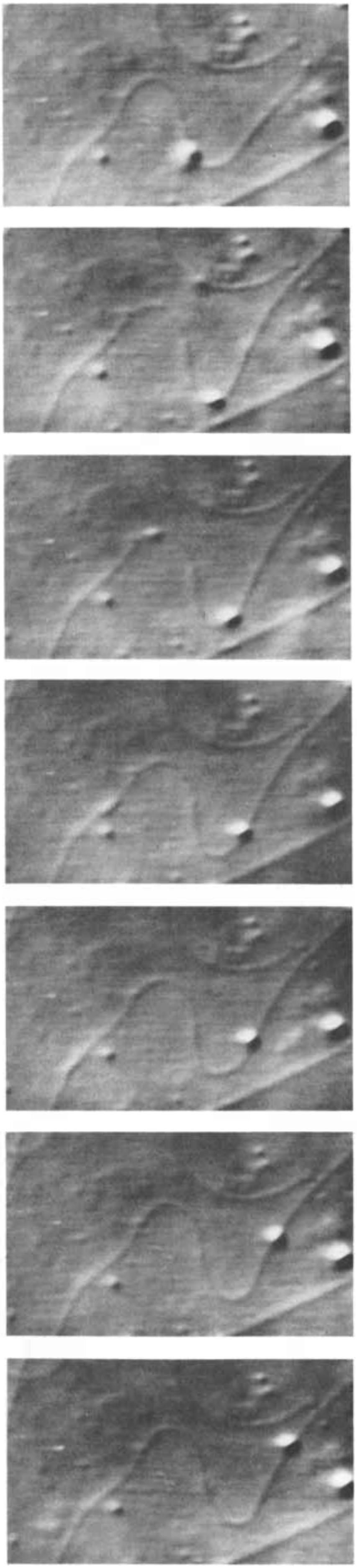
AXONAL TRANSPORT along microtubules allows the swift exchange of substances between the nerve cell body and the synaptic terminal at which the nerve fiber adjoins its target cell. Vesicles that will convey neurotransmitters are manufactured by the Golgi apparatus and carried toward the synapse. Surplus membrane at the synaptic terminal is packaged into multivesicular bodies, which return to lysosomes in the cell body for degradation. Mitochondria move in both directions, catering to the energy demands of the cell. Transport of the vesicles is continuous; that of the mitochondria is intermittent. It is clear that a single microtubule can support transport in both directions, but the transport mechanism is still undetermined. This illustration is not drawn to scale.

Some years earlier other investigators had established how to build microtubules from purified tubulin subunits [see "Microtubules," by Pierre Dustin; SCIENTIFIC AMERICAN, August, 1980]. Soon after we made our observations Thomas S. Reese of the National Institute of Neurological and Communicative Disorders and Stroke and his collaborators at Woods Hole found that such reassembled microtubules would not glide unless a soluble fraction of squid axoplasm was added to the medium. This evidence leads us to believe that the site of action of the FGE is on the microtubule surface and that a reserve supply of FGE must exist in the cytoplasm. The findings also suggest that the FGE is only loosely associated with the microtubule, not intimately coupled with it. Hence we call the FGE a microtubule-interactive protein.

How did our findings mesh with the established views regarding the mechanisms of axonal transport? David S. Smith of the University of Miami and others had shown in electron micrographs that microtubules have "arms," or cross bridges. It had also been demonstrated that some of the vesicles and organelles in nerve axons are connected to microtubules by cross bridges of varying lengths. Sidney Ochs of the In-



GLIDING AND FISHTAILING of isolated microtubules suggest that some kind of force-generating enzyme is directly associated with the fibers themselves. The course a microtubule follows is determined by its initial position, because each microtubule can move in only one direction. When it encounters an obstacle such as an organelle, a gliding microtubule may fishtail, curling on itself or writhing in continuous serpentine patterns. The motion is not sluggish: these AVEC images were made just 18 seconds apart.



PARTICLE MOVEMENT in opposite directions on a single microtubule is depicted in this series of AVEC micrographs. A large particle, presumably a multivesicular body, travels through the curving "valley" of the S-shaped microtubule and ascends the opposite slope. In the third frame a smaller, elongated particle attaches to the top of the microtubule's "bump" and appears to coast down the left side. Thirteen seconds elapse during the series, in which the microtubules are enlarged 9,000 diameters.

diana University School of Medicine had postulated in 1972 that these cross bridges are composed of a protein that would displace the organelle when an energy-dependent reaction caused the protein to change its conformation. Several investigators thought the contractile proteins involved might be actin and myosin, the generators of muscle contraction. But they predicted that, on individual filaments, movement would be stepwise and would proceed in only one direction, whereas we found some continuous movement as well as movement in both directions on single microtubules.

Our initial attempts to visualize the sidearms met with little success, but Brady and Lasek were conducting experiments that offered some hope. They were using an ATP analogue called AMP-PNP to clarify the microtubule transport mechanism. This substance replaces ATP at the active site of the FGE but cannot be hydrolyzed to provide the energy for motion. Brady and Lasek knew that AMP-PNP caused vesicles and organelles to adhere to microtubules and deduced that it probably stabilized the sidearms as well. By treating extruded axoplasm with the analogue, Brady succeeded in capturing the sidearms' image in electron micrographs. From similar preparations Brady, now at the University of Texas Health Science Center at Dallas, isolated an enzyme that hydrolyzes ATP, called an ATPase, that seems to be a good candidate for the FGE. Reese's group, working at Woods Hole, also took Brady's approach but arrived at a different protein.

The "home" position of the sidearms is also at issue. Are the arms that form the cross bridges actually on the microtubules, on vesicles and organelles or freely dispersed in the fluid phase of the cell? Again, different investigators have reached different conclusions. It is apparent, however, that in certain circumstances both microtubules and their cargo can bind the sidearms, and so we have begun to think about how the sidearms and the FGE might interact to induce microtubule-dependent transport.

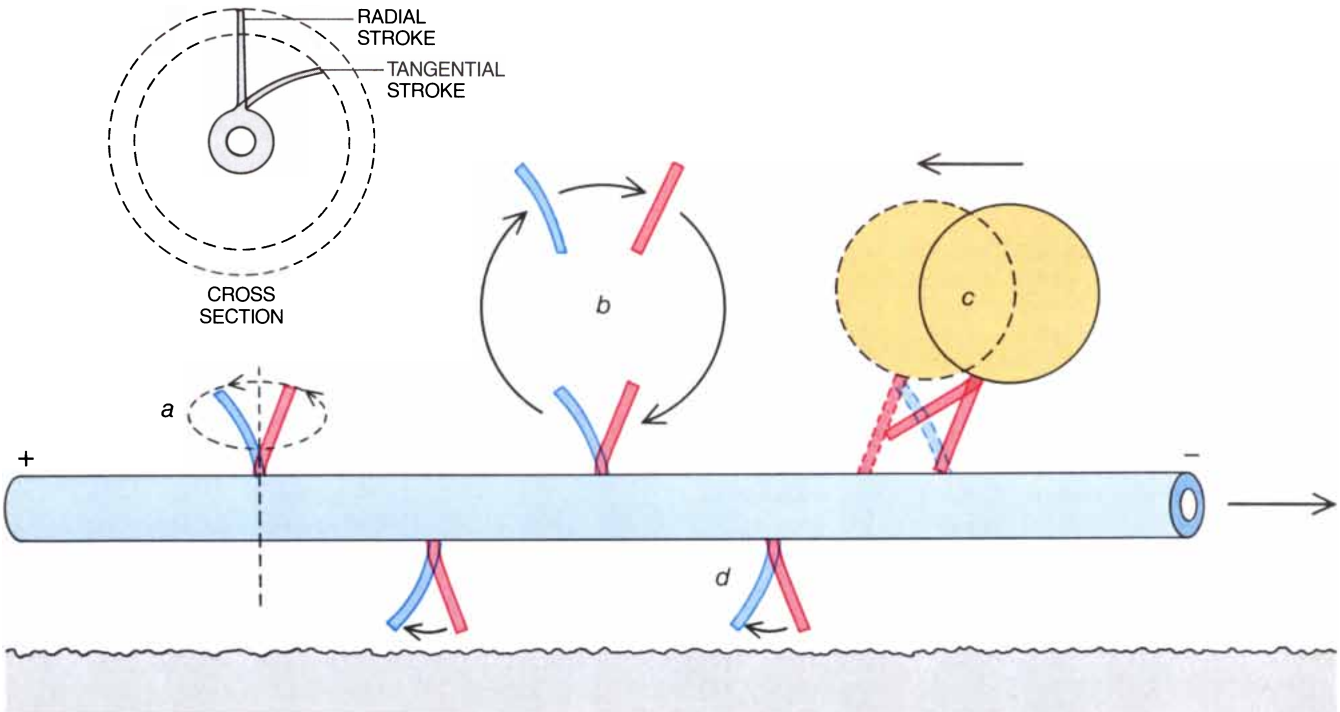
Compiling my own experience and hunches regarding this transport, I developed what I call the "active native microtubule" model during my last sojourn at Woods Hole in 1985. I had as my collaborators Weiss, George M. Langford of the University of North Carolina School of Medicine and Dieter Seitz-Tutter, who, along with Weiss, is now at the Technical University of Munich. Our model proposes that each sidearm contains a force-generating ATPase that, in the course of ATP hydrolysis, causes the sidearm to change its conformation from an "activated" state to a "resting" state. As the sidearm alters its configuration it either forms transient cross bridges between a vesicle and a microtubule or interacts with a nearby surface. The force of the sidearm's conformational change, generated by ATP hydrolysis, is utilized to move the vesicle along the microtubule—or to move the microtubule along an adjacent surface.

Whether the sidearms reside initially on the microtubules, on vesicles and organelles or in free form, we think contact with a microtubule must be necessary to trigger the conformational change, because vesicles never seem to move independently of the microtubule surface. Our model also assumes that the sidearms become activated before they attach to the microtubules. The cycle of conformational changes may be analogous to the actin-myosin mechanism of muscle contraction, which also involves the hydrolysis of ATP.

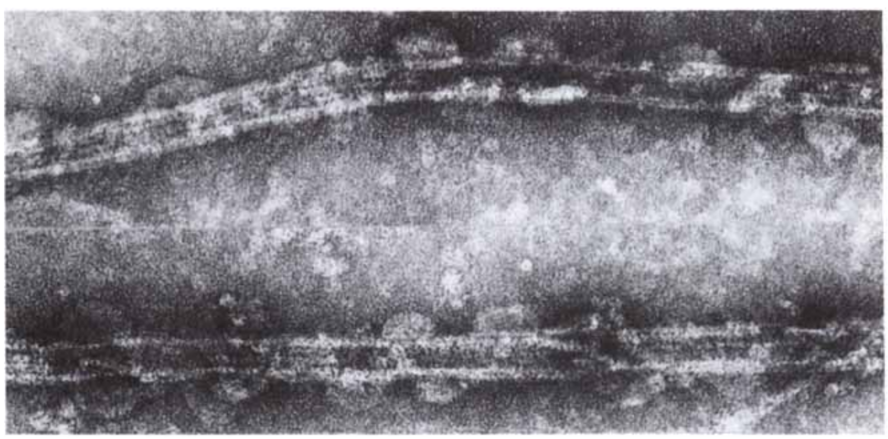
Although we know very little about the mechanics and hydrodynamics of particle translocation, we have found on vesicles certain V-shaped structures (which we interpret as pairs of sidearms) that offer an elegant explanation. While one arm of the V forms a cross bridge and goes through its cycle of conformational change, the other could become activated. Then, as the first finishes its cycle and dissociates, the second would bind, and so on. This "walking" mechanism would keep the particle from being swept away by the vigorous Brownian motion that is constantly threatening its course. Such a mechanism would also require fewer sidearms than a model involving only freely dissociable sidearms.

How can our model account for the simultaneous movement of organelles in both directions? I have introduced the "backstroke hypothesis," a scheme depicting particle transport in two directions by a single sidearm molecule. Each sidearm could undergo a pattern of bending such that the activated stroke propels an organelle along a path relatively distant from the microtubule surface; then the sidearm sweeps around closer to the fiber for the return stroke, which would carry larger organelles in the opposite direction. Particles traveling in different directions would not bump into one another because they would occupy different planes of motion.

The active-native-microtubule model is only hypothetical; there is plenty of room for alternative interpretations. Reese believes, for instance, that two kinds of sidearms exist, each responsible for just one direction of motion. Ambiguities such as these are sure to dissolve in the coming years now that we have the ability to visualize microtubule-dependent transport on the living cytoskeleton. In time we shall gain an understanding of the phenomenon that gripped us so many years ago: the two-way intracellular traffic on what appears to be a molecular one-way street.



MODELS FOR TRANSPORT proposed by the author assume that microtubule gliding and organelle movement are controlled by the same molecule: a force-generating enzyme (FGE) that undergoes a cyclic change of shape. The FGE resides in the sidearms that connect microtubules to organelles or interact with the substratum. It can alternate between an activated state (*red*) and a resting state (*blue*). The resting FGE could become activated while remaining on the microtubule (*a*), while circulating freely (*b*) or while attached to an organelle (*c*). The first model best fits the "backstroke" hypothesis, illustrated in the cross section. This theory suggests that the activated and resting strokes of an FGE occur in two different orientations, one radial and the other tangential, to transport particles in both directions without collisions. The last model (*c*) describes a "walking" mechanism of organelle movement, by which two or more neighboring sidearms coordinate activation and deactivation to keep the organelle in constant association with the microtubule. Because microtubules have an intrinsic polarity (*indicated by plus and minus signs*), sidearm binding may also be directional. Hence an individual microtubule will glide in only one direction, dictated by its own infrastructure (*d*).



VESICLES AND MICROTUBULES can both display the sidearms that have been implicated in microtubule-dependent transport. When chemical fuel in the form of ATP is available, the arms cling to vesicle surfaces (*left*); when the "false fuel" AMP-PNP is substituted for ATP, the sidearms stick instead to the microtubules (*right*). In these images the sidearms, enlarged 210,000 diameters, are conspicuous on the vesicle but barely visible on the microtubule, where they appear as rigid rods between large, globular structures. The electron micrographs were made by George M. Langford of the University of North Carolina School of Medicine.

The Oldest Pulsars in the Universe

These unusual pulsars are dense, compact stars spinning at the rate of several hundred revolutions per second. Why do they spin so fast? They are thought to have been resurrected from an early death

by Jacob Shaham

One sunny morning in the middle of November, 1982, David J. Helfand, an astronomer colleague at Columbia University, came into my office to tell me a remarkable story. He had just returned from a business trip to the Arecibo radio-telescope observatory in Puerto Rico. While he was there he learned that Donald C. Backer of the University of California at Berkeley and his collaborators had discovered a 1.558-millisecond radio pulsar in the constellation Vulpecula. That would make the radio pulsar, whose official name is 1937 + 214, the fastest one known.

Backer's discovery was a startling piece of news indeed. Radio pulsars are starlike celestial objects that generate many kinds of particulate and electromagnetic radiation. Among them are radio waves. When a radio telescope is directed at such a source, the radio waves are registered as periodic "beeps." The fastest radio pulsar known prior to the discovery of 1937 + 214 was the Crab pulsar: a 30-millisecond-period beeper buried inside a large, fuzzy structure resembling a cotton ball and called the Crab nebula. Now there appeared to be something beeping 20 times faster.

Since the discovery of the Vulpecula pulsar two more superfast pulsars have been reported. In 1983 a 6.13-millisecond pulsar (called 1953 + 29) was announced, and in 1986 a 5.362-millisecond pulsar (called 1855 + 09) was publicized. A candidate for a fourth has been mentioned. As more evidence becomes available, it seems increasingly likely that the superfast pulsars can be explained only as a part of a new class of pulsars.

Although many of the details of the class remain obscured, some general facts are emerging. Perhaps most interesting of all is the great age these new celestial objects are thought to have. Ordinary pulsars are relatively young, typically less than a million years old; the Crab pulsar, which is the youngest one known, is a mere infant of 932 years. The superfast pulsars, in comparison, are thought to be ancient. They are probably the result of evolutionary processes that could go back as much as a billion years, or one-twentieth of the age of the universe, and they are likely to live for several billion years more.

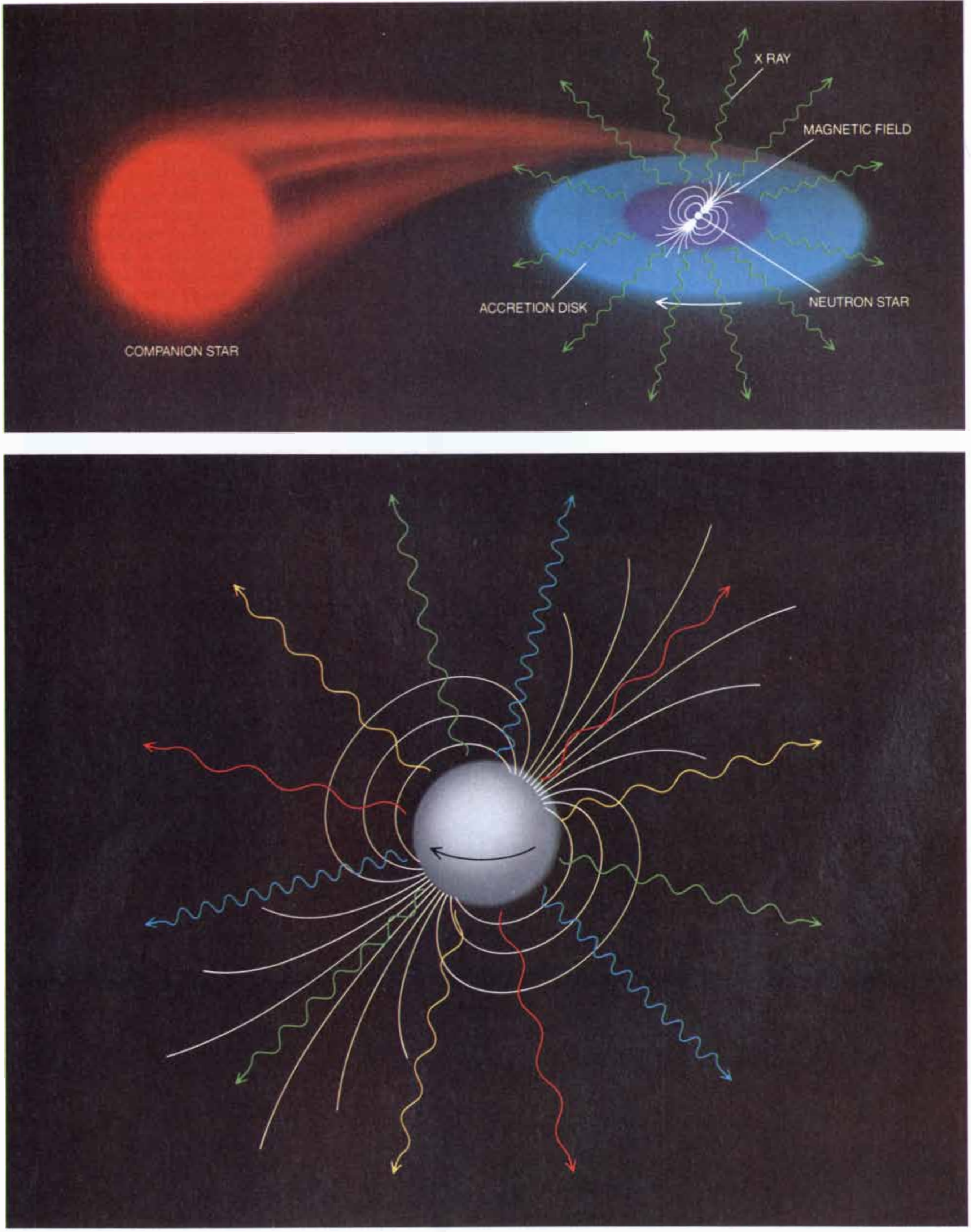
Ordinary radio pulsars are fascinating in themselves. They are thought to be rotating neutron stars: huge, spinning "nuclei" that contain some 10^{57} protons and neutrons. (The nucleus of a hydrogen atom, in contrast, consists of a single proton; an iron nucleus contains a total of 56 protons and neutrons.) The neutron star gets its name from the fact that it has about 20 times more neutrons than protons. The large clump of nuclear matter, which has a mass about equal to that of the sun, is compressed into a sphere with a radius on the order of 10 kilometers. Consequently the density of the star is enormous, slightly greater than the density of ordinary nuclear matter, which is itself some 10 trillion times denser than a lead brick.

Currents of protons and electrons moving within the star generate a magnetic field. As the star rotates, a radio beacon, ignited by the combined effect of the magnetic field and the rotation, emanates from it and sweeps periodically through the surrounding space, rather like a lighthouse beam. Once per revolution the beacon cuts past the earth, giving rise to the beeping detected by radio telescopes. The Crab pulsar, for instance, is a star as massive as the sun squeezed into a shell the size of an average city and spinning at the high rate of 33 times per second. (Every 30 milliseconds the Crab pulsar makes one complete revolution.)

One can amuse oneself for many hours thinking about the properties of a neutron star. One can easily calculate, for example, the force of gravity at the surface of a neutron star. It turns out to be some 100 billion times greater than the force of gravity at the surface of the earth. Indeed, the gravitational field is so strong that any object drawn toward a neutron star would fragment even before impact: objects cannot fall on a neutron star, they can only "rain" on it.

What processes could give rise to the birth of such an extraordinary star? Astrophysicists believe a neutron star is formed when an ordinary star, which has a mass several times greater than that of the sun, collapses violently. The event is called a supernova. The inner parts of the ordinary star implode to form the neutron star while the outer parts explode to form a surrounding nebula, or cloud of gas and dust. Some of the energy from the explosion is released as visible light. The display can sometimes be quite spectacular. In A.D. 1054, for instance, when the supernova that gave birth to the Crab pulsar and Crab nebula occurred, the light display was visible during the daytime for nearly three weeks and visible during the nighttime for more than a year. Energy continues to be released even after a pulsar is born: the spinning star ejects high-energy particles that are absorbed by the surrounding nebula, making it glow. As the pulsar ejects the particles it loses rotational energy and slows down. A general rule of thumb follows that the older the pulsar, the slower its rate of spin.

One might therefore expect that the recently discovered superfast pulsars must be very young—certainly younger than the 932 years of the Crab pulsar. In late 1982 and early 1983 M. Ali Alpar of the Research Institute for Basic Sciences in Turkey, Andrew F. Cheng of Johns Hopkins University and Malvin A. Ruderman of Columbia University joined me in analyzing the data from the superfast pulsar



MILLISECOND PULSAR designated 1937 + 214 is believed to have originated differently from other pulsars. It is thought to be a resurrected, or recycled, dense neutron star. Perhaps some dozens of million years ago or more the star was part of an X-ray binary: a spinning neutron star orbited by a companion star (*top*). As the stars orbited each other, the neutron star slowly accreted matter from its companion. The impact of the incoming matter on the surface of the star released X rays, and the star began to spin faster. When the accretion eventually stopped, and after the companion star had accreted itself "to death," the neutron star was presumably spinning fast enough to be a radio pulsar (*bottom*). Here the red wavy lines represent radio waves; yellow, visible light; green, X rays, and blue, gamma rays. At present it is not understood why the companion star should have disappeared.

1937 + 214 collected by Backer and his colleagues (S. R. Kulkarni and Carl E. Heiles of Berkeley, M. M. Davis of the Arecibo Observatory and W. M. Goss of the Kapteyn Laboratory in the Netherlands). We calculated, as others have, that the pulsar was rotating so fast that it was a factor of only two or so away from actually flying apart owing to centrifugal forces. It was clear that the star could not have been rotating much faster when it was born and therefore could not have slowed down by much.

A statistical consideration, however, would ultimately lead to the conclusion that even though the superfast pulsar has a high rate of spin, it is probably an object of great antiquity. In other words, it appears to have been formed during a long evolutionary process, not during the relatively quick explosion that is characteristic of a supernova.

The statistical consideration pertains to how fast a radio pulsar slows down. Simply stated, the shorter the period of a pulsar is, the more it slows down; the longer the period is, the less it slows down. As a consequence there should be more slow pulsars than fast ones. The existence of the 1.5-millisecond pulsar 1937 + 214, unless it is some freak accident, implies that if pulsars are born at a fairly constant rate over time, there should now be many pulsars that have slow periods of between 10 and 30 milliseconds. Yet in spite of a fairly extensive search of the heavens, no such pulsars have been found. Moreover, even though 1937 + 214 has such a short period, it is barely slowing down. Why has it continued to spin so fast?

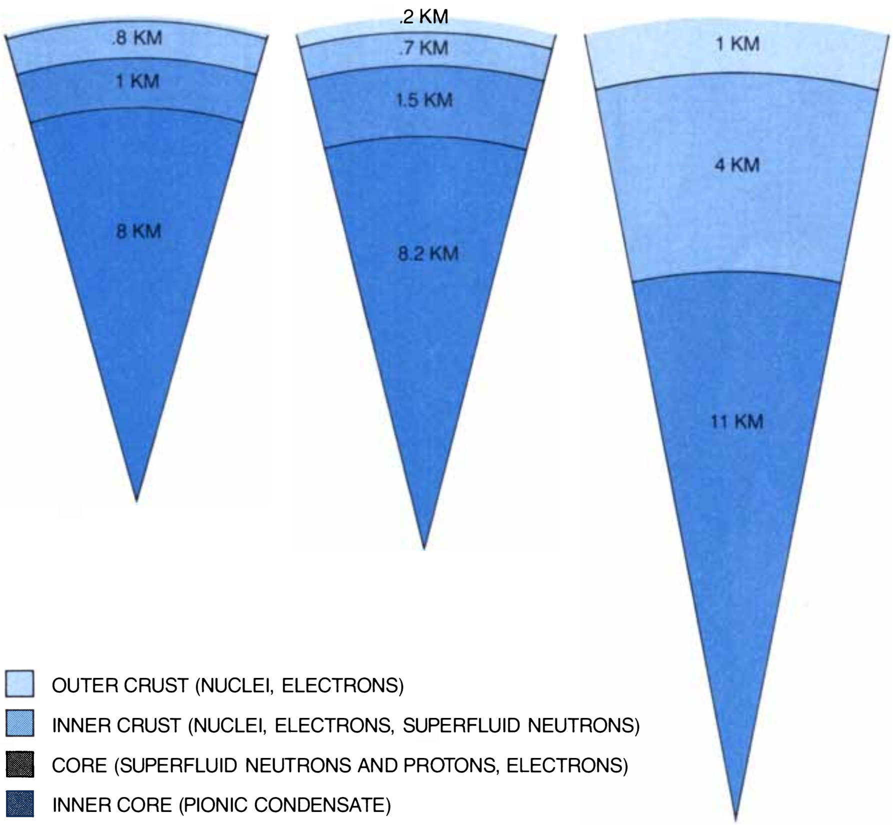
To account for the discrepancy between theory and observation, we made use of the fact that the rate at which a pulsar slows down also depends on the strength of the magnetic field at the surface of the star. The greater the magnetic field is, the more energy the pulsar radiates and the faster it will slow down. A typical neutron star has a magnetic field of about a trillion gauss, which is roughly 100 million times greater than the field of a heavy-duty magnet. We showed that if a 1.5-millisecond pulsar had a magnetic field 1,000 times weaker (about a billion gauss), it could conceivably require a time comparable to the age of the universe to slow down to a period of 10 milliseconds. This could explain why no pulsars have been seen in the 10-to-30-millisecond range: they simply would not yet have had time to slow down.

A few simple calculations also show that if the magnetic field of a pulsar is relatively weak, the star may even "turn off," or stop radiating, before it can slow down greatly. An ordinary pulsar—one that has a magnetic field of a trillion gauss—turns off when the spin rate is lower than about one revolution per second. A pulsar that has a magnetic field of 100 million gauss, in contrast, would turn off if its spin rate were lower than one revolution per 10 milliseconds. In other words, even if a pulsar with a weak magnetic field could have somehow slowed down to 10 milliseconds, it would not be visible anyway.

All these considerations amount to a remarkable result: the mere existence of pulsar 1937 + 214 means there must be a new class of radio pulsars—very distinct from mainstream radio pulsars—that have weak surface magnetic fields.

The realization that there is a new class of radio pulsars with weak magnetic fields took a great load off our hearts. Observations of the radiation coming from 1937 + 214 and its vicinity indicate that the pulsar gives off comparatively little energy. Yet if an ordinary, high-magnetic-field pulsar were spinning as fast as 1937 + 214, its "slowing-down time" (the time required for its period to double) would be a few years and the star would radiate enormous amounts of energy: it would be several million times more luminous than the Crab pulsar. We calculated that as a weak-field pulsar, 1937 + 214 has a slowing-down time of nearly a billion years and its energy output is less by a factor of some 100 million, which is consistent with observation. Thus our theory successfully explains what would otherwise be an enormous discrepancy.

We were then faced with a new twist in our analysis. How could a pulsar with a weak magnetic field be created? It so happened that approximately two weeks before we heard about the fast spin of 1937 + 214, new work by Roger D. Blandford of the California Institute of Technology, James H. Applegate of Columbia and Lars Hearnquist of Berkeley came to our attention. They showed that if a pulsar is formed at a high temperature—above about 100 million degrees Kelvin, which is actually below the characteristic temperature of a supernova—the pulsar will automatically acquire a high magnetic field. The implication of their conclusion was clear:



NEUTRON STAR is shown in cross section for each of three theoretical models. A neutron star consists of neutrons and protons packed so tight that its density exceeds even that of ordinary nuclear matter, which is itself some 10 trillion times denser than a lead brick. The star depicted here has a mass 1.4 times as great as the sun's, yet its radius is only somewhere between 10 and 16 kilometers, which is the size of an average city.

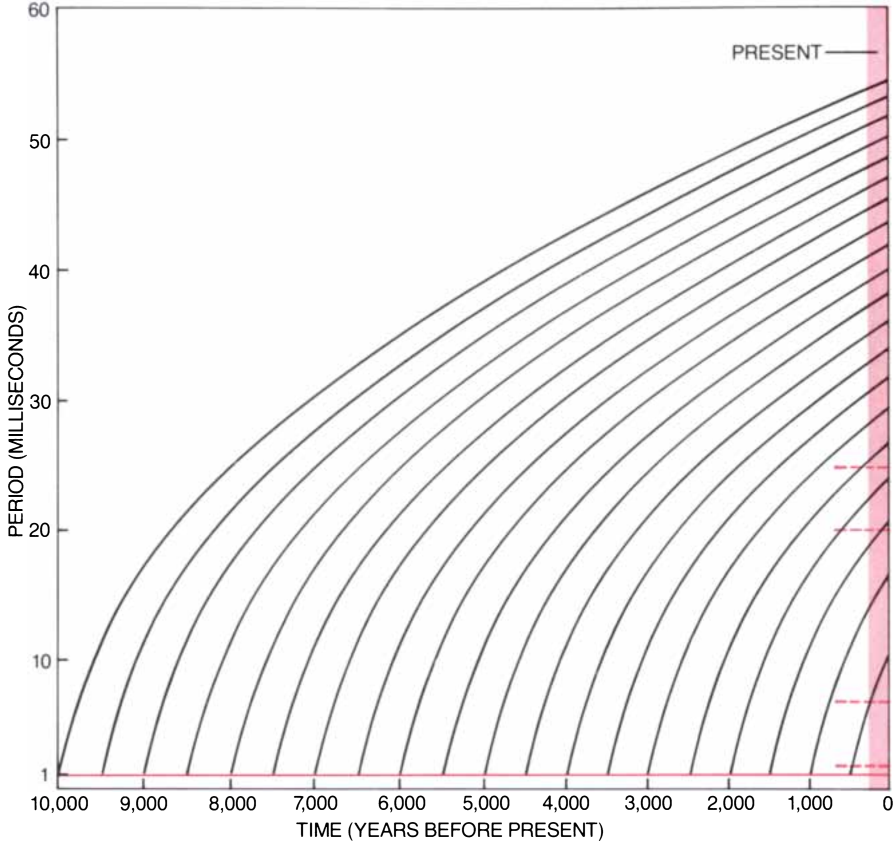
pulsar 1937 + 214 must have been formed "cold" (at temperatures below 100 million degrees), which indicates that it could not have been created by a supernova. Why? Because a "hot" formation implies a large magnetic field and a rapid rate of slowdown, which means that a 1.5-millisecond pulsar with a small magnetic field could not have been formed.

The only known way that a cold neutron star could spin fast is if it accreted slowly rotating material. The basic phenomenon is familiar to anyone who has watched an ice skater achieve fast spins. First the skater spins slowly with outstretched arms, and then he or she draws the arms in, causing the spin rate to go up. For the same sort of process to occur with a cold neutron star the rate of accretion must be low or the star would overheat and the magnetic field would jump too high. As a consequence pulsar 1937 + 214 must have been in the making for a long time, certainly for at least 10 million years and probably on the order of a billion years.

The slow accretion could have proceeded in one of two ways. According to one scenario, a so-called white-dwarf star (which has the mass of a neutron star but a radius 1,000 times greater) accreted matter from a companion star orbiting it. As the matter accumulated, the white dwarf collapsed into a neutron star. If the collapse were slow enough, the neutron star could have gradually "spun up" and become a millisecond pulsar. By accreting additional matter from its companion, the spin of the neutron star would have increased until it reached an equilibrium period.

Alternatively, a neutron star could first have been formed in the usual way (by a supernova) and behaved like a regular pulsar. After some 10 million years most of its crustal electric currents would have decayed, leaving a small magnetic field. If the "dead" pulsar were tightly orbited by a companion star, it could have slowly accreted the necessary material to acquire a fast spin and be resurrected, or recycled. Here and in the preceding scenario the companion would have to have been immolated in the process, since no star is found in the vicinity of 1937 + 214.

Underlying both scenarios, it is held, is a common phenomenon that guarantees the accretion will not proceed too quickly. As incoming electrons, atoms and molecules hit the surface of a star, some of their energy is transformed into light and other electromagnetic radiation. The radiation travels outward, striking other incoming electrons, atoms and molecules.



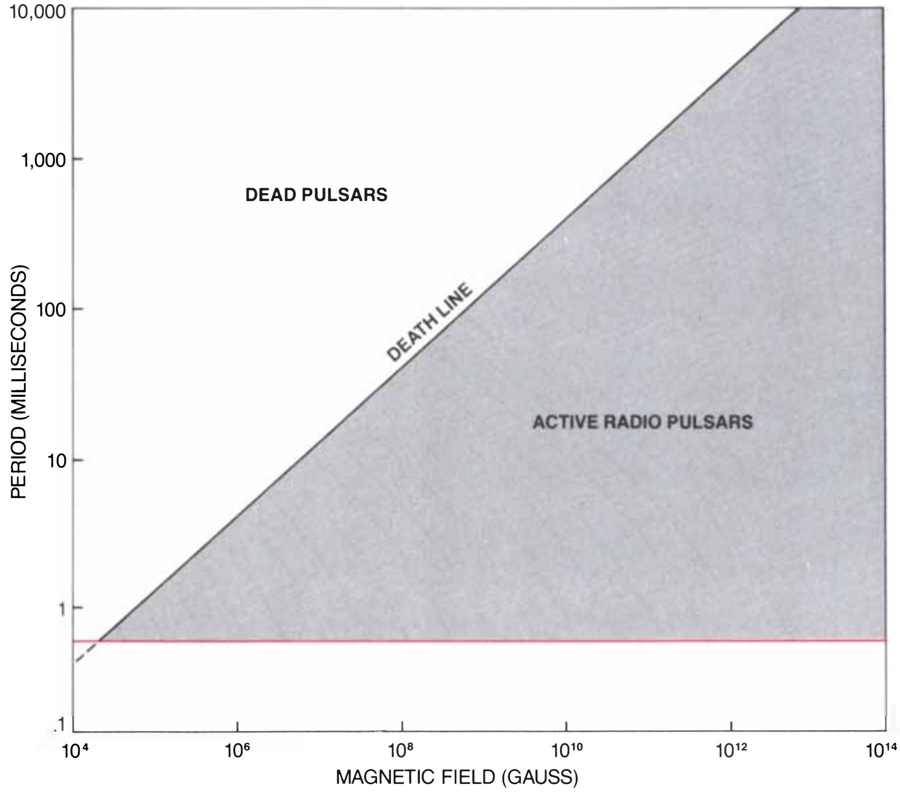
STATISTICS show that the millisecond pulsar 1937 + 214 must have a magnetic field that is much weaker than the field of an ordinary pulsar. The graph, which was prepared with the underlying assumption that the millisecond pulsar has a magnetic field of about a trillion gauss, traces the evolution of a family of millisecond pulsars (*black curves*). The uniform spacing of the curves reflects the supposition that pulsars are born at a fairly constant rate over time. As a pulsar ages, its period becomes longer: it slows down. In general the period of a pulsar changes roughly at a rate that is inversely proportional to the value of the period. In other words, the shorter the period of a pulsar is, the more it slows down; the longer the period is, the less it slows down. As a consequence the existence of the millisecond pulsar (*lower set of broken lines*) implies that there should be several pulsars with periods of 20 milliseconds (*upper set of broken lines*). No such pulsars have been found, however. The discrepancy can be explained by positing that the millisecond pulsars have weak magnetic fields, on the order of a billion gauss. Since their fields are weaker than those of ordinary pulsars, they radiate less energy and could require a time comparable to the age of the universe to slow down to a period of 10 milliseconds. They may also "turn off," or stop radiating, by the time their period is 10 milliseconds.

Because radiation carries momentum, it exerts a force on the incoming matter. The radiation force is directed outward and therefore opposes the inward gravitational pull of the star.

If the accretion rate becomes high enough (6×10^{17} grams per second, or 10^{-8} solar mass per year), the radiation force is so strong that it balances the gravitational pull and the accretion rate cannot go higher. The accretion rate at which the two forces balance is called the Eddington limit, after the English astronomer Sir Arthur Eddington. In simple terms, the Eddington limit means that any accretion process onto a neutron star has a built-in thermostat that ensures the rate of accretion is never too high. If enough matter is available for accretion, the rate will also never be too low. The thermostat would seem custom-tailored for the formation of millisecond pulsars, because limiting the accretion rate means limiting the stellar temperature to 100 million degrees or less.

Such an accretion-regulated process is already known to play an important part in the formation of "compact X-ray sources": small, rotating neutron stars that emit X rays. One kind of compact X-ray source is called a low-mass X-ray binary, in which the neutron star accretes mass from a light companion star. The accretion can last for 100 million years or more.

The fact that low-mass X-ray binaries have such long accretion epochs raised an interesting question: Could they be the progenitors of millisecond pulsars? We decided they



RADIO PULSARS turn off when their surface magnetic field is too weak or when their period of rotation is too long. The diagram shows the boundary between active and "dead" radio pulsars for various fields and periods. Note that because a neutron star will fly apart if its period is shorter than one millisecond, the surface field of an active pulsar must be at least 100,000 gauss, which is roughly 10 times the field of a powerful magnet.

simply had to be. Both the collapsing-white-dwarf scenario and the dead-pulsar scenario require a long-lived accreting binary star, and the only known candidate is the low-mass X-ray binary.

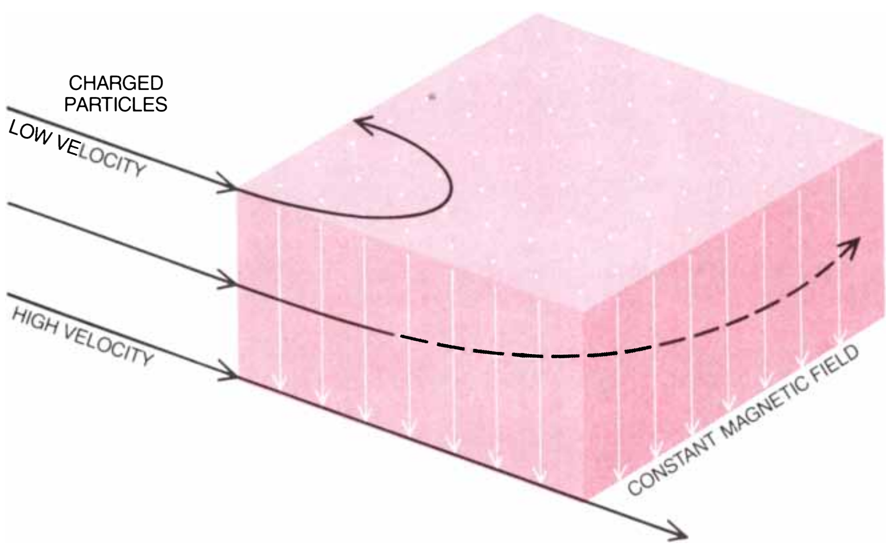
Statistical considerations seem to support the idea that low-mass X-ray binaries are the ancestors of millisecond pulsars. To give a flavor of the argument, I shall first discuss a simple example. Assume that all human beings live for 80 years. I shall define children as humans of age 10 or less; all the rest are adults. On the average, what should the ratio of adults to children be? Clearly there should be seven adults for every child, because each adult covers seven times the time span of a child.

Returning to stars, millisecond pulsars (the "adults") must have a very long life before they turn off; for purposes of this calculation I shall assume that they live for some 20 billion years, which is the age of the universe. Low-mass X-ray binaries (the "children") are estimated to have a relatively short life in comparison, about 100 million years. The statistics now indicate that the ratio of millisecond pulsars to low-mass X-ray binaries should be roughly 20 billion to 100 million, or 200 to 1.

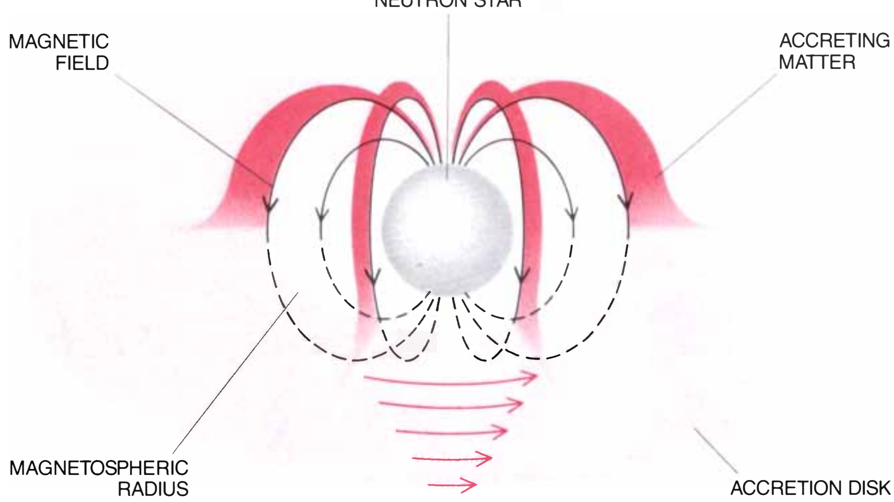
The conclusion is consistent with observation. I mentioned above that a total of three millisecond pulsars have been found so far. More are certain to be discovered, and so it is reasonable to expect that there are probably several thousand millisecond pulsars lurking undetected in the galaxy. (The pulsars cannot be seen from the earth unless they are close enough.) One would therefore expect to see a few dozen low-mass X-ray binaries in the galaxy, which is indeed what one finds.

Statistical considerations are naturally reassuring, but perhaps more important to us is that we were able to employ only basic principles of physics to develop a model that accurately predicts how fast the rate of spin of pulsar 1937 + 214 should change. According to our theory, 1937 + 214 is a resurrected radio pulsar. It began as a binary and was spun up by accretion. When the accretion stopped, the star had a high enough spin to become a radio pulsar in spite of its relatively weak magnetic field. We determined that its period should increase at the rate of 10^{-20} to 10^{-19} second per second, which is many orders of magnitude less than what one might expect for any fast pulsar known before 1982. Our prediction has been confirmed: the measured value of the rate at which the period increases is 1.24×10^{-19} second per second, with an error of $.25 \times 10^{-19}$ second per second.

A complete resolution of the mystery of 1937 + 214 is not yet in hand, however. In particular, if the pulsar was spun up by accretion, where is its companion star? The question became more vexing in 1983 with the discovery of the 6.133-millisecond pulsar 1953 + 29 (by Valentin Boriakoff of Cornell University, R. Buccheri of the National Research Council of Italy in Palermo and F. Fauci of the University of Palermo), which does have a binary companion.



CHARGED PARTICLE moving in a direction perpendicular to a magnetic field is deflected by the so-called Lorentz force in such a way that it would "like" to move in a circle. As long as the magnetic field is weak and the particle has high velocity, the deflection is not substantial (*bottom*). As the ratio of field strength to velocity increases, larger deflections occur (*middle*), until finally the particle is turned back by the field (*top*).



ACCRETION of material onto a neutron star occurs primarily on the polar caps because it is there that most magnetic field lines begin and terminate. Charged particles moving parallel to the field lines are not deflected, and so they tend to accumulate at the poles. Charged particles moving perpendicular to the field lines are deflected, however, although collisions with other incoming particles keep them from turning back. The point at which the particles tend to be stopped is called the magnetospheric radius. In the region inside, "the magnetosphere," particles move mostly along field lines and rotate with the star.

Soon after the discovery, Helfand, Ruderman and I set out to ponder why 1953 + 29 should be part of a binary system and 1937 + 214 should be isolated. We made use of the fact that there are two types of low-mass X-ray binaries: bright ones and weak ones. Bright binaries can be understood only in terms of a "normal" star swelling up to become a red giant and subsequently losing its outer envelope of material to a companion neutron star. A weak binary, in contrast, might result if two stars were drawn together by losing orbital energy to gravitational radiation.

In the past five years or so a number of investigators have shown that if a bright low-mass X-ray binary gives birth to anything, it can only be a binary system—in this case a pulsar and a leftover core of the "normal" star. (Among the investigators reaching this conclusion are Paul C. Joss and Saul A. Rappaport of the Massachusetts Institute of Technology, Edward P. J. van den Heuvel and G. J. Savonije of the Astronomical Institute in Amsterdam, Ronald E. Taam of Northwestern University and Ronald F. Webbink of the University of Illinois at Urbana-Champaign.) We consequently hypothesized that 1953 + 29 may have started out as a bright low-mass X-ray binary and 1937 + 214 may have originally been a weak one.

The situation became more complex in 1986 with the discovery of the 5.362-millisecond pulsar 1855 + 09 (by D. J. Segelstein, L. A. Rawley, D. R. Stinebring, A. S. Fruchter and Joseph H. Taylor of Princeton University), which also has a binary companion. The histories of 1855 + 09 and 1953 + 29 appear in principle quite similar both to each other and to those of five other binary pulsars that, while they are not superfast, have periods of several tens of milliseconds. As a result 1937 + 214 really begins to be conspicuous in not having a companion. Several scenarios to account for how the companion might have disappeared have been proposed (including at least one in which the pulsar is held to be relatively young, about a million years old). To date, however, none is entirely successful, and the quest for a satisfactory explanation continues.

Perhaps the most crucial missing link in our story was and still is the fact that no observed low-mass X-ray binary seems as yet to have exhibited a superfast neutron star as one of its two members. Indeed, if a low-mass X-ray binary is to give birth to a superfast pulsar, one would expect it to contain a rotating neutron star that has a period of a few milliseconds. Why have no such periodicities been seen?

Several reasons suggest themselves. The most plausible one pertains to the fact that any neutron star found in a low-mass X-ray binary is likely to be quite old. Its magnetic field will therefore be a relic of an earlier and stronger magnetic field, which is just what one needs to get millisecond periods by the accretion process. David S. Eichler and Zhengzhi Wang of the University of Maryland at College Park have shown that the strength of the relic magnetic field varies widely over the surface. Consequently whereas an ordinary neutron star has two magnetic X-ray hot spots (polar caps from which X rays emanate), an old neutron star may have many X-ray hot spots on its surface.

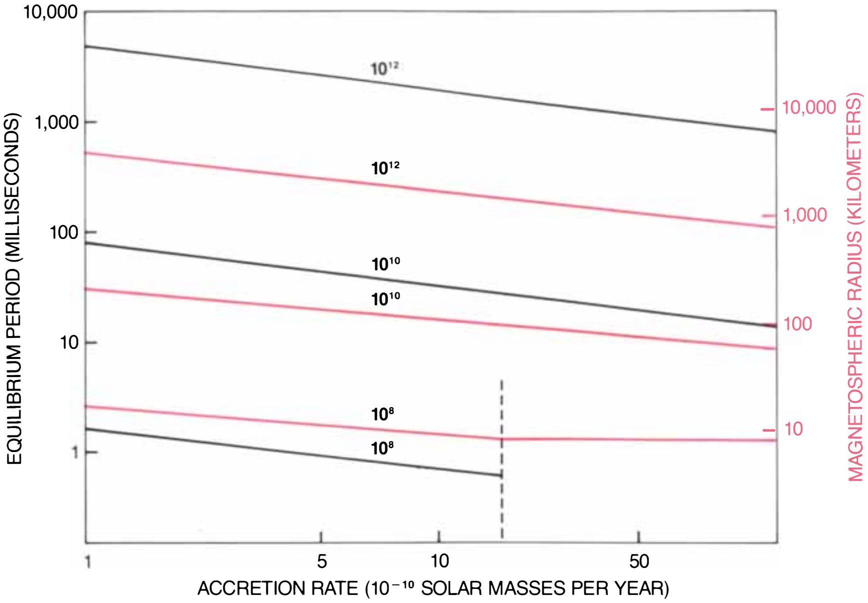
If the star rotates and an X-ray telescope is aimed at it, the hot spots will show up as variations of X-ray intensity. The variations will not be very strong or sharp, however; they will look like erratic flickerings at much higher frequencies. Even though the intensity pattern will repeat once per stellar revolution, it will be so smeared out that the period, and hence the spinning star itself, may not be detectable against the background of X-ray noise from that source.

Other effects might also make the reputed rotation of the neutron star of a low-mass X-ray binary difficult to see. Accreting matter from the companion star might scatter the X rays from the hot spots, thereby reducing the amount of modulation. In addition the neutron star probably acts as a "gravitational lens": it bends the X rays coming out of it, further reducing the degree of modulation.

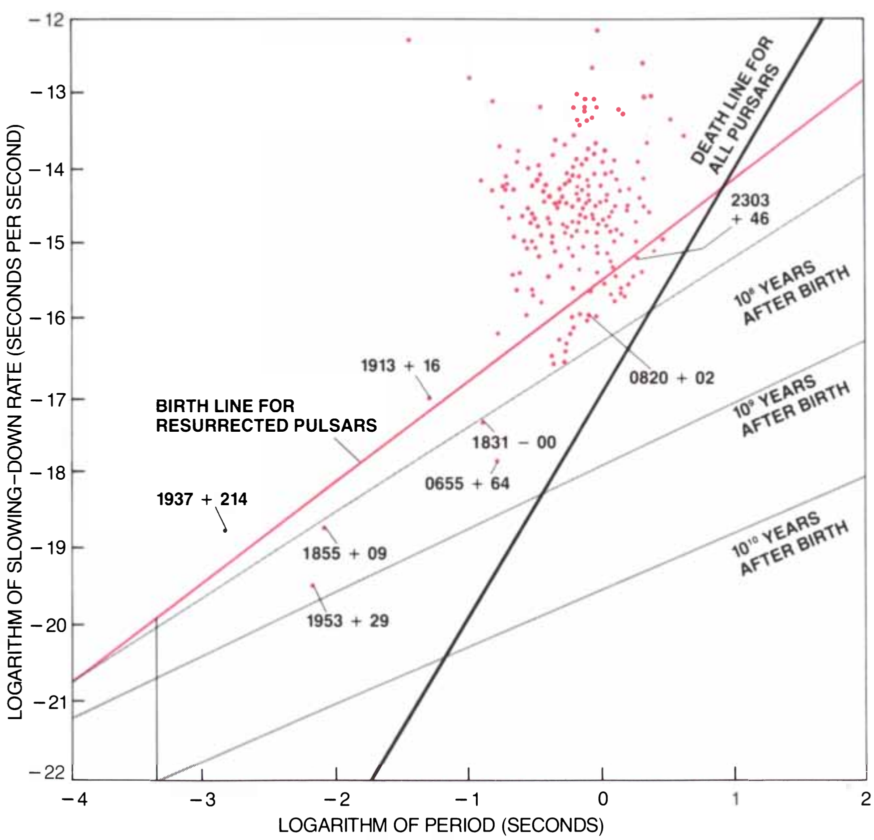
Even though such considerations are well taken, one would still be glad to actually find a spinning neutron star in a low-mass X-ray binary. In 1983 the European Space Agency launched *EXOSAT*, an X-ray observatory satellite, into orbit around the earth. One of the experiments on the satellite was to systematically search for very short X-ray periods in low-mass X-ray binaries. Although no single well-defined period was actually found, some quasi periodicities, or intermittent flickerings of radiation, were picked up in the galactic source called GX5-1. The time between "bursts" ranged from 25 to 50 milliseconds.

The quasi periodicities have some unusual properties, as Michel van der Klis of the European Space Agency and his colleagues (F. Jansen of the Laboratory of Space Research, J. van Paradijs and M. Sztajno of the Astronomical Institute in Amsterdam, J. Trümper of the Max Planck Institute for Physics and Astronomy in Munich and W. H. G. Lewin of M.I.T. and the Max Planck Institute) reported in International Astronomical Union circular No. 4043 on March 13, 1985, and subsequently in *Nature*. First, the quasi periodicities are not constant; in fact, they show large variability. The quasi-periodic signal is strongest at the longest quasi period observed and falls off consistently for shorter periods. The most intriguing feature of the phenomenon is that the total luminosity of GX5-1 correlates directly with the length of the period: the shorter the period is, the brighter the source shines.

Several weeks before the IAU circular came out, Alpar, who was then at



EQUILIBRIUM PERIOD AND MAGNETOSPHERIC RADIUS are plotted as a function of accretion rate for various magnetic field strengths. The equilibrium period is the period at which matter at the magnetospheric radius rotates with the star. The calculations are for a 1.4-solar-mass star (a star that has a mass 1.4 times that of the sun).



OBSERVED SLOWDOWN RATES of various pulsars are plotted against their observed periods. Most pulsars reside in the region of the diagram at the top right. They represent the standard class of ordinary pulsars created during a supernova: the explosion of a star. As the pulsars age, they drift down and to the right until they hit the "death line" (*see top illustration on page 54*). Six pulsars have been found that lie outside the main group. They (and two others that happen to lie within the main group) belong to a new class of pulsars that were spun up by a slow accretion process. All the members of the new class, which are identified by name, lie above the death line, of course. All except 1937 + 214 have binary companions. The colored and gray diagonal lines designate various stages in the life of pulsars that were spun up at an accretion rate of 10^{-9} solar mass per year. The vertical line indicates the shortest period a neutron star can have before it flies apart.

Illinois, called to tell me about it. A friend of his in Europe had attended a seminar on the findings, and the rumors then spread. As bizarre as the findings were, they sounded genuine, and we knew they contained a significant clue to our story of the superfast pulsars. We worked hectically to develop a model that would account for the observations.

In our model the quasi periodicities resulted from matter revolving about the neutron star and attempting to "hook" onto the magnetic field, which rotates with the star, of course. The "bursts" in themselves come at a frequency that equals the difference between the orbital frequency of the accreting matter and the rotating frequency of the star. The spread in the rates of orbital frequencies leads to quasi-periodic behavior rather than well-defined behavior. According to our analysis, GX5-1 must spin at a rate of about six to 10 milliseconds and have a surface magnetic field of roughly a billion gauss. We became excited indeed, because both values closely match those of the binary pulsar 1953 + 29.

How wonderful the news was! At last there was a clear indication of a millisecond periodicity inside an active low-mass X-ray source. The entire picture of the formation of millisecond pulsars seemed to come together.

X-ray astronomy has been thrown into a whirlwind with the discovery of the quasi-periodic X-ray source GX5-1, which is now known most generally as a quasi-periodic oscillator. To date more than 10 quasi-periodic oscillators have been discovered, of which seven exhibit short periods (from 25 to 250 milliseconds). All seven are associated with highly luminous low-mass X-ray binaries. The textbook behavior of GX5-1 has turned into a large zoo of phenomena in the other quasi-periodic oscillators. All the complications are certain to trigger further discussion and modeling of quasi-periodic oscillators and millisecond pulsars in coming years.

Much more remains to be done. Yet a fascinating picture of the resurrection of superfast, long-lived pulsars from ancient neutron stars is emerging. The picture brings with it a new perspective on some of the oldest stellar systems in the galaxy. Pulsars have already led to fresh insights in such areas as particle, nuclear, solid-state, plasma and superfluid physics as well as electromagnetism and general relativity. Now the oldest pulsars light up the path of evolution of the galaxy. "Twinkle, twinkle, little star, we still wonder what you are."


YOU HAVE TO BE BLIND TO GET ON THIS PLANE.
Or you have to be an eye surgeon.
That's because Project ORBIS is a flying eye hospital. And its mission is to train doctors while treating patients throughout the world.
Last year alone, ORBIS helped over 1000 blind men, women and children regain their sight.
It also helped more than 1000 surgeons learn and use the latest ophthalmic techniques. Techniques that restore sight or save it.
That's important. Important because at least thirty million blind people and 350 million other people with potentially blinding diseases can be cured.
That's right, they can be cured. All of them. They just need more doctors with more training.
And ORBIS provides that training to those doctors. Right on board its aircraft.
What's more, as ORBIS spreads its treatment and training, it spreads goodwill. People regain their faith as they regain their sight.
So give people a better vision of the world. Send your tax-deductible contribution to Project ORBIS, Suite 1900, 330 West 42nd St., New York, NY 10036.
Project ORBIS

NASA: No Flight Plan

In mid-December James C. Fletcher, head of the National Aeronautics and Space Administration, was asked by SCIENTIFIC AMERICAN what role the military would play in the space station. Fletcher replied without hesitation: "There is no military role in the space station." Apparently the word failed to get across the Potomac. That same week the Department of Defense asked NASA to delay discussions with other countries interested in the space station until military officials determine how they might use it.

The incident suggests that, in spite of the extensive reexamination that was *Challenger's* legacy, not much has changed at the space agency in the past year. There remains a tenacious commitment to manned space missions, evidenced in the priority assigned to returning the shuttle to service as quickly as possible and to refining plans for the space station. Otherwise there seems to be no well-defined program to guide NASA into the future. That state of affairs, some observers believe, combined with NASA's sensitivity to the needs of the politically powerful Defense Department, promises further deprivation for a space science program that has already been postponed indefinitely by budgetary pressures.

While conceding that they were surprised by the Defense Department request to defer space-platform discussions with other governments, NASA officials say they will honor it. "We're Americans too," NASA executive officer Henry E. Clements said. "The President's program is for the peaceful uses of space. That does not preclude doing DOD research."

Even as Fletcher said that the DOD would not participate in the space station, he emphasized NASA would continue to serve the military. The Defense Department, he said, "will always be a major customer" of the space shuttle. In addition to launching military satellites, Fletcher pointed out, the shuttle will provide a platform for research on the Strategic Defense Initiative, the program to build a shield against nuclear missiles. Recent reports have suggested that the needs of SDI researchers may explain the Defense Department's sudden interest in the space station.

***CHALLENGER* DISASTER a year ago left the U.S. space program in disarray. The photograph, from the report of the presidential investigating commission, shows the undamaged left booster soaring away from the cloud that envelops the disintegrating orbiter.**

This turnabout has disturbed space scientists, many of whom are already skeptical about the station's value to science. "If civilians and the military try to share the space station, there will be conflict," said Thomas M. Donahue, chairman of the National Academy of Sciences Space Science Board. "If there is conflict, the military will always win."

Peter M. Banks of Stanford University, who heads a committee that advises NASA on the needs of the space science community and who has been a vigorous defender of the space station, called the military's newfound interest in the station "very disturbing." If the military becomes involved in the space station, Banks said, Japan will drop out immediately and other countries may soon follow. "The space station will be a success because of international participation," he added, "not because of military participation."

Militarization of the space station might devastate a space science program that is, in Fletcher's words, "in terrible shape." The *Challenger* accident has postponed the scheduled launching of such scientific spacecraft as the Hubble Space Telescope, the *Galileo* Jupiter probe and the *Ulysses* solar probe for five years or more and has left other projects in limbo.

Donahue said there is a "grave danger" that continued disarray in the space science program will deter graduate students from entering astronomy and cause the research community to shrink. Many scientists, notably James A. Van Allen of the University of Iowa, discoverer of the radiation belts that bear his name, argue that NASA should concentrate on development of alternative launchers for relatively cheap, unmanned scientific spacecraft rather than on revamping the shuttle program and establishing the space station (see "Space Science, Space Technology and the Space Station," by James A. Van Allen; SCIENTIFIC AMERICAN, January, 1986).

Fletcher rejects this argument; unmanned scientific missions, he contends, do not generate enough public interest and support to sustain NASA. "Unfortunately," he remarked, "middle America is more interested in the astronauts, the human element." He expressed a desire to help space scientists, however: "They were hurt the most" by delays stemming from the *Challenger* accident, he admitted.

As a step toward rebuilding NASA's ability to launch scientific spacecraft, Fletcher said, NASA intends to establish

M I T S U B I S H I

Soothe your nervous system.

There are those who measure distance in miles. While others choose to measure distance in terms of time.

At Mitsubishi Motors, we take an altogether different view. We believe that the best measure of distance in a luxury sedan is the distance you never notice.

And to prove it, we created the Mitsubishi Galant.

Galant's engineered to deliver a unique driving experience. Patiently crafted to be an island of calm amid the storm of demands success often brings. Galant's onboard computer, ETACS IV™, monitors all secondary functions from power windows and door locks to a new theft-deterrent system. Its electronically fuel-injected 2.4-liter engine responds with a silent surge of power. There's even an available ECS™ electronically controlled suspension that self-adjusts to changing road conditions.

Of course a sedan this sophisticated demands luxury as well as performance, so inside you'll find a haven of comforts and appointments. Adjustable front bucket seats, in velour or optional leather. Electronic power steering for smooth, confident handling. An available six-speaker AM/FM cassette stereo system with graphic equalizer and remote controls mounted on the steering wheel. And to ensure that passengers in back are every bit as comfortable, Galant offers a unique reclining rear seat with adjustable headrests.

The Mitsubishi Galant. Road car performance in a luxury sedan. Patiently engineered to soothe your nervous system and to keep an impatient world at a comfortable distance.

Call 1-800-447-4700 for your nearest dealer.

Mitsubishi. Perfecting the Experience.™

MITSUBISHI MOTORS

G A L A N T

a fleet of unmanned rockets, called expendable launch vehicles (ELV's), to provide an alternative to the shuttle. NASA's decision to eliminate its ELV programs in the late 1970's in order to concentrate all its resources on the shuttle, Fletcher acknowledged, "was probably a mistake."

Yet NASA may not gain the funding necessary to have its own stock of expendable launchers, according to John L. McLucas, a former secretary of the Air Force and adviser to NASA and now head of an aerospace consulting firm. McLucas speculates that NASA will have to use rockets built for the Air Force, which has contracted for its own ELV's.

Representative George E. Brown, Jr., of the House Science and Technology Committee also questions whether NASA has the political clout to gain a new ELV program. What is lacking, Brown says, is "a clean indication of support from the Administration." He attributes to the Executive Branch much of the blame for NASA's recent difficulties in setting its house in order. The White House offices of Science and Technology and of Management and Budget, Brown maintains, have been slow and ineffective in presenting NASA's requests for another shuttle, for expendable launch vehicles and for the space station; meanwhile the military has moved "smartly" to fill the void.

In addition to the formidable distraction of the military presence in NASA's plans, space scientists must also contend with the agency's commercial preoccupation. Even though President Reagan has ordered NASA to stop launching commercial satellites and although only $27 million of NASA's 1987 budget is devoted to commercial research and development, NASA still promotes commercial exploitation of space. Materials processing, Fletcher said, "is still a very big part of not only the shuttle but also the space station. Some of this is science, but a lot has commercial potential."

There are important dissenters from this view. Robert A. Frosch, a former NASA administrator and now vice-president for research at General Motors, says: "I think it's unrealistic to expect commercial benefits from something that hasn't demonstrated it can lead to a product. Even if it turns out that space processing is valuable, I'm not sure that will be enough [to justify the space station]."

Former NASA administrator Thomas O. Paine, who chaired a commission that recommended "future directions" for the U.S. space program last spring, argues that the space station should be seen as a step toward manned exploration and settlement of the solar system. Paine's ambitious views have been received coolly by the Administration. He insists, however, that without such a bold program, driven by higher funding, NASA is doomed to continue its "downward spiral." A less grandiose vision might do, Van Allen and others maintain: one that also nurtures the needs of space science. The price of *Challenger's* replacement would provide adequate seed money.

Animal Welfare

After several years of activism the animal welfare issue has begun to make some political fur fly. An animal-experimentation ordinance proposed in Cambridge, Mass., has prompted more than 1,000 letters to the city clerk's office since its introduction last September. The proposal, which is awaiting final judgment from the city council, would establish an advisory committee to regulate all animal experimentation within city limits, including more than 50,000 animals in laboratories at Harvard University, the Massachusetts Institute of Technology and local biotechnology companies. Many people doubt that the ordinance will become law, but it has raised the hackles of academic and corporate investigators alike.

The proposal was drawn up by the Cambridge Committee for Responsible Research, a fledgling organization formed under Gul A. Agha of M.I.T., a specialist in artificial intelligence. If it is passed, it would require the city government to review all experimental protocols and proposals involving experimental animals and allow it to halt any experiment involving "pain with inadequate anesthesia."

Proponents of the ordinance, including the Massachusetts Society for the Prevention of Cruelty to Animals, say it would fill gaps in the monitoring practices dictated by state and Federal law. Rats and mice, for instance, which constitute more than 90 percent of the laboratory animals in the area, do not fit the U.S. Department of Agriculture's definition of "animals." And even the proposal's opponents concede that most inspections are concerned primarily with animal facilities rather than with the procedures to which the animals are subjected.

Critics nonetheless argue that the proposal is both ambiguous and impractical. Its definition of "animals" could extend to fruit flies and sea slugs, one administrator at Harvard complained. Delays in protocol review could jeopardize grant funding at universities, and delays seem inevitable: Harvard investigators alone would generate about 2,000 such protocols. Making experiments public could also infringe on proprietary procedures in company laboratories.

Critics contend too that the Cambridge committee's proposal is redundant. In 1985 the Animal Welfare Act was strengthened and the Department of Health and Human Services required all publicly supported institutions to establish animal-research "watchdog" committees. A year ago the Office of Technology Assessment produced a report emphasizing the need to exploit alternatives to animal research.

The objections typify the kind of resistance the animal welfare movement has recently encountered. Many scientists believe major abuses have been identified and corrected. The Cambridge committee's proposal in particular may lack some impetus because it was prompted not by any instance of abuse but by a wish to bring animal experimentation into the public domain. The issue has received so much local publicity that this objective may already have been achieved.

PHYSICAL SCIENCES

Core Questions

What little is known of the earth's core—it has a solid inner layer and a liquid outer one, both made predominantly of iron—has been gleaned indirectly, from the analysis of seismic waves that bounce off or pass through it. By borrowing a technique from medical diagnosis, computer-aided tomography (CAT scanning), geologists have found a richer way to read such data, one that has begun to provide a more sophisticated view of the core. The new knowledge is likely to have ripple effects in other areas of geophysics as well.

Early results from the tomographic studies include the first topographic maps of the boundary between the core and the overlying mantle. The maps, made by research groups at the California Institute of Technology, Harvard University and the Massachusetts Institute of Technology, suggest that the core is not a smooth ellipsoid. Instead its surface is marked by broad swells that are probably between five and 10 kilometers in height, and by broad and equally deep depressions.

Tomographers obtain these and other results by analyzing the time it takes seismic waves to travel from an earthquake on one side of the planet to a seismometer on the other side. The travel time, say, of a compressional wave that bounces off the core (a *PcP* wave) will be shorter than expected if the wave happens to hit a swell on the core surface. On the other hand, a wave that penetrates the outer core (a *PkP* wave) will have a travel time that is longer than normal if it hits a swell, because it will have a longer path through the core, which transmits seismic waves less rapidly than the solid mantle does. Features that speed up or slow down seismic waves are found by analyzing thousands of crisscrossing waves.

Interpreting the results is not a trivial matter. A *PkP* wave, for instance, may be slowed by a bump on the core; it may also be slowed by its passage through a region of the mantle that is hotter (and therefore less rigid) than normal, or through a region of the mantle or core that has a chemical composition slightly different from that of its surroundings. The difficulty of untangling such effects, along with the fact that investigators have worked with different data sets, helps to explain why the various tomographic maps agree that the core has topographic features but do not agree on precisely where the features are.

The presence of any kind of topography on the core-mantle boundary bears on a number of fundamental geophysical questions, according to Adam M. Dziewonski of Harvard. A rough boundary would presumably perturb the flow of liquid iron in the outer core, much as a mountain influences winds. The flow in the outer core is thought to generate the earth's magnetic field; hence to understand the geodynamo and its many fluctuations (such as the wandering of magnetic north) it will probably be necessary to have a precise topographic map of the core.

The topography of the core also has implications for the flow of energy in the earth. A rough boundary would suggest that the geodynamo uses a lot of energy, but also that a lot of energy escapes the core as heat. It would further imply that the slow convective circulation of rock in the mantle, which is believed to drive the motions of the tectonic plates at the earth's surface, is itself driven largely by heat emanating from the core.

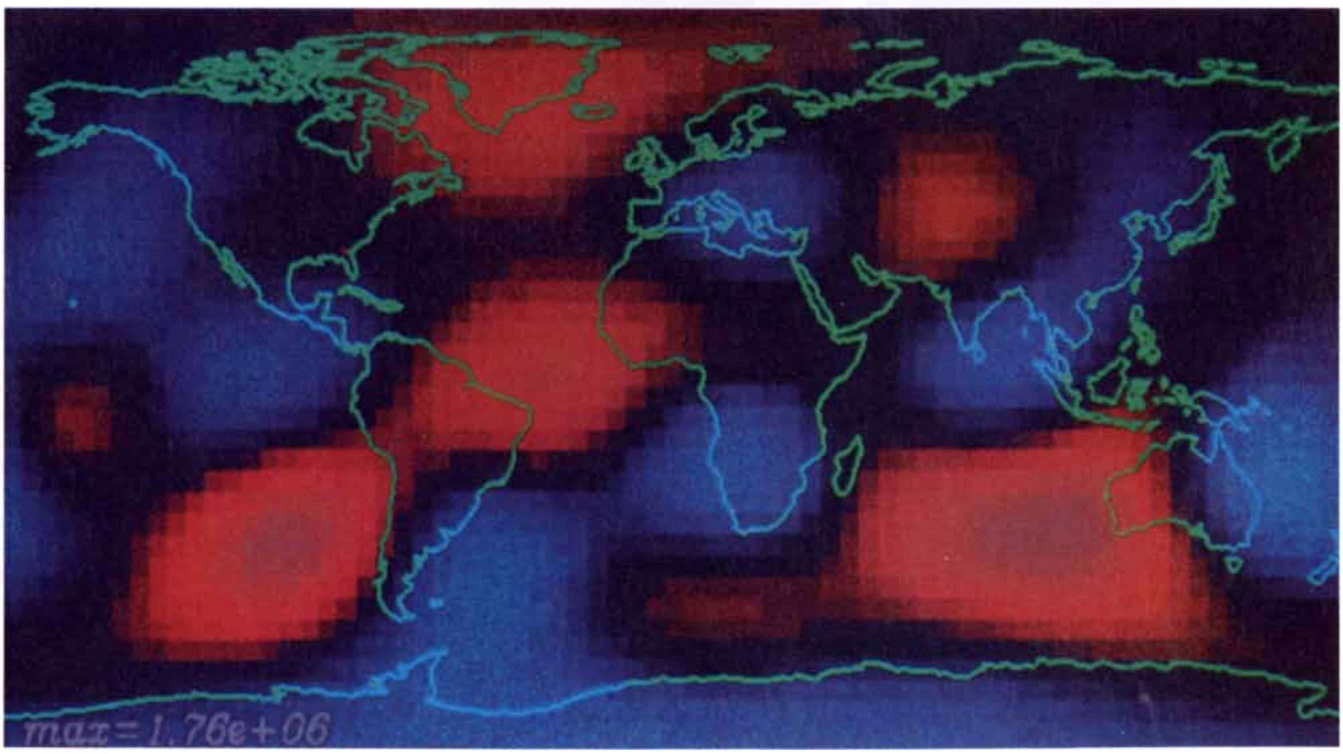
Unmashing Chaos

Physicists seeking order within chaotic systems often rely on fractals: geometric forms with irregular patterns that repeat themselves at different scales. Many complex phenomena, such as crystal formation and changes in animal populations, can be represented with fractals (see "Fractal Growth," by Leonard M. Sander; SCIENTIFIC AMERICAN, January). But to the extent that fractals represent a range of possible events without regard for the order in which they unfold, they say nothing of the actual history of a chaotic system. Now Mitchell J. Feigenbaum of Rockefeller University and Mogens H. Jensen and Itamar Procaccia of the University of Chicago have found a way to probe fractals for insights into the events that lead to chaos.

The group tested the technique on fractals derived from fluid-turbulence experiments carried out at Chicago. Mercury in a chamber was simultaneously heated and electrically perturbed. As the mercury became chaotic, readings from a thermometer in the chamber fluctuated; plots of the frequencies of these fluctuations revealed a pattern of fractal repetition.

By analyzing these fractal "measures" the investigators deduced how the velocity of a typical section of mercury oscillated prior to the onset of turbulence. They also delineated the respective roles that heat and electric current played in stirring the mercury. Extracting this information from fractals, the workers comment in *Physical Review Letters,* "is like retrieving whole potatoes from mashed potatoes."

The technique involves an analysis of what is called the fractional dimension of a fractal. The concept can be understood, Feigenbaum says, by considering a fractal curve. The reiteration of irregular details at progressively smaller scales, he explains, makes the curve "fatter than a straight line" but still "not as fat as a solid blob." The fractional dimension of the curve,



TOPOGRAPHIC MAP of the earth's core was derived by means of seismic tomography by Robert W. Clayton and his colleagues at Caltech. Red areas are swells on the core; blue areas are depressions. The variation from mean altitude is about 10 kilometers.

then, lies between 1 and 2; the dimension approaches 2 as the density of the details and the frequency with which those details are repeated at different scales increase.

As a whole a fractal will have a single fractional dimension. The dimension will also vary across a narrow spectrum depending on the scale at which the fractal is examined. It is from this spectrum of dimensions that Feigenbaum and his colleagues extract information about the causes of chaos.

The technique "still has some serious shortcomings," Feigenbaum acknowledges: the data processing is complicated and "there is genuine compression of information" in constructing the spectrum of a fractal. But this very compression, he adds, makes the technique an ideal "bridge" between the two mathematical methods—one extremely precise and the other statistical—usually employed to predict chaotic behavior.

Statistical theory has been hard pressed to describe chaos in any but the most general terms. Extremely precise descriptions, on the other hand, work well for nearly chaotic motion but not for fully chaotic motion. They are also difficult to verify experimentally, Feigenbaum says, because "real experiments inadvertently have some noise in them." The more compressed descriptions obtained from spectrum analysis are less sensitive to this noise.

Could the technique help to "unmash" systems somewhat larger and more complex than a chamber filled with mercury? Could it help astrophysicists, for example, to deduce the initial conditions of the universe from fractal representations of its present state? Not yet, says Feigenbaum, but he adds that "some people are thinking about" that possibility.

Halley's Legacy

Nearly one year after Halley's comet hurtled past the earth, investigators are still pondering data gathered by the six spacecraft and countless ground observatories that tracked its journey. The investigators have reached a consensus on several key issues, but many questions remain.

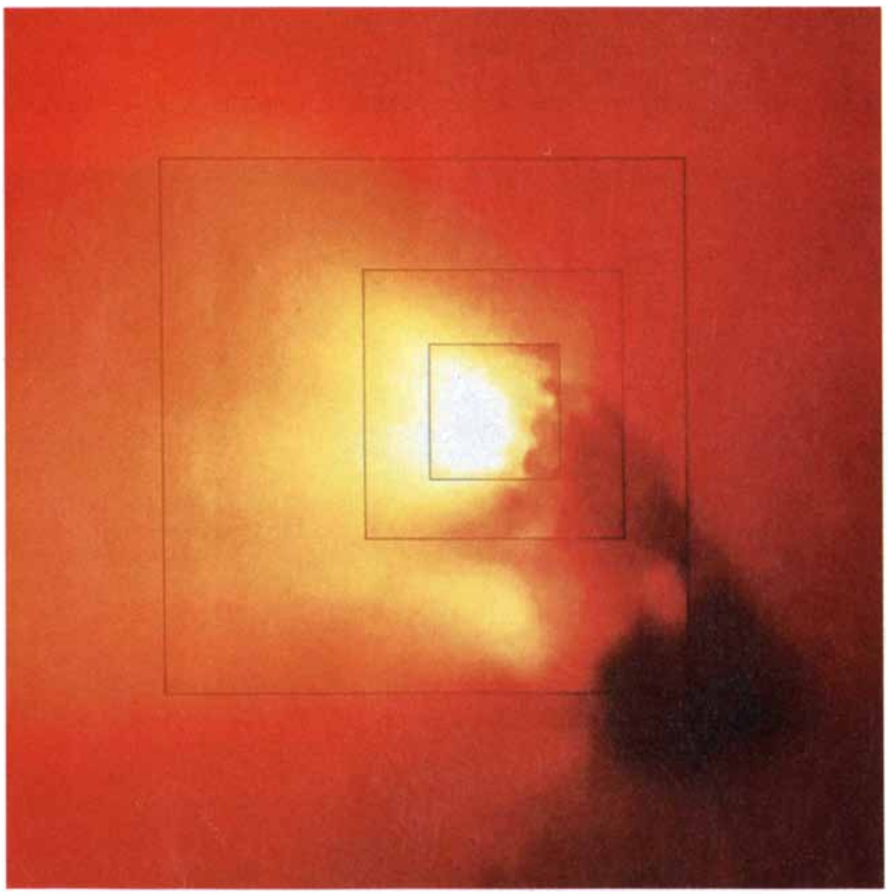
From the spiraling of dust jetting from the comet's nucleus, workers have deduced that the nucleus rotates; uncertainty lingers over the rate of rotation. Last summer it was generally believed the nucleus rotates once every 2.2 days, but new evidence favors a rotation period of 7.4 days. The answer to the apparent contradiction may be that the nucleus rotates about its long axis once every 7.4 days and that this long axis precesses, or wobbles, once every 2.2 days.

There is now agreement that the nucleus, which some observers have described as a 10-mile-long peanut, is encrusted by a layer of dust. Estimates of the thickness of the layer range from less than one centimeter to several tens of meters. The dust's thickness may vary from place to place and may change as time passes; this variation might explain why almost all the gas and dust forming the comet's tail issues in powerful jets from a few points on the nucleus.

Some investigators have proposed that the dust layer is a porous "lattice," formed when water evaporates from a mixture of dust and ice on the surface of the nucleus. Porosity could account for a major mystery: why the relatively low-density nucleus appears to be so dark. The lattice of dust particles may be porous enough to admit solar photons and yet fine-grained enough to trap many of these photons inside the nucleus. Investigators have also proposed that the nucleus contains simple organic molecules—made up of hydrogen, carbon, nitrogen and oxygen—that form dark, tarlike substances.

The latter theory may be impossible to confirm. By analyzing clouds of gas and dust expelled from the nucleus, investigators have deduced that from 80 to 90 percent of the mass of the comet's nucleus consists of dust and frozen water. Because the comet's tail only imperfectly represents the composition of the nucleus, however, the investigators cannot be sure about the remaining 10 to 20 percent. Whereas dust expelled from the nucleus remains essentially unchanged, gases undergo a complex series of chemical reactions as they travel outward that obscure their molecular ancestry.

This report is condensed from an account by Rüdeger Reinhard of the European Space Agency of findings presented at an international symposium on Halley's comet held in Heidelberg. Reinhard's full account is scheduled to appear soon in *Spektrum der Wissenschaft*, the German-language edition of SCIENTIFIC AMERICAN.



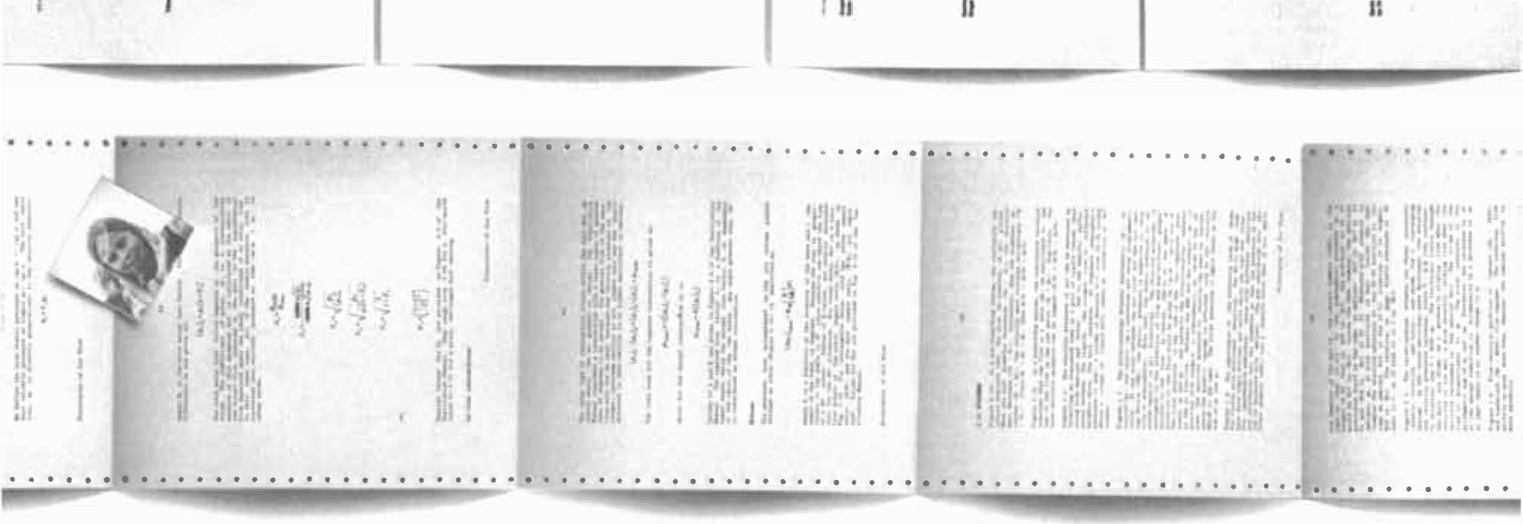
NUCLEUS of Halley's comet was photographed by the European Space Agency's probe *Giotto*. Photographs made at distances ranging from 20,000 to 2,500 kilometers were combined to form a single image. The innermost frame corresponds to four kilometers.

Probing the Permafrost

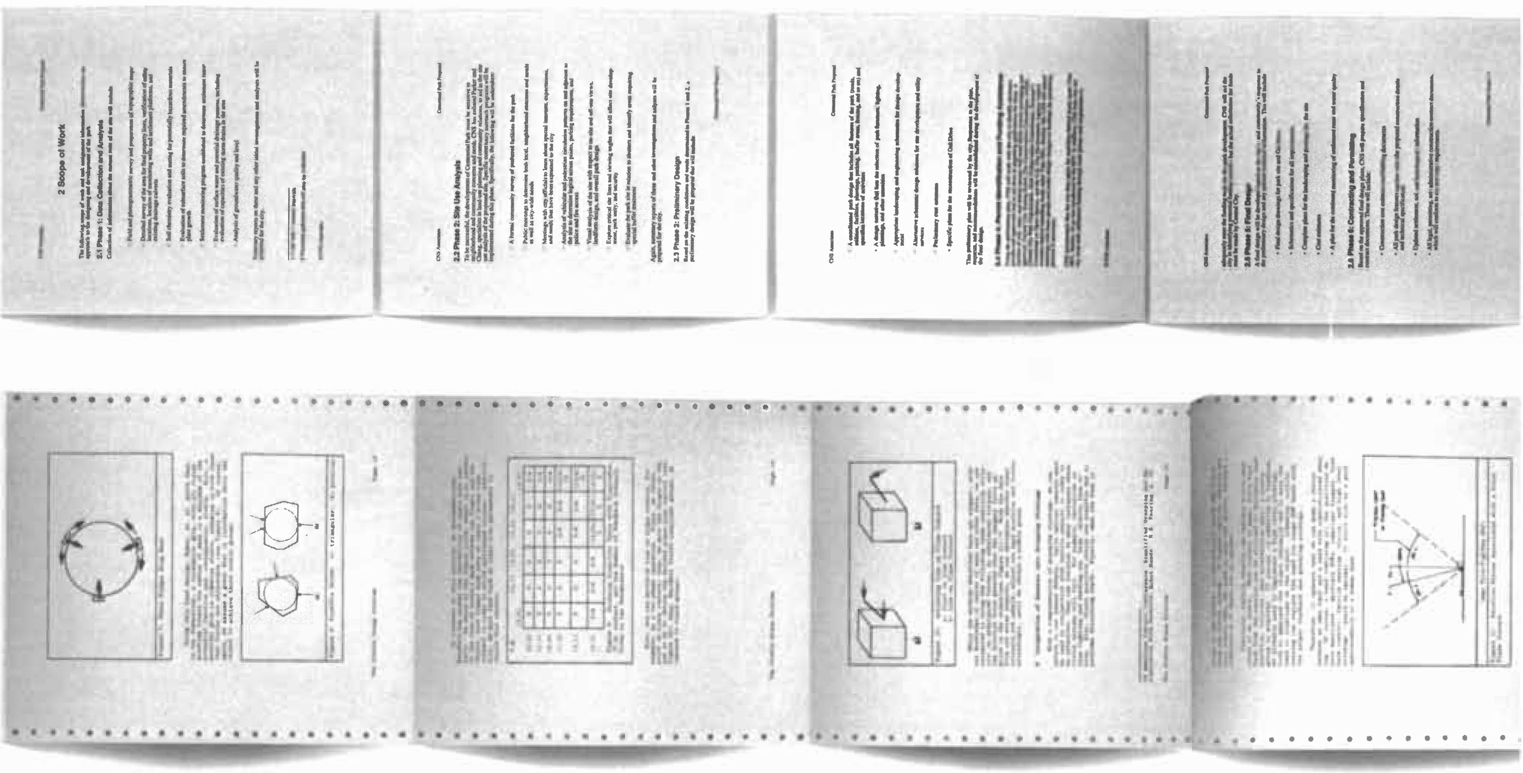
Confirmation of the greenhouse-effect theory, which holds that a buildup of carbon dioxide and other industrial by-products is warming the earth's atmosphere, has been thwarted

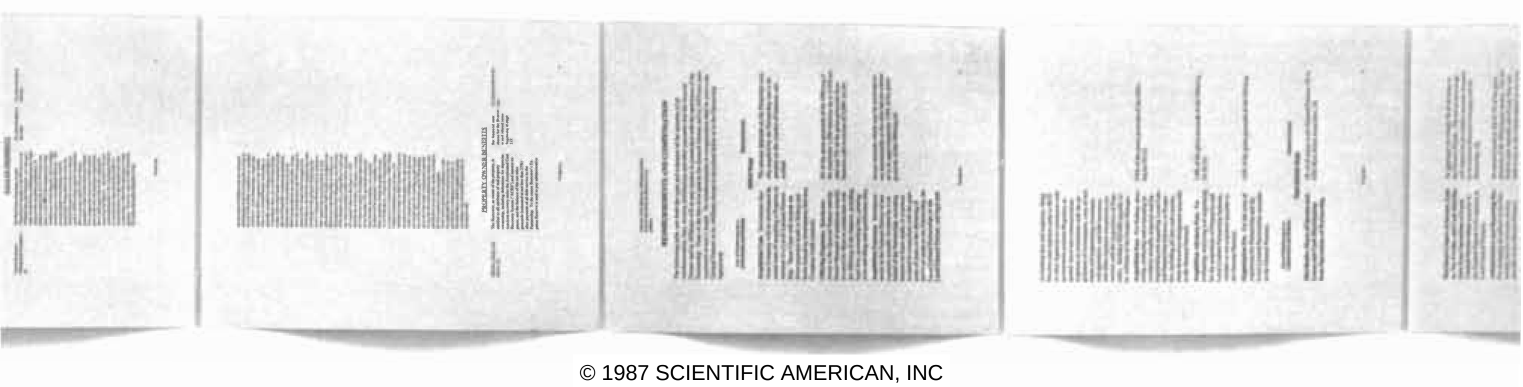
Your word processing problems

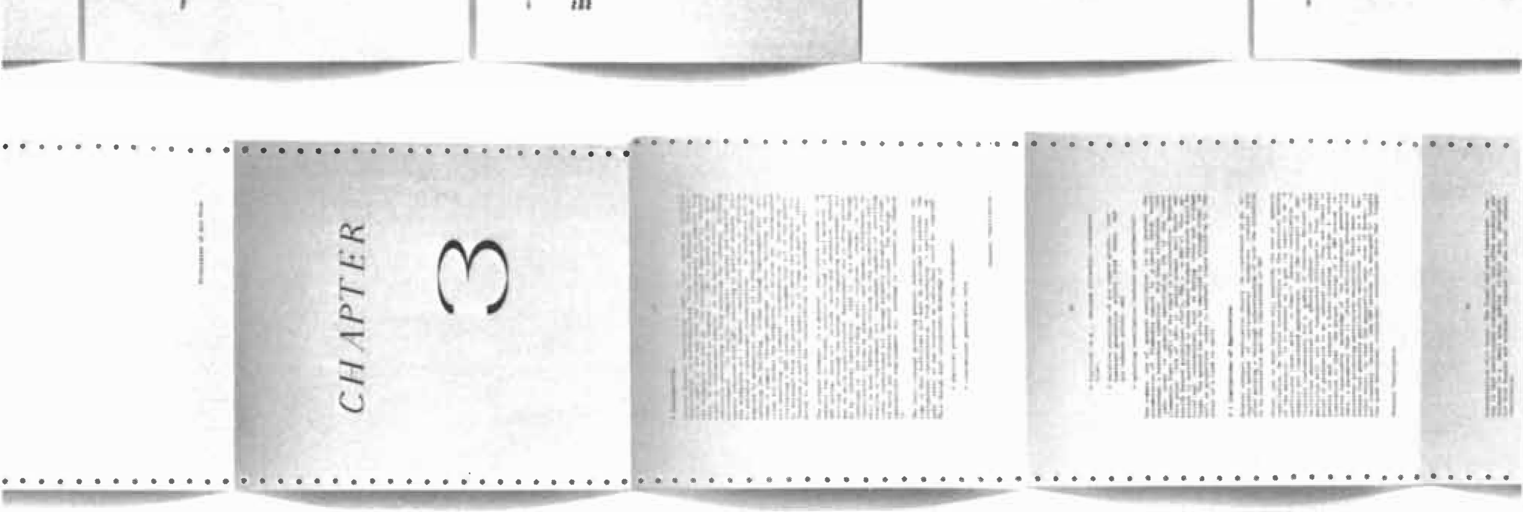
will be a thing of the past if you



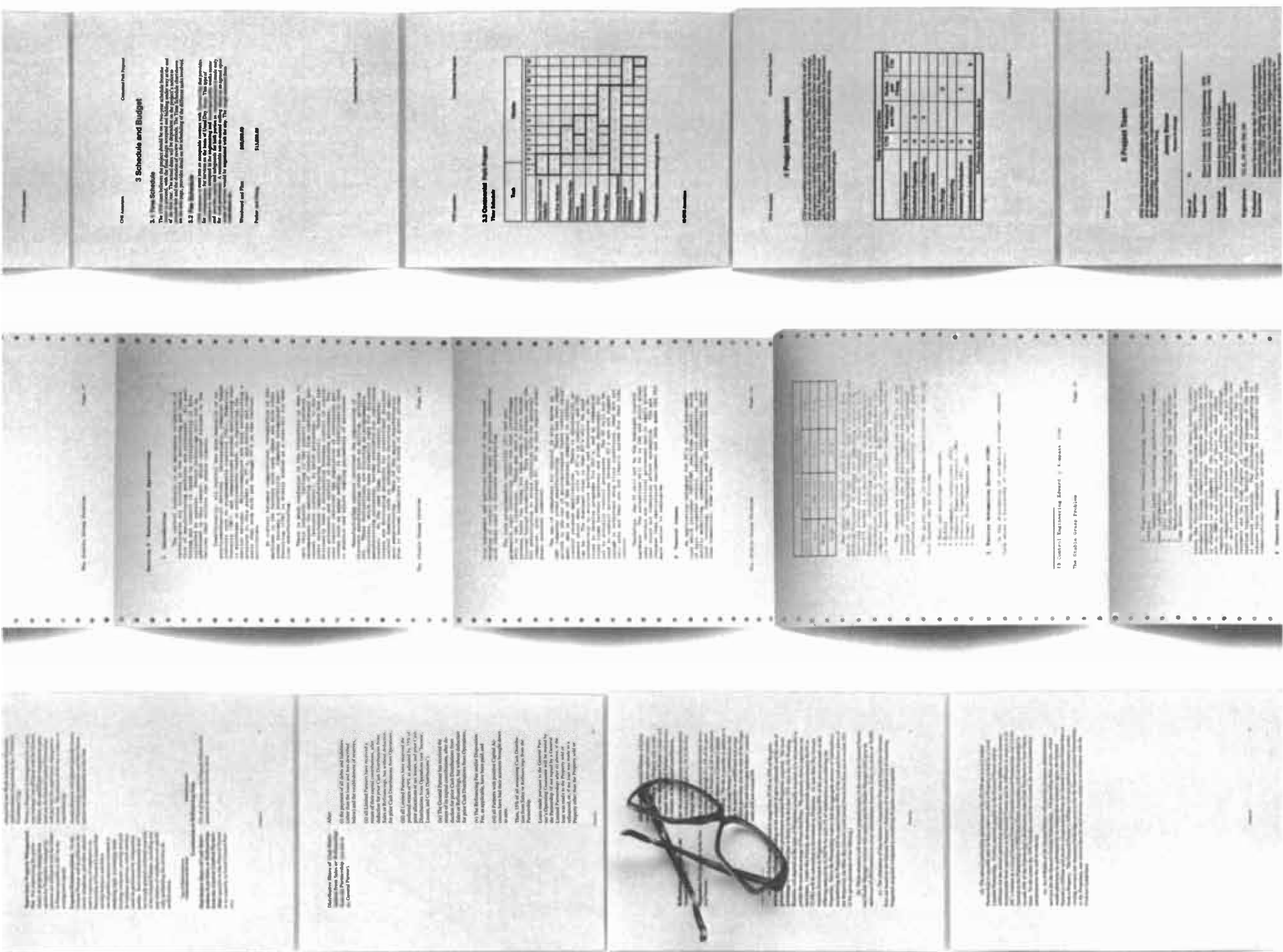
Was your word processor

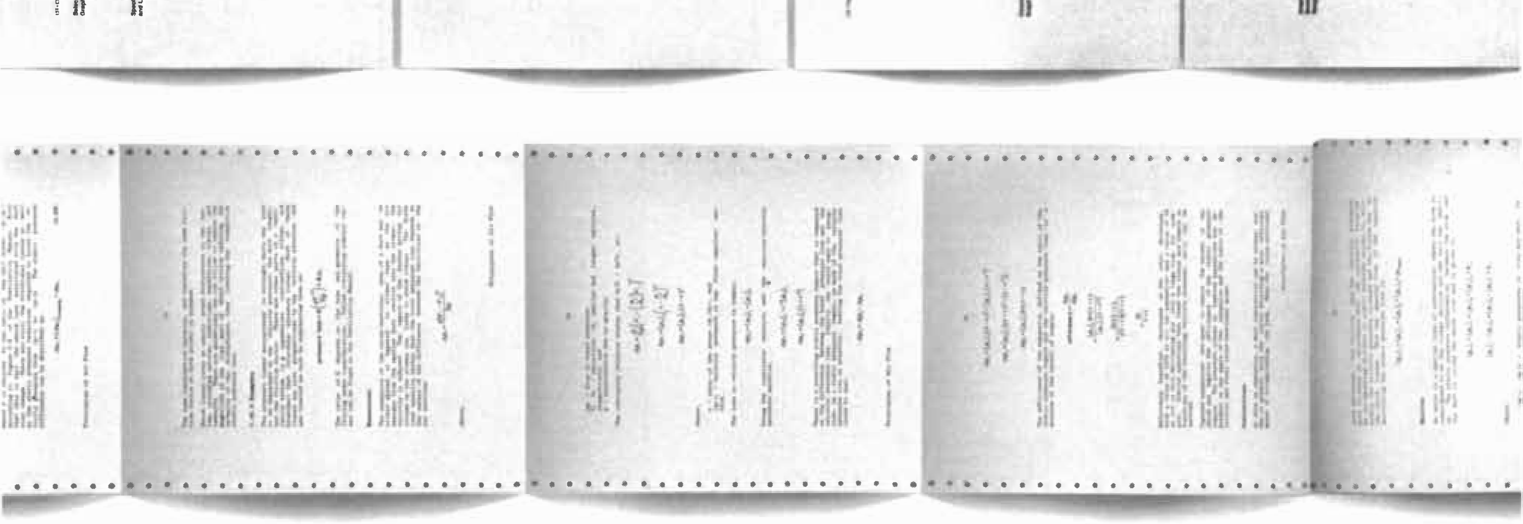




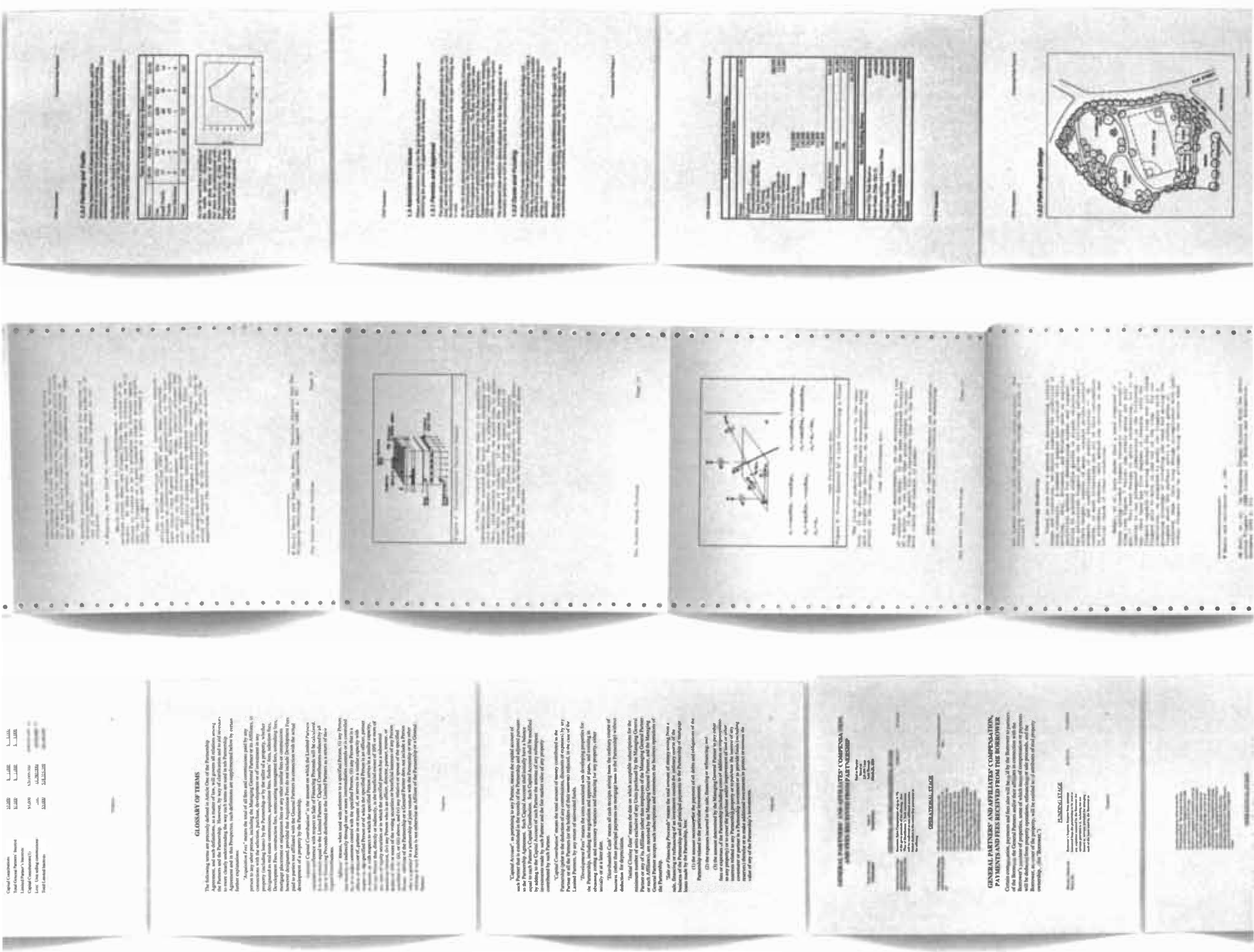


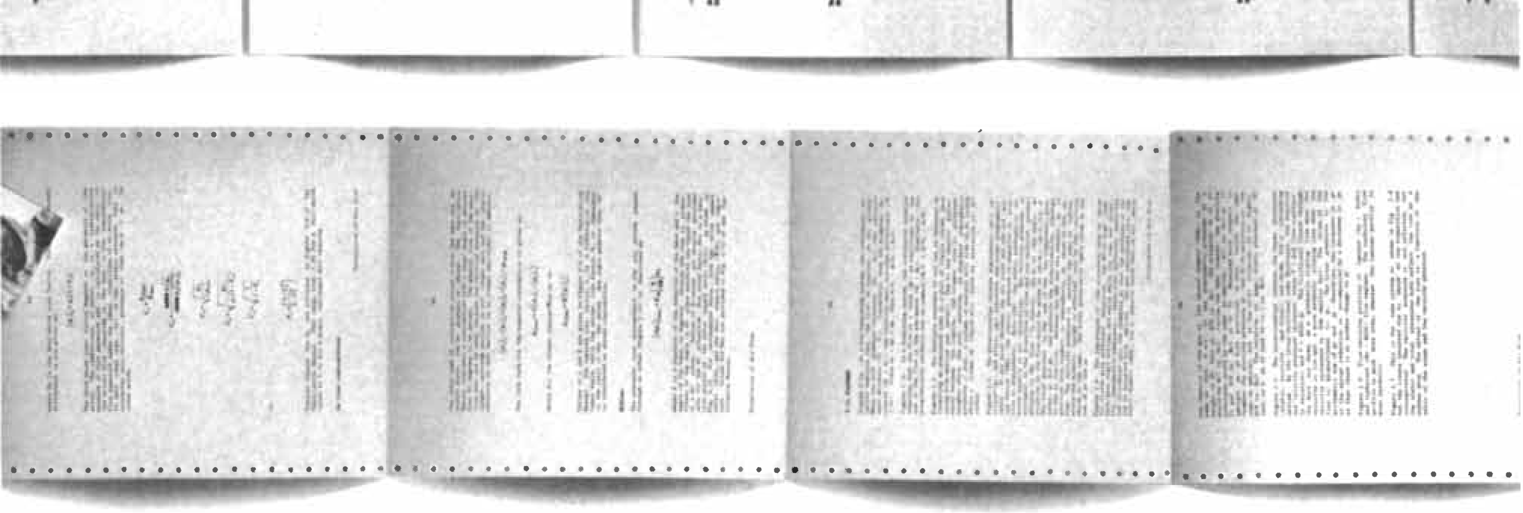
designed to write like this?



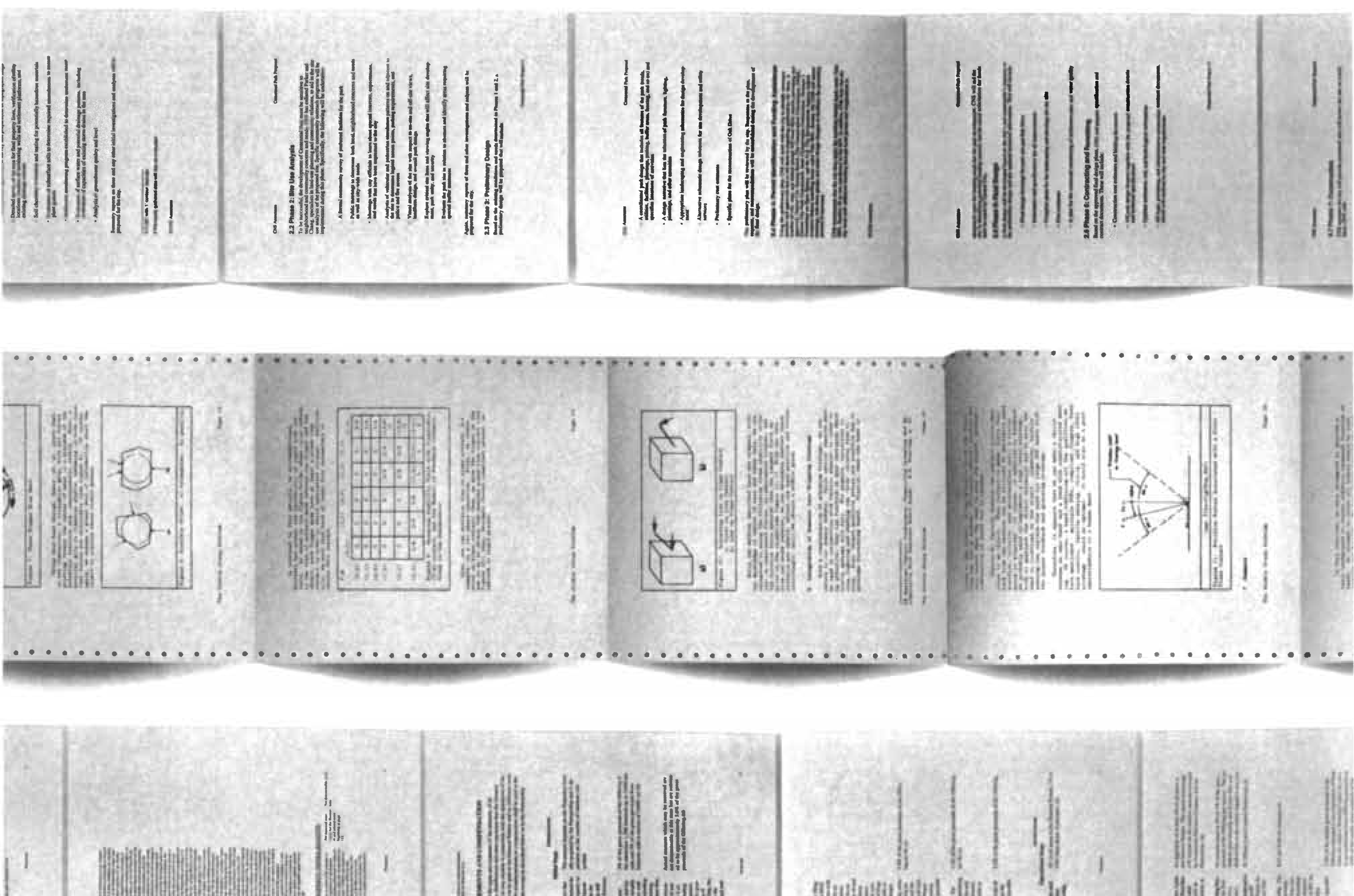


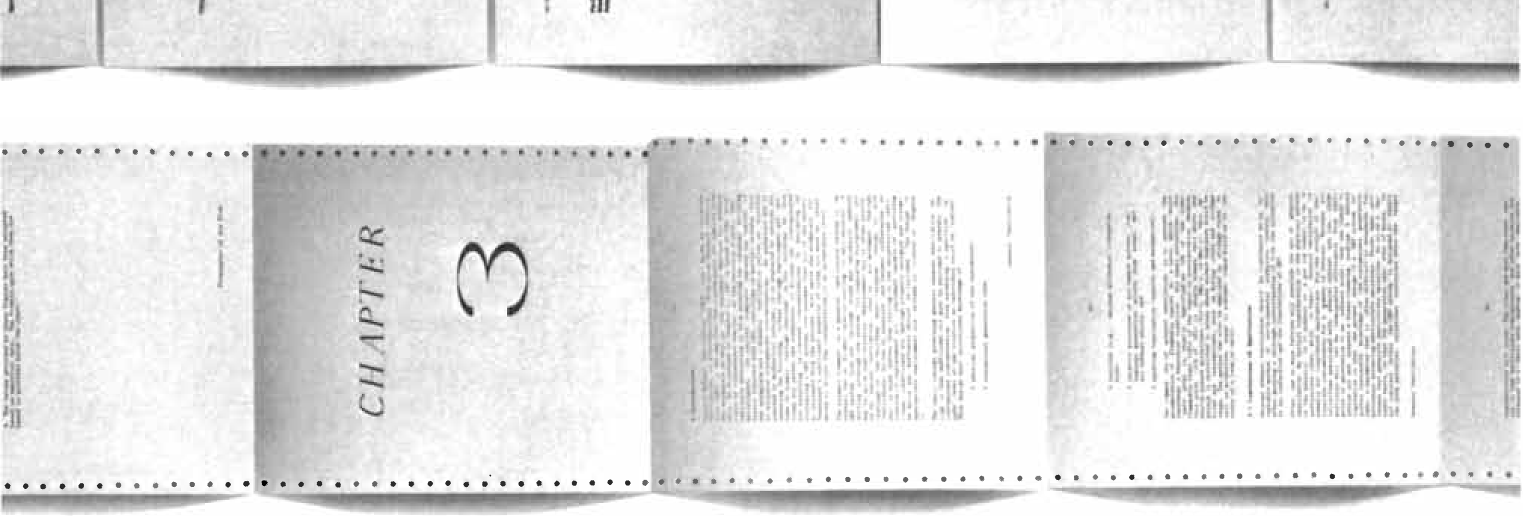
Let's say you have to do



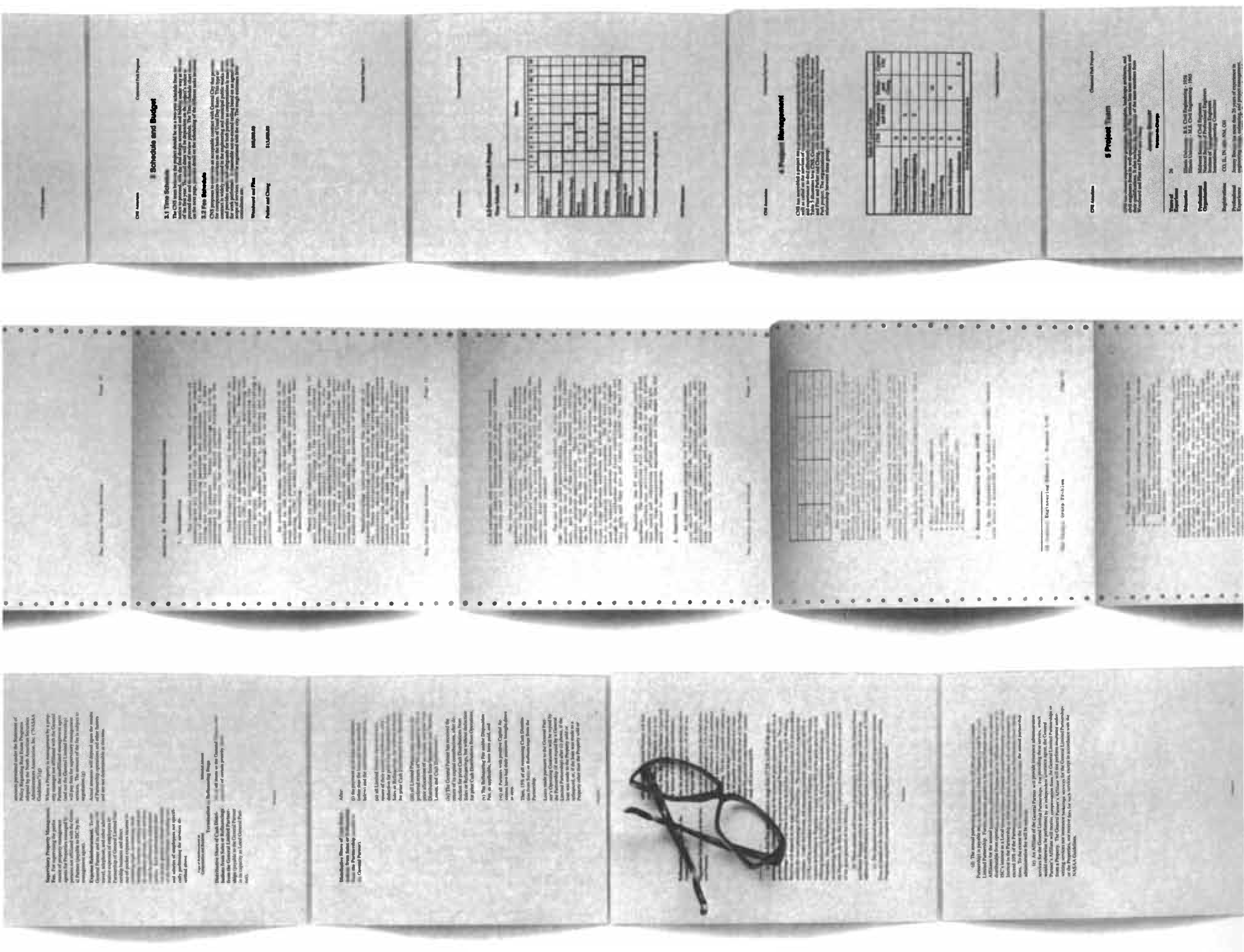


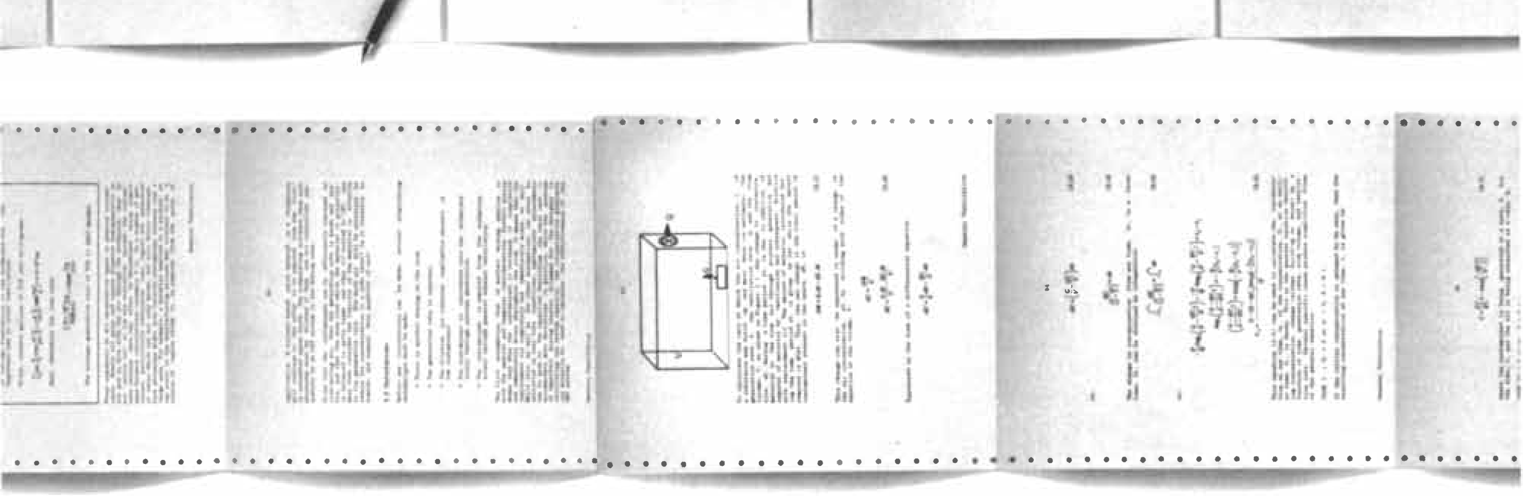
a very long, very complex,



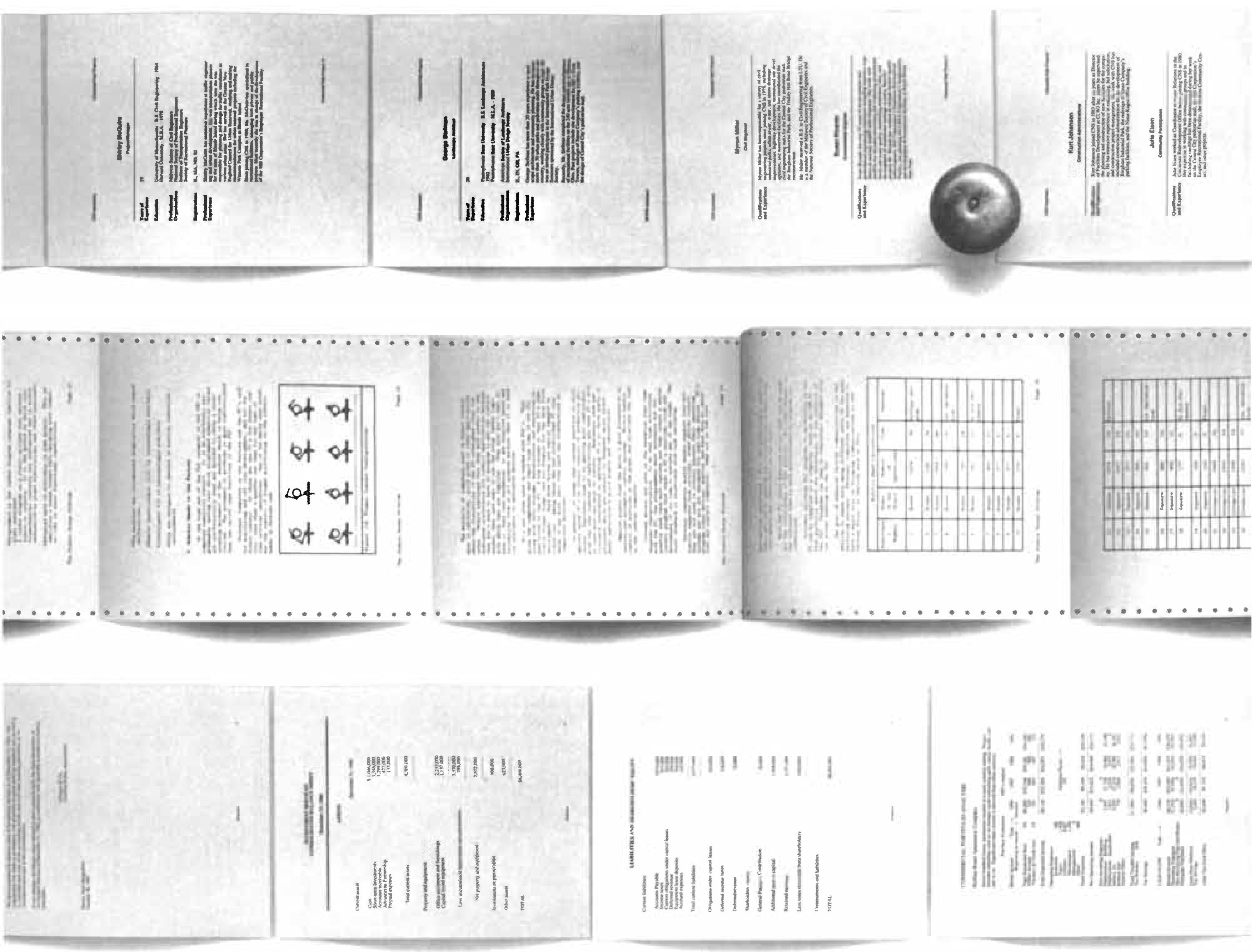


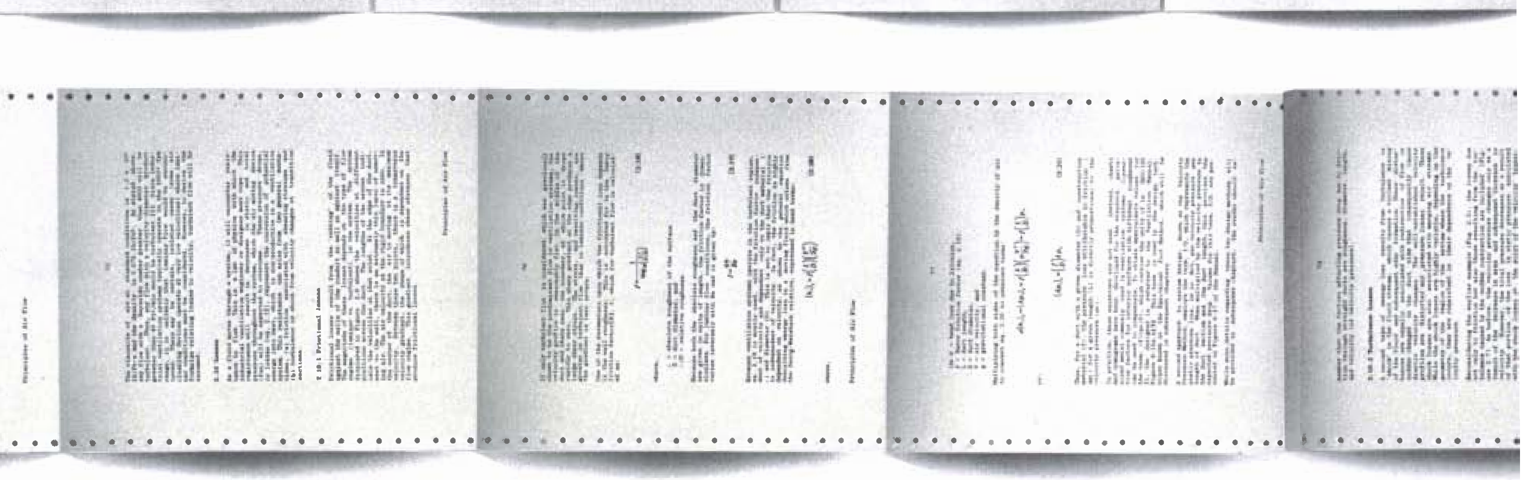
very technical document,



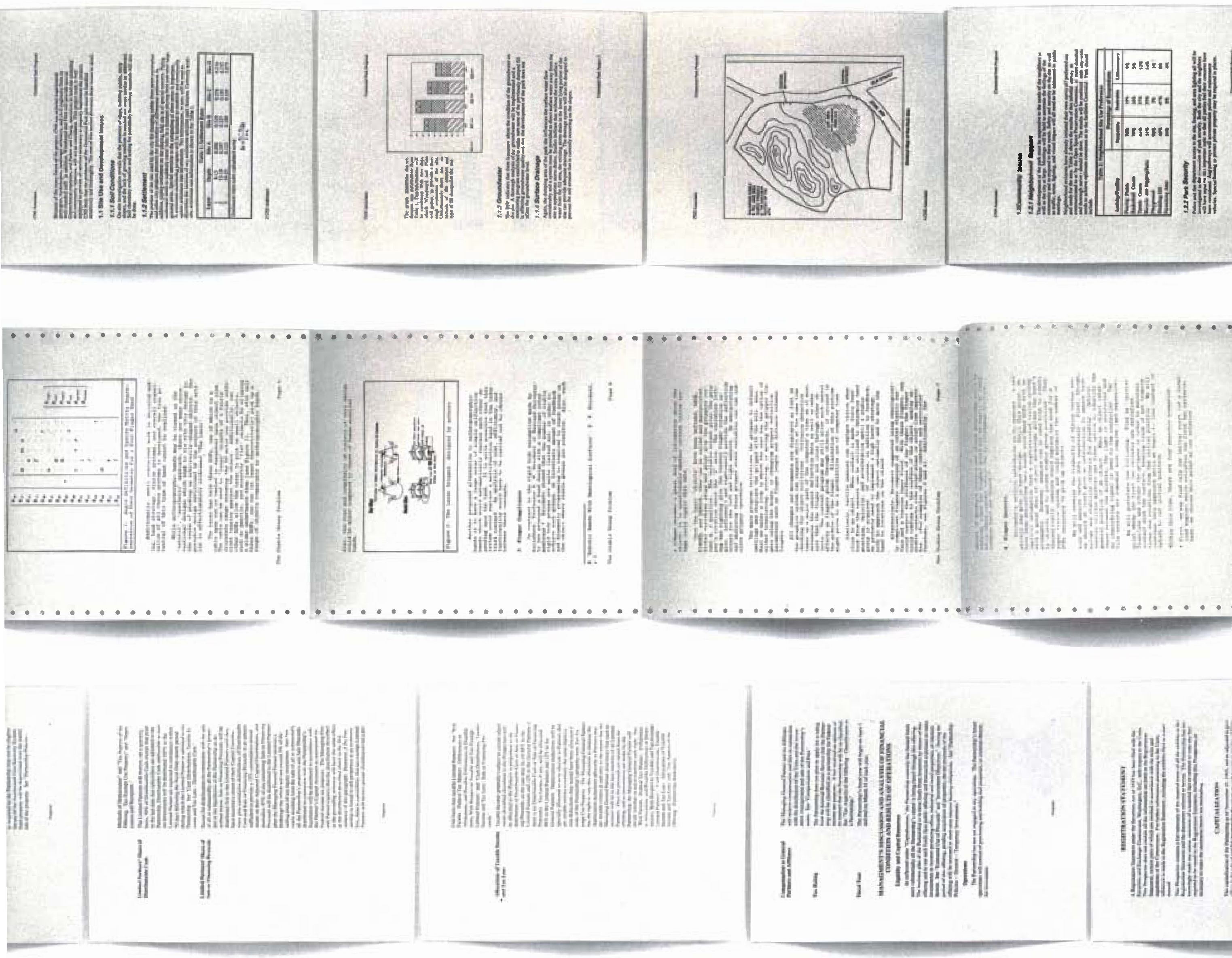


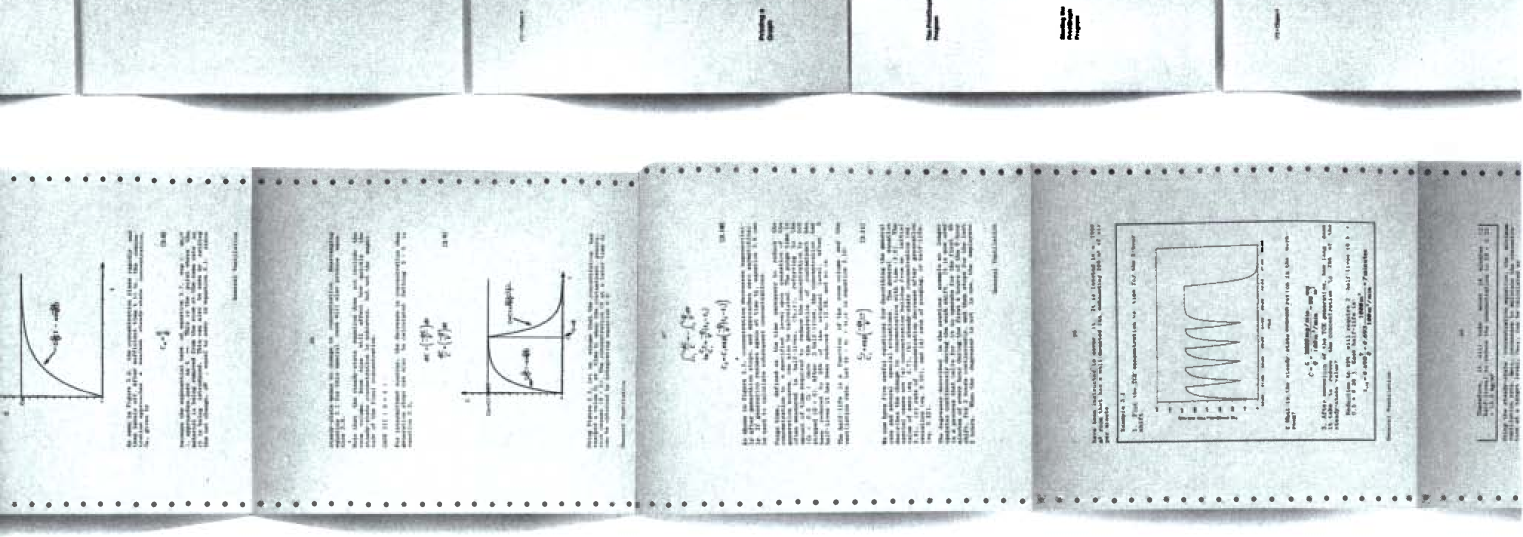
in a short amount of time.



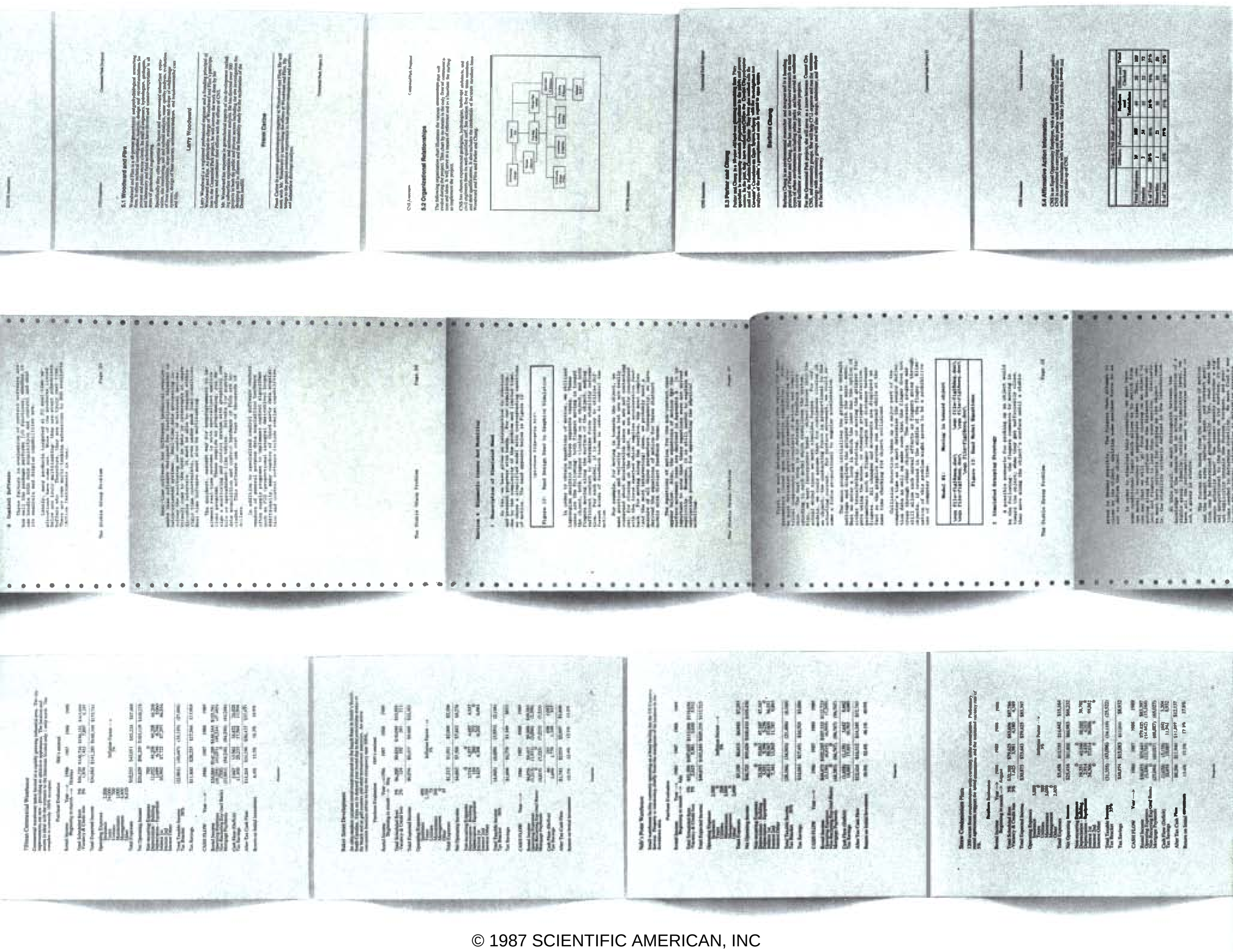


But wait, it's even worse,

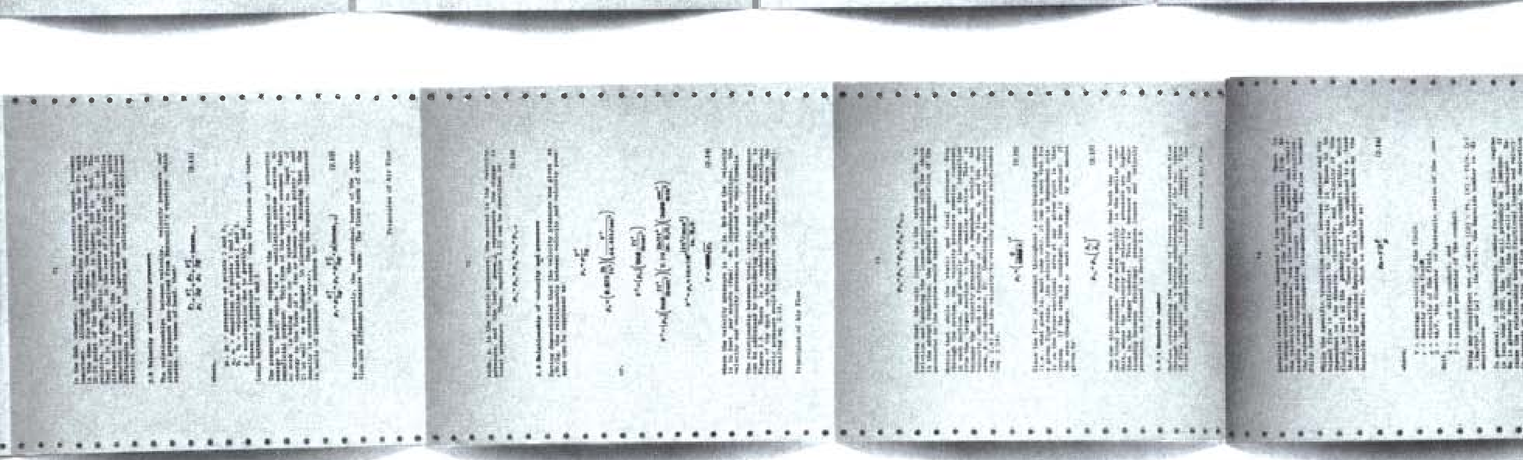


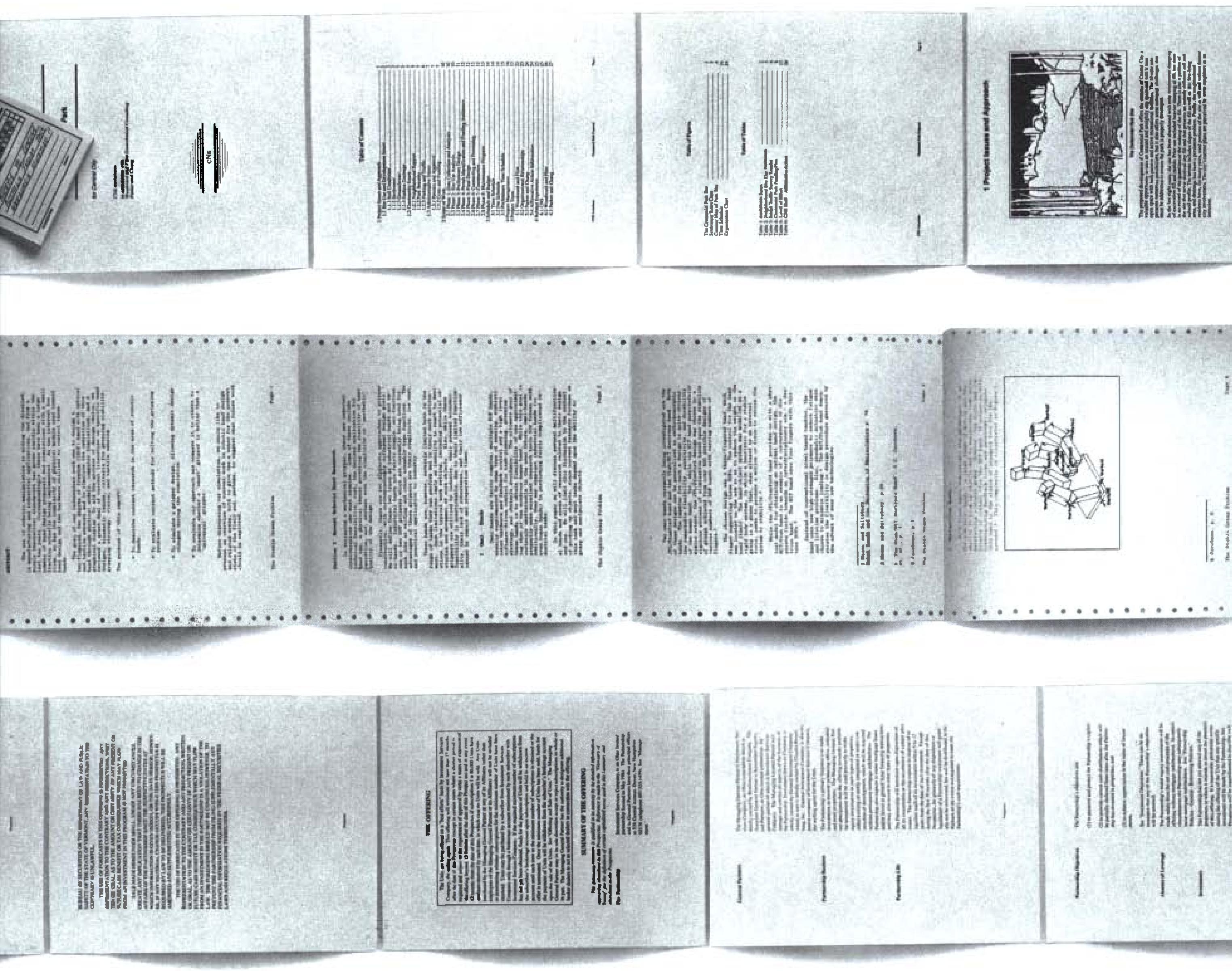


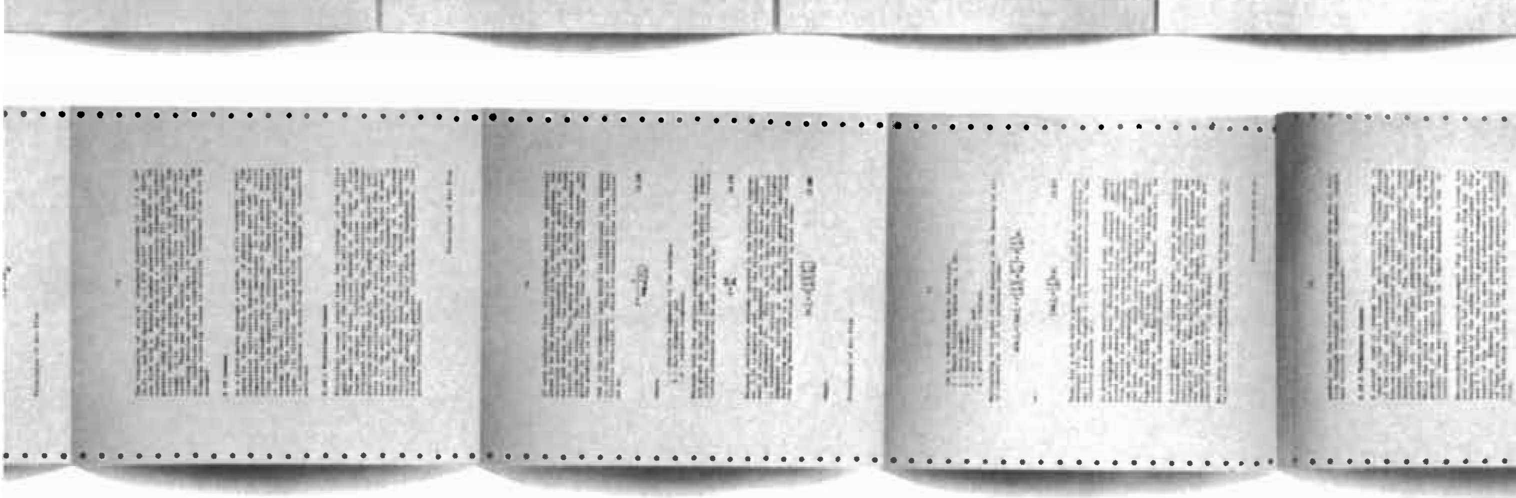
with plenty of revisions.



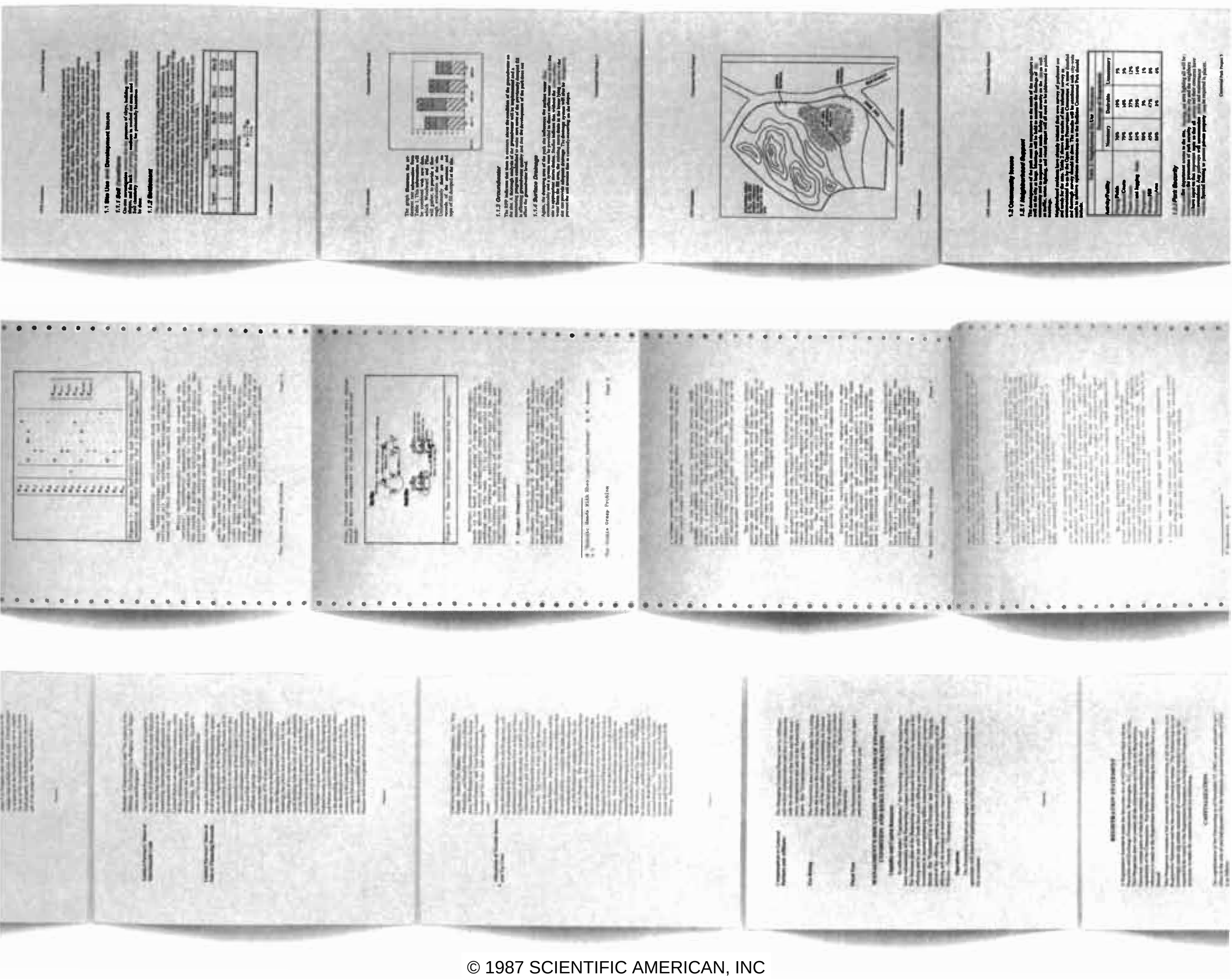

Chapter 4: Graphs

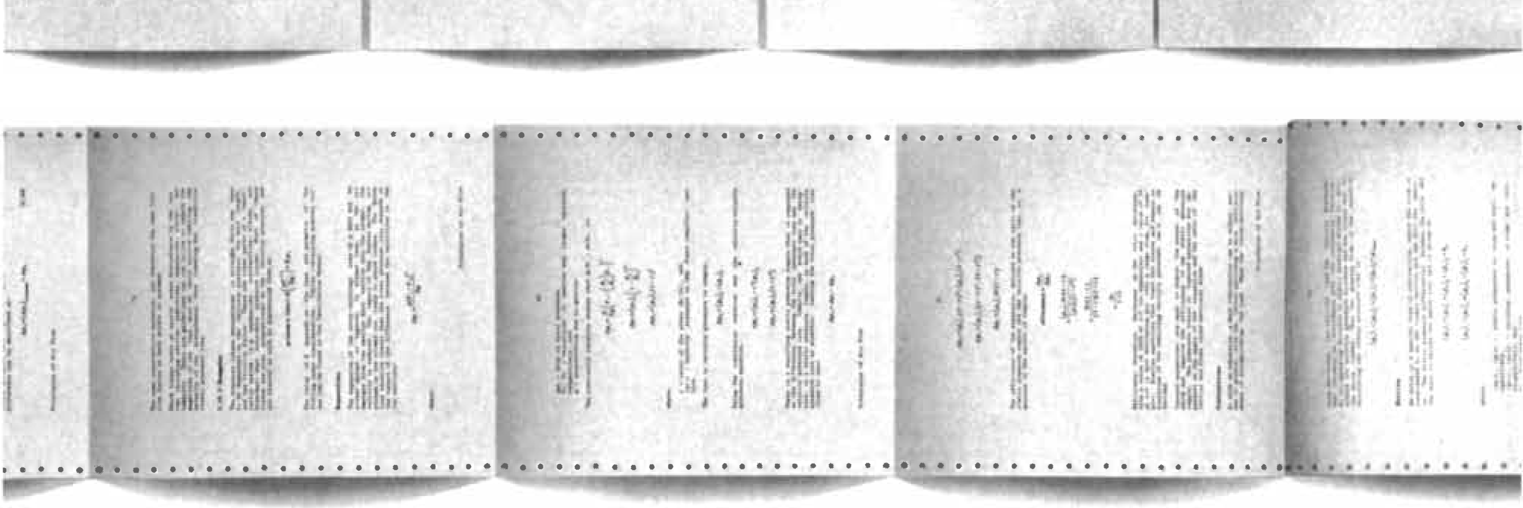




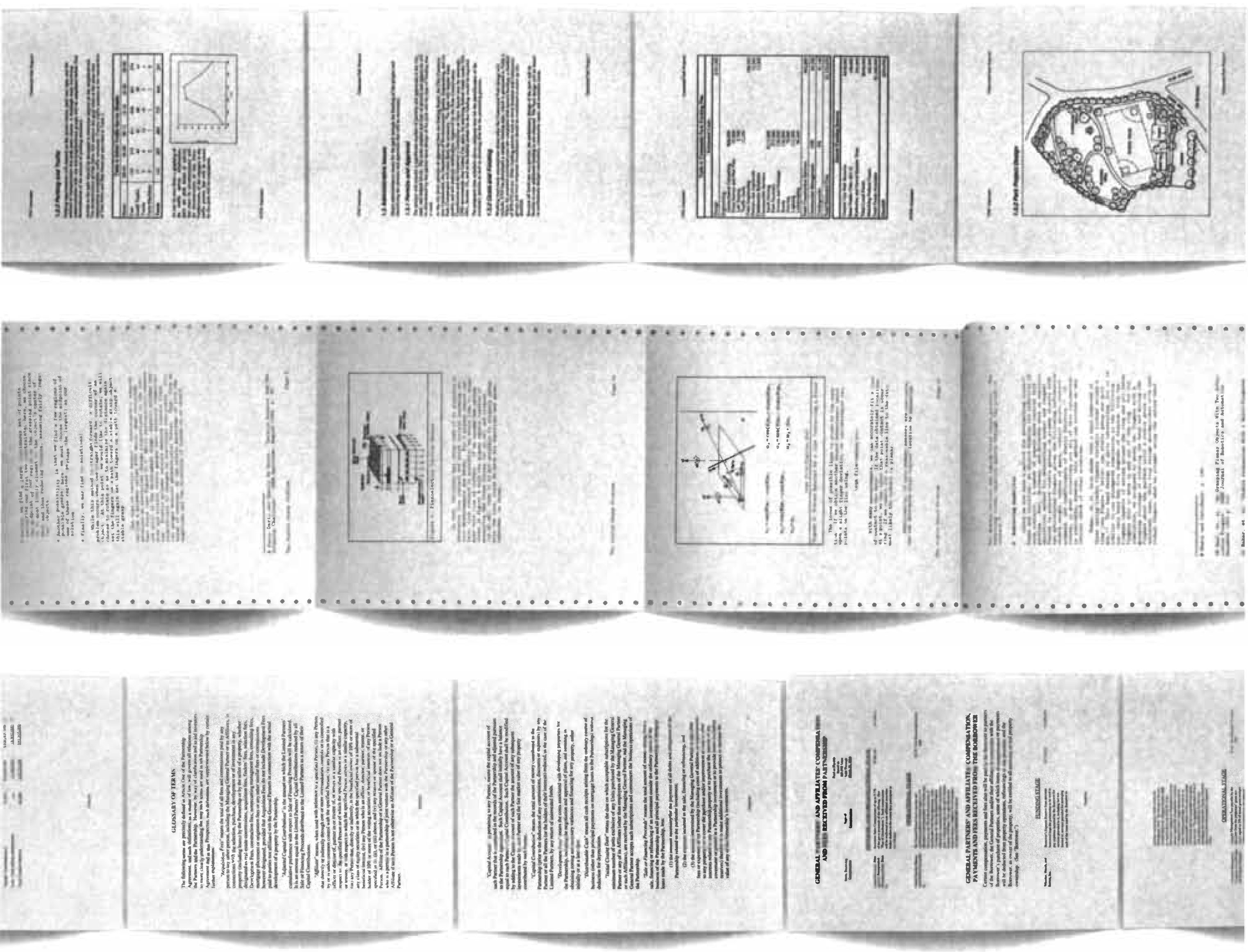


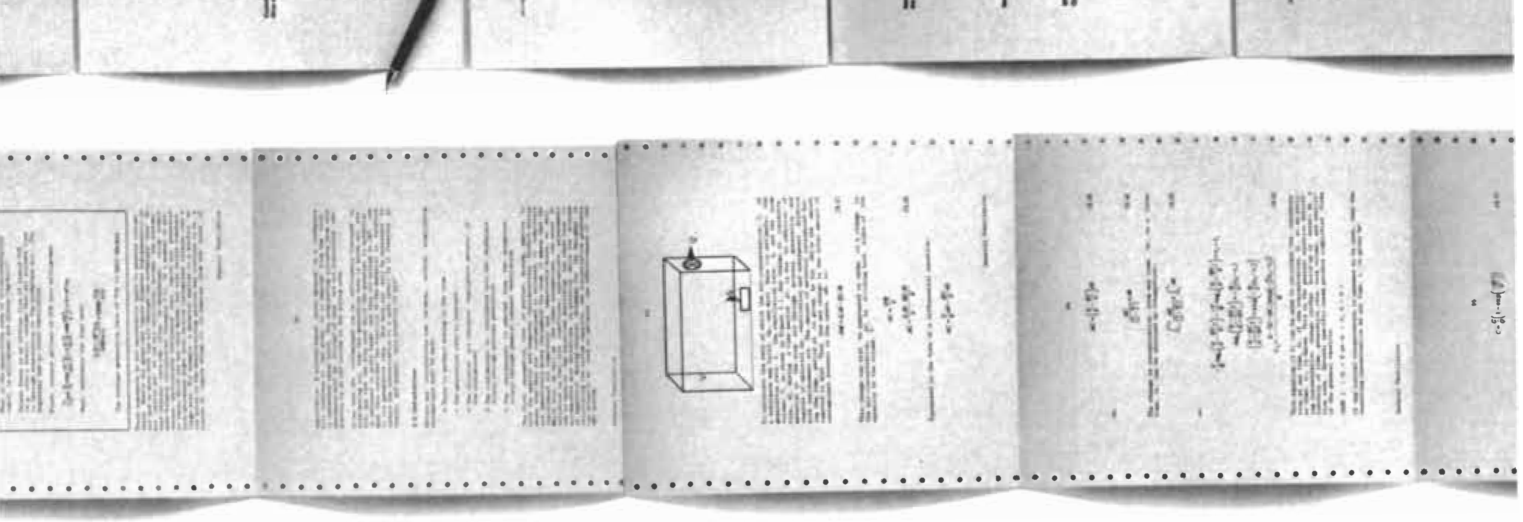
enough of a challenge,



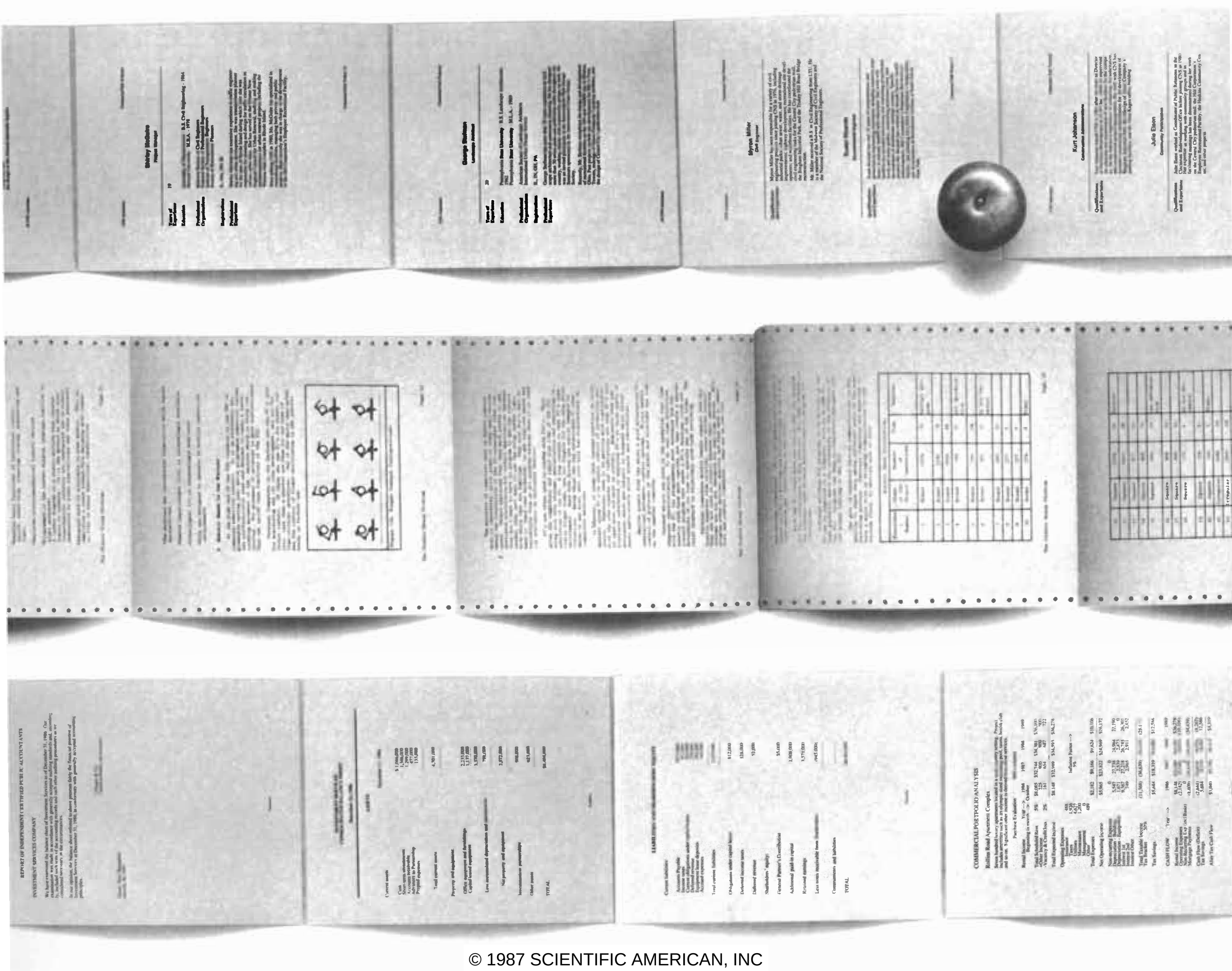


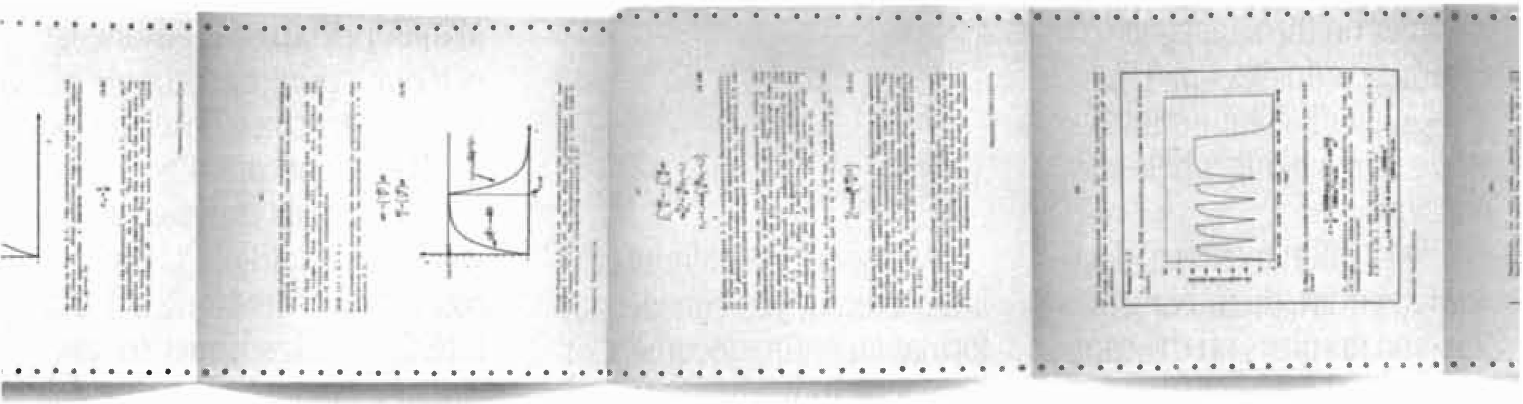
this time you have to do



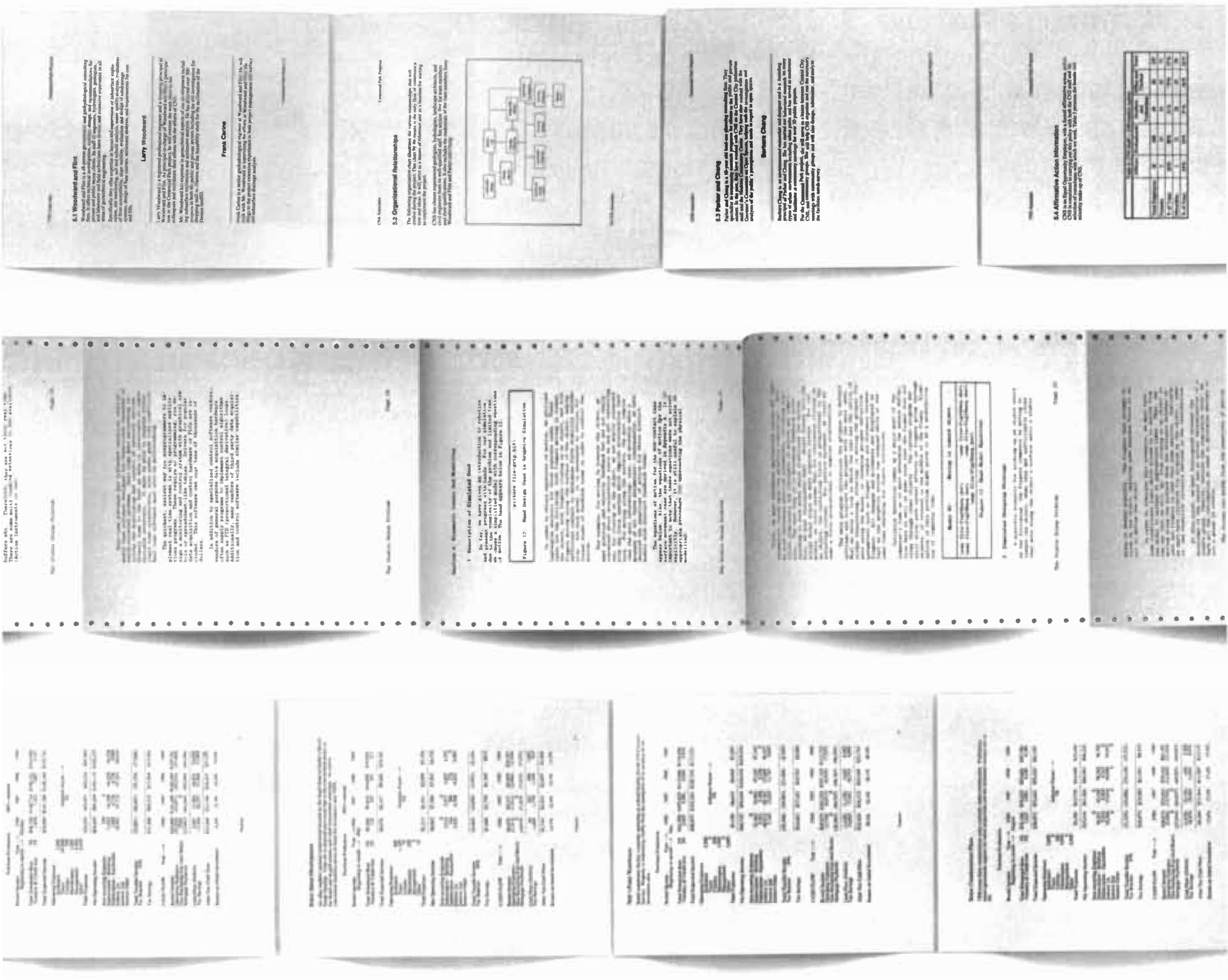


a structured document,





mixing text and graphics.

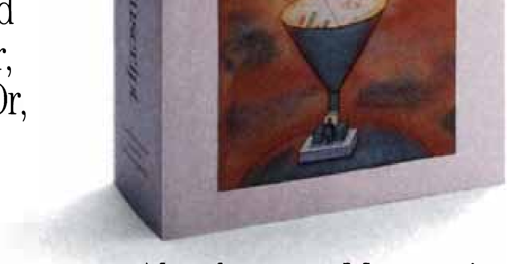


May we suggest Manuscript.

Needs like mixing text and graphics on the same page. Or, creating complex equations. Or, working with columns and tables. Or, making wholesale revisions.

Very often technical documents require the mixing of text and graphics on the same page. Graphics from Lotus® 1-2-3,® Symphony® and Freelance® Plus. With Manuscript you can import spreadsheets and charts, as well as diagrams and scanned images.

Another feature designed specifically for the writer of technical documents is Manuscript's integrated Outliner. When you collapse the document you're able to navigate the outline much easier than wading through the entire piece. You can move a page, or



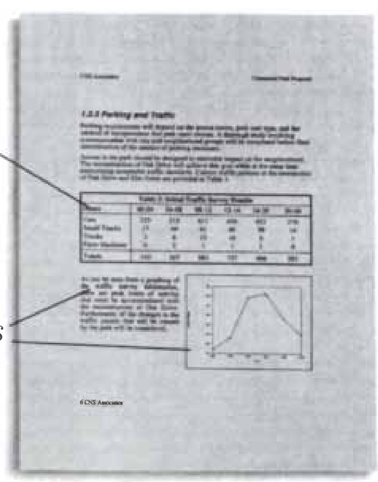
Also, because Manuscript is structured, you can globally format an entire document or format by individual sections.

Imagine changing all headlines from 14 point plain text to 16 point bold, by a simple menu selection.

How about the age old dilemma of proof reading revisions? Manuscript's Document Compare feature highlights changes between revisions. You might call it fool-proofreading.

Also, Manuscript automatically sizes and generates math equations as well as upper and lower case Greek symbols, diacritical marks and brackets. No longer do you need to literally cut and paste to create equations.

With our powerful Print Formatter you have complete control

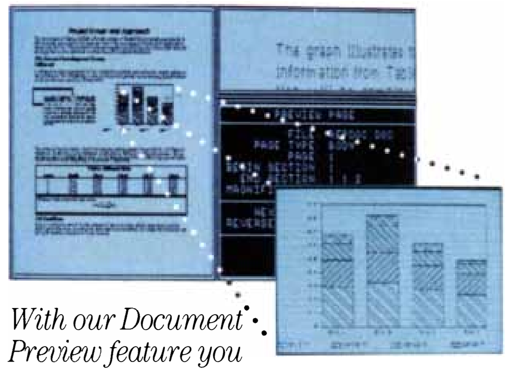


Spreadsheets, graphs and charts from 1-2-3® and Symphony® can easily be mixed with text on the same page.

You can throw away your scissors and glue, cut and paste are a thing of the past.

As for the quality of output, Manuscript takes full advantage of today's printing technology, from dot-matrix to laser, including PostScript® printers.

Manuscript is designed to work on most IBM® PCs and compatibles.* Its familiar 1-2-3 interface makes it easy to use. And our Manuscript evaluation kit makes it easy to try. For $10.00, you'll get a presentation disk, working software, and a tutorial manual. To get your evaluation kit, call 1-800-345-1043, ask for lot BF-1450. Or, for more information, see your authorized Lotus Dealer, or write Lotus Development Corp., 55 Cambridge Parkway, Cambridge, MA 02142.



With our Document Preview feature you can see how any page will look before it's printed. Zoom capability lets you take a closer look at graphics and equations.

Lotus Manuscript™

*Manuscript and Manuscript Evaluation Kit require 512K and a hard disk.

by the unreliability of historical temperature records. Now two investigators with the U.S. Geological Survey have found what may be evidence of global warming under the frozen tundra of northern Alaska.

Permafrost provides an excellent medium for recording fluctuations in surface temperatures. The deeper one digs, the further one travels back into thermal time. The absence of water seepage in permafrost makes it easy to calculate the rate at which heat descends through it.

By lowering thermometers into several dozen abandoned oil-exploration wells on Alaska's North Slope, Arthur H. Lachenbruch and B. Vaughn Marshall found that the surrounding permafrost has warmed up by as much as seven degrees Fahrenheit in the past 100 years. Writing in *Science,* they suggest that although the warming may be a strictly local phenomenon, it could turn out to be an early warning of the greenhouse effect.

Records of surface temperatures gathered around the world for the past century, however incomplete, do indicate a warming of perhaps one degree F. This global warming should be accentuated at the poles, Lachenbruch explains, by various "positive feedback" phenomena. When sunlight melts snow, for example, the dark ground under the snow absorbs more heat, causing additional melting and warming. Moreover, Arctic air convects slowly, keeping relatively warm air trapped near the surface.

Other experts on Arctic geology, while praising the permafrost study, caution that it does not prove conclusively that the Arctic is warming, let alone the entire earth. Thomas E. Osterkamp of the University of Alaska's Geophysical Institute notes that records of temperatures in the Arctic are "pretty spotty." Such records, says Osterkamp, could corroborate the findings reported by Lachenbruch and Marshall, and they could also help to establish more precisely the relation between changes in temperature at the earth's surface and changes within the permafrost.

Osterkamp thinks the case for long-term permafrost warming is "solid," but he points out that his own research has uncovered evidence of a recent cooling. Working with shafts drilled specifically to record permafrost temperatures, he has discovered a cooling of nearly two degrees F. over the past three years, particularly near Alaska's coast. He speculates that the recent cooling might be caused by shrinkage of the thickness of the snow cover, which insulates the ground. Osterkamp emphasizes, however, that his data demonstrate "everything is not proceeding in the same way."

Gunter E. Weller, also of the Geophysical Institute, suggests the greenhouse effect may even be counteracted by an approaching cooling trend. Noting that the last ice age ended about 12,000 years ago, he says: "We are due for a change toward cooler climates."

Volcanic Winter

Kilometer-deep core samples of ice from Greenland, tree rings from eastern California and historical records from China, Korea and ancient Rome—from these clues a group of investigators has reconstructed the climatic effects of the enormous cloud of dust and ash thrown up more than 2,000 years ago by volcanic eruptions of Mount Etna in Sicily. The workers, Kevin D. Pang and David Pieri of the Jet Propulsion Laboratory (an astronomer and a volcanologist respectively) and Hung-hsiang Chou, professor of East Asian languages and cultures at the University of California at Los Angeles, presented their work at a recent meeting of the American Geophysical Union.

The Greenland ice cores are stratified: every year a new layer of snow covers the last. It is possible to count the number of layers in a core sample because the chemical composition of the atmosphere (and therefore of the top surface of the snow) changes periodically with the seasons. The Etna eruptions appear as peaks in the acid content of the layers because the sulfur dioxide emitted by the volcano combined with water in the atmosphere to form sulfuric acid. Ice cores show that the volcano erupted three separate times during a period within about 10 years of 50 B.C. The rings of 5,000-year-old bristlecone pines in California show that the trees suffered extensive damage from frost (probably because of cold weather brought on when the volcanic cloud blocked the sun) in about 42 B.C.

Chinese, Roman and Korean histories date the event more precisely. Some Roman chroniclers viewed the eruptions and their effects as signs of divine displeasure at the murder of Julius Caesar. According to Plutarch, after Caesar's death there were earthquakes and "the obscuration of the sun's rays. For during all that year its orb rose pale and without radiance." The volcanic cloud must therefore have formed after Caesar's assassination on March 15, 44 B.C. Other Roman records state that a red comet, so bright that it was visible in daylight, appeared in the northwestern sky later that year. According to a history written by Dio Cassius in the third century A.D., "the majority...ascribed it to Caesar, interpreting it to mean that he had become a god." Chinese and Korean histories record a red daylight comet in the northwest in May and

ABANDONED OIL WELL on Alaska's North Slope provides a way to measure temperatures in permafrost. Geophysicists fill wells with diesel oil and lower sensors into them.

June of 44 B.C. The comet's color can be attributed to volcanic dust in the atmosphere, fixing the date of the first eruption at sometime between mid-March and mid-May of 44 B.C.

The volcano's climatic effects were so severe that as far away as China six consecutive grain harvests (three years' harvests) failed. According to bookkeeping records maintained on certain bamboo sticks that have recently been excavated, in some areas the price of grain was inflated by as much as 1,400 percent between 44 and 42 B.C. The chronicles of the Han dynasty state that in April of 43 B.C. "it snowed. Frost killed mulberries." In May "the sun was pale blue and it cast no shadows."

Historical records confirm that the effects were due to several distinct eruptions. The Han dynasty chronicles indicate that by October of 43 B.C. the sun seemed to have recovered, but in the spring of 42 B.C. the sun, moon and stars again appeared "veiled and indistinct." In attempting to match such historical records with the peaks and valleys of acidity in ice cores, the workers noted what they think is a "missing summer": the ice cores show no record of a summer (that is, no summerlike change in chemical composition of the atmosphere) between acid peaks representing two of the eruptions that, according to histories, were about a year apart (in May of 43 B.C. and March of 42 B.C.).

BIOLOGICAL SCIENCES

Borrowed Time

Once destroyed, a neuron in the adult central nervous system is not regenerated. Therefore the loss of neurological function that often accompanies stroke, severe head and spinal-cord injuries and diseases that attack the brain is generally irreversible. There is hope that this grim state of affairs will not persist forever. Work with hamsters suggests that at least one critical function, control of circadian, or daily, rhythms, can be restored with transplanted brain tissue.

The hamster study involved two groups of brain cells, nestled in the hypothalamus, that act as a biological clock. Called the suprachiasmatic nuclei, they spur animals to wake, eat and engage in other activities at regular times in a 24-hour period even if the animal is isolated from a normal light-dark cycle. An animal that suffers damage to this region loses its ability to establish circadian rhythm; in human beings the loss may lead to chronic insomnia and other disorders.

At the annual meeting of the Society for Neuroscience, investigators from Barnard College, the University of Cincinnati College of Medicine, the Mount Sinai School of Medicine and the University of Massachusetts at Amherst described experiments done at Barnard on adult hamsters whose suprachiasmatic nuclei had been surgically destroyed. When they were put in the dark, the brain-damaged hamsters ran on wheels in their cages at random times. When the investigators implanted suprachiasmatic nuclei removed from hamster fetuses into the brains of 14 adult hamsters, 10 regained their daily running rhythms.

Perhaps the most dramatic finding was that the growth of new neural tissue followed implantation and apparently played a role in the restoration of the circadian function. As time passed, axons extended from the grafted tissue out into the brain. In some hamsters axons also grew from the retina back to the transplant.

By doing further experiments the workers hope to elucidate the workings of the biological clock more completely. They also hope to learn how the suprachiasmatic nuclei adjust their signals to the brain in response to environmental cues. Jet lag results from a temporary conflict between the setting of the biological clock and a changed light-dark cycle; the suprachiasmatic nuclei, which receive input from the optic nerves, can only gradually cause the body to adjust to a new cycle.

More fundamentally, the experiments should advance understanding of how brain-damaged animals respond to transplanted neural tissue. The information should help investigators who hope to treat victims of Parkinson's and Alzheimer's diseases with brain grafts. "If you understand how one system recovers," Rae Silver of Barnard notes, "that should help you understand others."

Abzymes

Antibodies are a class of immune-system proteins that can recognize and specifically bind to a vast range of compounds. In molecular biology and its technological outgrowths they serve in the detection and purification of specific molecules. For many of its tools, however, molecular biology must draw on the much more limited repertory of enzymes, proteins that not only bind to particular targets but also transform them by catalyzing chemical reactions. The toolbox of molecular biology would be enlarged enormously if antibodies could be given the chemical talents of enzymes.

Writing in *Science,* groups working independently at the Research Institute of Scripps Clinic and the University of California at Berkeley describe the first examples of antibodies that act as enzymes, catalyzing chemical reactions. The Scripps workers suggest that the new catalytic antibodies be called abzymes.

The success of both groups is founded on a principle of enzyme action proposed in 1948 by Linus Pauling. An enzyme hastens a chemical reaction by lowering an energy barrier along the reaction path. One way it does so, Pauling proposed, is by binding to the transition state: an unstable chemical intermediate formed partway through the reaction. The transition state is thereby stabilized, reducing the energy needed to form it. The Scripps and Berkeley workers reasoned that since tight and specific binding is the forte of antibodies, an antibody with an affinity for a molecule that resembles the transition state of a chemical reaction should catalyze the process.

The Scripps investigators—Alfonso Tramontano, Kim D. Janda and Richard A. Lerner—built on an earlier, partial success they reported in *Proceedings of the National Academy of Sciences.* In that work they identified a molecule that mimics the probable transition state in the hydrolysis (a breakdown process entailing a reaction with water) of certain simple organic molecules known as carboxylic esters. They injected the transition-state analogue into laboratory animals and then isolated and cloned cells secreting an antibody to the substance. The resulting monoclonal antibody displayed some but not all of the behavior of an enzyme: it hastened the breakdown of the esters it was tested against but was itself transformed in the process. The antibody was in effect used up, whereas a true enzyme is unchanged after catalyzing a reaction.

On the basis of presumed details of the interaction between the transition state and the antibody's combining site, the workers modified the esters. In *Science* the group now reports true enzymatic activity. Acting on the new esters, the antibody accelerated the hydrolysis several hundredfold and retained its activity even after each molecule of antibody had catalyzed the reaction hundreds of times over.

Instead of isolating a new catalytic antibody, Scott J. Pollack, Jeffrey W. Jacobs and Peter G. Schultz of Berkeley began by demonstrating the catalytic abilities of an existing antibody. The antibody binds to a molecule whose structure is analogous to that of the transition state formed during the hydrolysis of an organic carbonate compound. The Berkeley group found that the antibody's binding affinity gave it catalytic activity. Later the investigators generated a new catalytic antibody by injecting experimental animals with an analogue of the transition state in another carbonate-hydrolysis reaction.

Both results point the way to a general strategy for enzyme design: in virtually any reaction for which the transition state is known, an antibody capable of binding to and stabilizing the transition state could in principle be generated. The combining site of the antibody might then be chemically modified to bolster the catalytic effect.

The practical flowering of the technique may begin, the workers suggest, with the making of catalytic antibodies that can split the peptide bonds linking the building blocks of proteins. Such antibodies could, for example, provide molecular biology with an array of protein-cutting tools and give medicine a set of weapons targeted to the distinctive proteins of cancer cells, viruses and blood clots.

Probing the Mind's Eye

Investigators at the National Institute of Neurological and Communicative Disorders and Stroke have quietly begun experimenting with human volunteers to learn more about how electronic stimulation of the brain can induce visual sensation. F. Terry Hambrecht, head of the neural prosthesis program at the institute, says two experiments have been conducted, one last July and the other in December. Both experiments involved sighted patients at the University of Western Ontario who were undergoing brain surgery and were locally anesthetized.

During the operation the neurosurgeon inserted a microelectrode into the section of the brain known as the visual cortex. As the surgeon stimulated different areas of the cortex, the patient indicated where points of light, called phosphenes, appeared on a calibrated screen before him.

The experiments, disclosed at a National Institutes of Health workshop, mark a resumption of controversial research begun about 20 years ago by Giles S. Brindley of the University of Cambridge. He opened the skull of a blind volunteer and placed a cap lined with 80 electrodes on the subject's brain. Brindley ran a wire bundle from the electrodes through a hole in the subject's skull to a second cap, studded with 80 miniature radio receivers, and sewed the scalp up. When Brindley transmitted pulses to the apparatus, the patient reported "seeing" white points of light.

This first attempt at electronically inducing visual sensation in the blind was followed by similar experiments at a handful of U.S. institutions. The experiments continued for about 10 years. Then, Hambrecht says, peer-review groups concerned by the disparity between the experiments' modest results and sensational media coverage "lost their enthusiasm" for the research. "It stirred up a lot of skepticism," he recalls.

He says the main difference between the recent experiments and those carried out by Brindley is that "our electrodes are about 100 times smaller than his." Brindley's relatively large, flat electrodes sat on the brain's surface and needed strong pulses to generate phosphenes, Hambrecht says. The new electrodes, thinner than a human hair, can be inserted directly into the cortex, bringing them in closer contact with specific nerve cells. The electrodes therefore require much less power than the earlier ones, according to Hambrecht. The smaller electrodes might also be placed closer together without causing their phosphenes to merge. The older electrodes, Hambrecht remarks, gave rise to "severe interaction between the visual phosphenes."

Eventually, says Hambrecht, if results from the latest round of experiments prove favorable, investigators will implant arrays of electrodes in blind subjects for long-term testing. Hambrecht stresses that the research is "very preliminary" and that it is too soon to know whether a visual prosthesis can ever be built.

MEDICINE

Social Agenda

In recent months the U.S. National Academy of Sciences has produced a series of reports that amount to a social agenda, or perhaps even a lumberyard in which planks for a political platform could be found. Such topics as crime and career criminals, the menace of AIDS and the human impact on the environment have received the academy's attention, mostly in the form of reports by authoritative panels assembled by the National Research Council, the academy's research arm.

To this list the academy has now added teen-age pregnancy. The facts are familiar, but they are presented in a way that keeps them from being dismissed as the parochial concern of a single subpopulation or social class.

This year, notes a National Research Council panel assembled to study the problem, about a million U.S. teen-agers will become pregnant. In round figures the outcome will be 470,000 births, 400,000 abortions and 130,000 miscarriages. These facts, the panel notes, constitute a major social problem to which "no coherent policy" is addressed.

The panel, whose chairman was Daniel D. Federman of the Harvard Medical School, consisted of physicians, social scientists and public-health clinicians. Five foundations commissioned the study and provided the money for it. The foundations (Ford, Rockefeller, William and Flora Hewlett, Robert Wood Johnson and Stewart Mott) have for the past decade put money into programs on teen-age pregnancy.

The panel found that life is hard for many teen-age mothers and their children. The mothers are at high risk for complications of pregnancy and for having low-birth-weight babies, who are more than usually likely to face physical and developmental problems. Teen-age parents have "severely limited career opportunities" and so are likely to become dependent on public assistance. Indeed, teen-age pregnancy exacts a high cost from society, the panel stated, noting that $17 billion in Federal assistance (Aid to Families with Dependent Children, Medicaid and food stamps) went in 1985 to families begun by births to teen-age mothers. The panel emphasized, however, that the problem is not confined to the poor: teen-age pregnancies occur at all social and income levels.

The panel followed this hard truth with another: efforts to dissuade teenagers from sexual activity are fruitless. Contraception, the panel recommends, should be "the major strategy" and contraceptives should be made widely available to both girls and boys at low cost or no cost, with some experimentation in making them available in schools. Advice on abortion and assistance to girls who choose to have their babies are also crucial, the panel said.

In view of the response by Secretary of Education William J. Bennett ("a dumb policy—school-based birth-control clinics") it seems unlikely that the Reagan Administration will give the panel's recommendations much attention. The report has stirred interest in Congress, however, and probably will lead at least to committee hearings on the subject in the near future. Federman expresses the hope that it will also stimulate action at all levels of government and induce private organizations to devote money and effort to the problems the panel has identified.

What of the other items on the

National Academy's social agenda? Frank Press, the academy's president, said he believes the studies are having a "big impact" in terms of congressional response and the allocation of resources in the Federal budget. "We are careful about taking on social questions," he said. "We limit ourselves to issues that are researchable, so that we can present data on what does and does not work."

Bottle to Throttle

Ten Navy pilots recently spent two hours downing pure alcohol mixed in diet soft drinks. They stopped when the highway patrol of many states would consider them legally intoxicated: when their blood contained .1 gram of alcohol per deciliter. Fourteen hours later the alcohol (the equivalent of what is in from five to seven six-ounce glasses of wine) was gone from their blood. Feeling perfectly fit to fly, the pilots took the controls of a flight simulator that mimics a craft they regularly operate: a four-engine P-3C Orion antisubmarine aircraft. How did they fare?

Not well. On almost every measure they handled the takeoff and instrument landing of a disabled craft less effectively than they did after having avoided alcohol for 48 hours. The plane lacked power in two engines on one side, making it difficult to keep the vehicle from yawing. The pilots' performance was measured by the degree to which the subjects strayed from an ideal course.

The study was done by Jerome A. Yesavage and Von Otto Leirer of the Stanford University School of Medicine. They conclude that a "hangover effect" may significantly impair a pilot's "ability to perform critical flying maneuvers"—even 14 hours after the individual stops drinking. The finding contradicts the opinion of private pilots who, according to a survey, believe it is safe to fly four hours after drinking. The study further highlights the inadequacy of subjective feelings as a barometer of whether a skilled person is too hung over to handle a complex task, such as flying an airplane, operating heavy equipment or perhaps driving a car in a storm.

Just how having been intoxicated can impair performance after alcohol has disappeared from the blood is not entirely clear. Nevertheless, Yesavage and Leirer point out that alcohol can temporarily impair "working" memory and hence the ability to divide one's attention among several tasks. It also reduces the ability to perform nonroutine tasks. Both abilities become crucial in a flight emergency.

The workers emphasize that their report is preliminary; they are now conducting a four-year study to evaluate how such factors as a pilot's age, fatigue and health influence the effects of consuming varying amounts of alcohol. For now, they say, it is too early to determine whether the existing Federal Aviation Administration's "bottle to throttle" rule should be changed. That rule requires all pilots to put a minimum of eight hours between drinking and flying. (Commercial airlines and the armed services tend to have stricter rules.) It further prohibits flight when a pilot has a blood-alcohol level of .04 gram or higher.

In the meantime the FAA is already considering adding bite to its regulations by authorizing random alcohol testing. In a call for comments on the proposal the agency reports that from 1968 through 1985 some 7.5 percent of 5,853 pilots who died in accidents had a blood-alcohol level higher than .04. The FAA and the Air Line Pilots Association point out, however, that no fatal accidents involving scheduled flights by a major U.S. commercial airline have been attributed to a pilot's abuse of alcohol.

In another study of pilots, Martin C. Moore-Ede of the Harvard Medical School has found that commercial pilots frequently suffer from intense drowsiness, sometimes to the point of nodding off in the cockpit. Indeed, entire crews have been said to have fallen asleep as a plane cruises along on automatic pilot.

AIDS: Hope...and Warnings

AIDS researchers around the world seem to describe new findings daily. Some reports suggest that therapies or vaccines against AIDS may be in hand soon. Others suggest that the disease has not yet played all its epidemiologic and genetic cards.

A team including Robert C. Gallo of the National Cancer Institute and led by investigators at the Repligen Corporation in Cambridge, Mass., has isolated a fragment of protein from the AIDS virus that could stimulate the production of antibodies. The fragment may have potential as a vaccine; it is being tested in chimpanzees.

A discovery that may lead to a therapy for those already infected with the virus has been reported by a group from the University of California at San Francisco. In the blood of three men who had tested positively for AIDS antibodies but had not contracted the disease, the workers found a group of lymphocytes (white blood cells) called *T*8 cells that seem to suppress the replication of the AIDS virus. The San

Francisco workers think *T*8 cells grown in culture might serve as a treatment for AIDS carriers.

At the National Institute of Mental Health researchers hope to exploit a molecule on the surface of the AIDS virus that enables it to bind to receptors on cells it infects. These investigators have found that a string of amino acids called peptide-*T*, which resembles the AIDS attachment molecule, could prevent the AIDS virus from invading the cells by blocking the cells' receptors. Other investigators have not yet confirmed the findings. Tests of a drug based on peptide-*T* are expected to begin this year.

Perhaps the most startling report concerning therapy comes from Zaire, where workers from the University of Kinshasa and the University of Paris have been injecting human subjects with a substance that it is hoped will protect them from AIDS. The subjects harbor the AIDS virus but have not yet developed symptoms. The team has not revealed the nature of the substance, but it is reportedly a protein fragment that stimulates the production of lymphocytes able to kill the virus. If the preliminary tests are successful, the program will be expanded.

Investigators stress that progress should not breed complacency; indeed, even widely accepted theories about the disease AIDS may still be imperfect. It has been generally believed, for example, that the virus can enter the body only through the bloodstream. But in vitro research at the National Institute of Allergy and Infectious Diseases suggests that the virus may infect cells directly during anal intercourse rather than first passing into the bloodstream through breaks in rectal tissue, as has been widely thought. Rectal cells, the investigators say, have receptors like those found on *T*-4 cells, which the AIDS virus attacks.

Is there only one AIDS virus? Investigators headed by Luc Montagnier of the Pasteur Institute have isolated a retrovirus that has apparently caused some inhabitants of West Africa to develop AIDS symptoms. The virus is different enough from the first known AIDS virus to be undetectable with standard tests, according to the Montagnier team. The existence of the new virus lends weight to warnings that developing a universal vaccine may be very difficult.

A dire note has also been sounded by Mark E. Whiteside, codirector of the Institute of Tropical Medicine in Miami, Fla. Based on his five-year observation of Belle Glade, Fla., whose population shows the highest rate of AIDS mortality in the U.S., Whiteside says he believes environmental factors, including frequent exposure to disease-carrying insects, may play a more important role in the transmission of AIDS than has been generally thought. He says investigators from his organization and from others are beginning to study the potential role of insects in the spread of AIDS.

Therapy by Mimicry

Among the factors promoting coronary heart disease is a high level of cholesterol in the blood. For 30 years it has been known that thyroid hormones reduce circulating cholesterol by stimulating liver function, but efforts to administer agents that mimic thyroid-hormone activity therapeutically have been stymied by the fact that they increase heart rate. A. H. Underwood and his colleagues at Smith Kline and French Research Limited in England may have found a solution to the problem. They have developed a thyromimetic that reduces cholesterol levels but does not cause an increased heart rate. In *Nature* the workers report that they started with the natural thyroid hormone T_3, which has a known effect on the heart and the liver. The team then measured the effect of a series of small changes in that hormone's molecular structure on heart and liver activity in animals. One of the thyromimetics they tested, SK&F L-94901, was as effective as T_3 in stimulating liver function and lowering the cholesterol level and yet was a thousandfold less active in its effect on the heart.

The liver is the primary organ where the uptake of low-density lipoprotein (LDL) particles, the major carriers of cholesterol in the blood, takes place. Thyroid hormones and mimetics increase the expression of LDL receptors on the liver-cell surface, thereby facilitating passage of the LDL particles through the cell wall; once it is inside the cell, the LDL is broken down. The number of LDL receptors on the cell surface depends on the rate at which the thyroid hormone binds to its own receptors in the cell nuclei. Unfortunately the cardiac-cell receptors for thyroid hormones are similar to, if not identical with, the liver-cell receptors. Increased LDL destruction by thyroid-hormone therapy, in other words, is coupled with increased cardiac output—not a desirable result in patients with coronary heart disease.

The mechanism by which the thyromimetic SK&F L-94901 uncouples heart and liver response rates is, says Underwood, more significant right now than are possible therapeutic applications of the agent some years down the line. In assays of the thyroid hormone's interaction with cultured heart and liver cells, SK&F L-94901 does not have an organ-selective effect. Only the animal studies showed a huge difference in the binding rate of the agent in the heart and the liver.

This selectivity suggests that the thyroid-hormone receptor's binding affinity is not the limiting factor in hormone effectiveness—that factors controlling how the hormone gets inside the cell to the receptor exert a stronger influence. According to Underwood, the study supports earlier work describing a "nuclear pump," or concentrating mechanism, for thyroid hormone in the liver. "This nuclear pump," concludes the Smith Kline team, "could be the source of selective access if the liver was more active *in vivo* in pumping SK&F L-94901 into the nucleus than was the heart."

TECHNOLOGY

Tagging Nuclear Weapons

As nuclear weapons grow smaller and more mobile, officials who seek arms reductions or limitations face an increasingly difficult problem: how to verify that the other side has not exceeded treaty limits. Some arms controllers have proposed tagging warheads or missiles. Weapons without tags, spotted by a spy satellite or other means, would represent a violation of the treaty. Tagging might also aid verification of treaties that restrict weapon systems to certain locations.

"A number of people are working on this," says Roger L. Hagengruber, who oversees verification research at the Sandia National Laboratories. "It won't be a revolution when it's done, but it could be very important." For a tag to be effective, Hagengruber says, it must be impossible to copy, alter or remove from a weapon. A tag could take the form of an electronic chip that transmits a coded signature. Or a small area on the weapon's surface could be scanned acoustically; the pattern of microscopic bumps and scratches would serve as a "fingerprint."

If human inspectors were unacceptable, tags could be monitored by satellites, robots or unmanned drones that fly over missile sites. Steve Feder of Harvard University's John F. Kennedy School of Government, who studied tagging of mobile missiles for Lawrence Livermore National Laboratory a year ago, notes that in a "very sophisticated system" a drone might have sensors that could remotely detect plutonium (a component of all nuclear bombs); wherever the drone found plutonium it would check for a tag.

In another scheme, Feder says, missiles might be outfitted with radio receivers equipped with memories and tuned to signals from the U.S. Navy's Navstar satellite system, with which planes and ships determine their position. The device would record movements of the missile; periodically each side would interrogate the other side's tags and check the information against satellite intelligence.

U.S. military planners have objected to the idea, Feder observes, because it might reveal predictable patterns in the movement of mobile missiles. Indeed, the main challenge for adherents of tagging, according to Milo D. Nordyke, head of verification programs at Lawrence Livermore, is designing "a system that doesn't permit targeting." The motivation for building mobile missiles such as the Soviet SS-20, Nordyke emphasizes, is to create a relatively invulnerable retaliatory force.

Concern over targeting would be most acute, experts say, if tags emitted signals that were monitored by a satellite; such a system would otherwise be desirable because it would involve minimal intrusion. Hagengruber believes a monitoring satellite could be designed so that it is unable to pinpoint a transmission's source; a time delay could also be built into the satellite's signal processor.

Some experts advocate tagging and monitoring weapons as they are built rather than after they are deployed. A plan called the perimeter portal system, reportedly being studied by the U.S. Department of Defense, would require that fences outfitted with tamperproof sensors, including seismic sensors that could detect tunneling, be built around missile-production facilities. All weapons leaving the facility would have to pass through a portal, where tags would be checked either automatically or by an inspector.

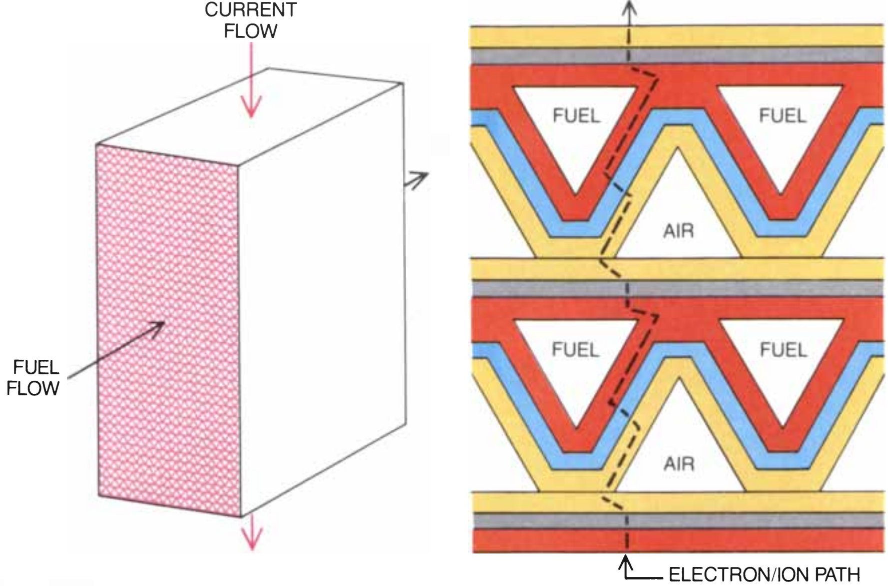
The same distrust that spurs calls for such highly intrusive verification will make it hard—some say impossible—to achieve. Manfred Eimer, assistant director of verification and intelligence at the U.S. Arms Control and Disarmament Agency, says the U.S. and the Soviet Union have not discussed tagging in any formal arms talks. He adds: "You have to have a practical idea before you approach the Soviets."

A Better Battery

Engineers at the Argonne National Laboratory say a fuel cell they are developing for military use could also power electric cars and planes and serve as a highly efficient generator.

Fuel cells, which are batteries that consume hydrogen-based fuels, have been considered too costly and heavy for all but specialized military or industrial applications. Argonne's fuel cell contains inexpensive materials and produces 2.2 amperes of current per square centimeter, almost twice the output of its nearest competitor, according to laboratory officials. The prototype cell is only the size of a book of matches and generates a few watts of electricity. Darrell Fee, who heads the project, claims scaled-up versions could produce megawatts of electricity with the same efficiency.

The highly compact cell is made of sheets of a solid electrolyte (a nonmetallic material that transports ions) lined on each side with ceramic conductors; one side is the anode and the other is the cathode. The sheets are bonded together in a structure resembling corrugated cardboard; air and fuel feed into grooves lined with the anode and the cathode respectively. The charged cathode ionizes oxygen in the air; the oxygen ions flow across the electrolyte barrier and mix with the fuel, which has already vaporized in the extremely hot (1,400 to 1,800 degrees Fahrenheit) cell. Hydrogen from the fuel reacts with the oxygen ions to produce water, which is expelled as exhaust, and free electrons that generate current in the anode.



FUEL CELL designed at the Argonne National Laboratory is made of corrugated sheets of an electrolyte (*blue*) lined on each side with conductors serving as the cathode (*yellow*) and anode (*red*). The sheets are stacked, with connecting layers between them (*gray*). Current is generated when oxygen in the charged cathode channels is ionized; the ions traverse the electrolyte and react with hydrogen in the fuel channel. Electrons released in the reaction cross the connecting layer to the next cathode layer, and the cycle repeats.

The fuel cell's electrolyte is made of yttria-stabilized zirconia, a material car manufacturers use to make oxygen sensors in pollution-control devices. With raw materials worth $2.50 the fuel cell can produce one kilowatt, Fee says; other fuel cells need materials costing from $70 to $80 per kilowatt.

"We'd like to compete head to head with internal-combustion engines," he continues, "and really have an impact on oil imports." The technology will probably be exploited first by utilities, he says, estimating that 50-kilowatt cells twice as efficient as ordinary gas-burning generators could be available in from three to five years.

Marvin Warshay, head of fuel-cell programs for the National Aeronautics and Space Administration, suggests the Argonne fuel cell may work in larger earthbound systems, such as buses and power plants. But he thinks it is "unlikely" that the fuel cell will oust the internal-combustion engine in cars. Edward A. Gillis of the Electric Power Research Institute, which does research for U.S. utilities, is even less optimistic, pointing out that many utilities are already committed to other fuel cells for boosting efficiency. Security restrictions, Gillis adds, have kept the private sector from evaluating Argonne's research.

Fee acknowledges that even though the project is unclassified "we've been limited in what we could say" by sponsors of the program. Although the Department of Energy (DOE) funds Argonne, the fuel cell's primary sponsor is the Department of Defense. The cell is being considered as a power source for cruise missiles, for high-altitude drones that gather intelligence and for the military's planned National Aerospace Plane, an advanced space-transport vehicle.

The Energy Department owns all patents for the fuel cell but is now transferring rights to the patents to a business arm of the University of Chicago, which operates Argonne for the DOE. Within a year, Fee says, the university should be able to sell patent licenses to companies that want to investigate the fuel cell's potential.

"No—Bug Lite"

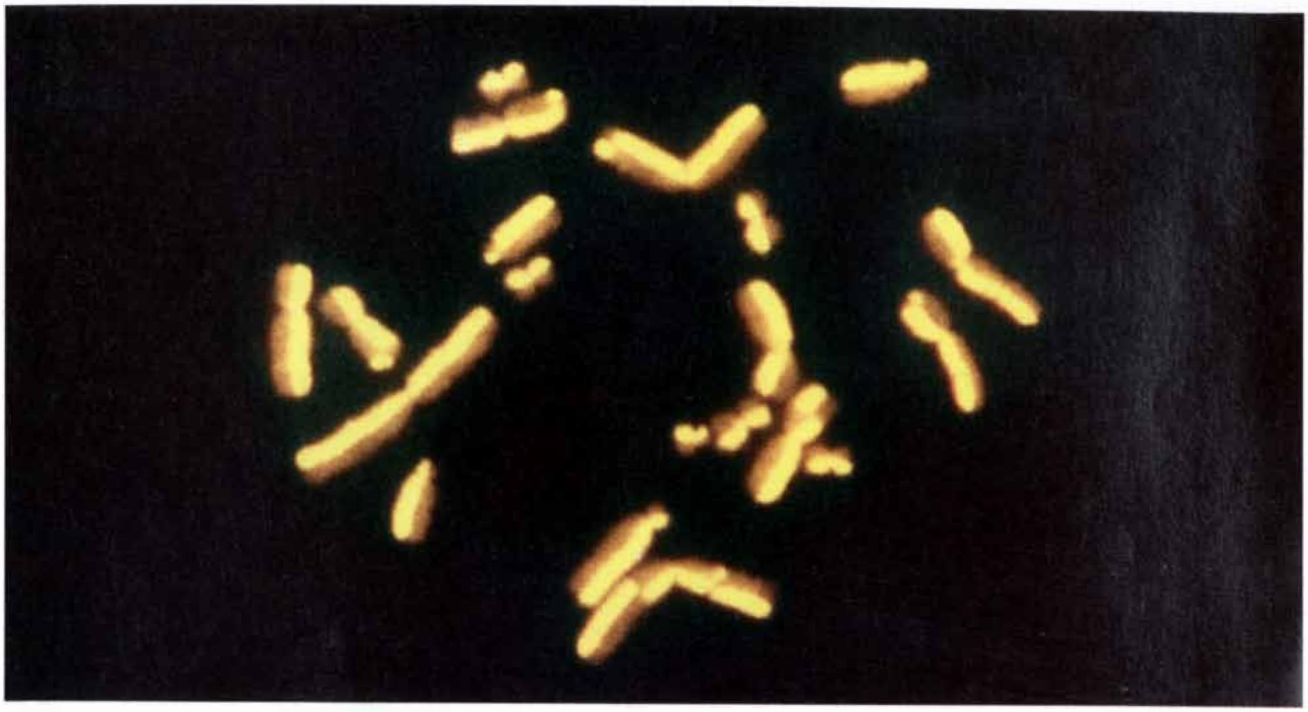
Genetic engineering has made a jump from the clinic to the corner bar. Insulin, growth factor and hepatitis vaccine may soon be joined by a less distinguished product: low-calorie beer. A Cambridge, Mass., company is testing a recombinant yeast that could make "light" beer without costly chemical additives or time-consuming changes in barley preparation.

Beer begins with barley, and the primary constituent of barley is high-calorie starch. Malting processes convert all but 25 percent of that starch into sugar, and then yeast ferments the sugar into alcohol. About a third of the calories in ordinary beer come from the excess, unconverted starch. To make light beer, brewers get rid of surplus starch by either extending the malting period or adding starch-degrading enzymes. As a result another 15 percent of the starch becomes alcohol; the light brew is then diluted to match the potency of regular beer.

Researchers at BioTechnica International, Inc., took from a common industrial bacterium a gene that codes for a starch-degrading enzyme and spliced it into the DNA of brewer's yeast. The engineered yeast can turn excess starch into sugar on its own, eliminating the need for enzyme additives or lengthy malting. The effect is akin to breeding cows that give skim milk; the resulting brew, however, must still be diluted. BioTechnica says its yeast achieves the same conversion efficiency as enzyme additives.

Still, it may be at least five years before the products of genetic engineering come by the case. Recombinant food products face an uncertain regulatory climate in the U.S., where most precedents are being set by medical research. And the company, a biotechnology start-up that has no intentions of marketing its own brand of beer, may have trouble introducing a microbial half-breed to a business that prides itself on tradition.

The key factor in the fate of the yeast may not be an objective one, however. This month at the Brewery Research Foundation in Surrey, England, brewery experts will get a taste of BioTechnica's high-tech beer. The yeast's success will probably depend on how well its beer pleases the palate.



HARLEQUIN CHROMOSOMES testify to the reciprocal exchange of genetic material between sister chromatids (two identical filaments making up a chromosome during cell division). The images resulted from a technique by which the chromatids were chemically altered so that one fluoresces more brightly, when it is stained with a fluorescent dye, than its sister chromatid (*top*). An agent that damages DNA, the constituent molecule of chromosomes, fostered the multiple exchanges seen in the lower image. Such exchanges are analogous to genetic recombination in that the chromatids are left complete: they neither gain nor lose genetic information. Unlike recombination, which creates new assortments of traits because the two chromosomes differ, sister-chromatid exchanges usually have no genetic effect. Sheldon Wolff of the University of California at San Francisco provided the images.

Genetic Recombination

Before an organism reproduces it often reshuffles its genetic information. Chromosomes trade parts in a recombination process whose molecular intricacies are now being unraveled

by Franklin W. Stahl

To reproduce itself successfully a creature must transmit a faithful copy of its genetic information to its offspring. Oddly, it is also part of the basic pattern of reproduction for some of that information to be reshuffled among its various repositories before it is passed on. The repositories are chromosomes: giant, double-chain molecules of DNA on which genetic information is encoded in a sequence of paired chemical units known as nucleotide bases. The reshuffling is known as genetic recombination.

Long before 1953, when James Watson and Francis H. C. Crick described the molecular basis of heredity, the DNA double helix, it was known that individual genes—the units of genetic information—can vary their travel companions. Two traits that were associated with a single chromosome in one individual sometimes turned up on different chromosomes in its offspring. Indeed, measuring the frequency with which any two genes on a chromosome are transmitted together became a standard means of gauging their physical separation: the closer together the genes lie, the more likely they are to be transmitted together.

Without the conceptual framework Watson and Crick supplied, however, the question of exactly how the reshuffling takes place could not be addressed. By now, more than three decades later, studies carried out in the simplest of all organisms, viruses and bacteria, have revealed some of the elegant and intricate molecular pathways of recombination.

Analogous mechanisms can be expected to operate in other organisms, including human beings; genetic recombination, like the DNA it operates on, is almost universal. In complex organisms it generally occurs during meiosis, the process of cell division that yields germ cells (sperm or eggs).

In the precursors of germ cells, as in the individual's other tissues, the chromosomes form two sets: each chromosome has a "homologous" counterpart carrying an almost identical array of genetic information. One chromosome set came from the mother's egg and the other from the father's sperm. During meiosis homologous chromosomes form pairs, and the members of each pair recombine—they swap parts. The swapping leaves each chromosome complete, so that neither one loses any genetic information. Yet the single sets of chromosomes that are parceled out to the germ cells (each of which contains half as many chromosomes as other cells do) represent a mingling of the two parental sets.

Why Recombine?

Homologous chromosomes are by no means identical, and therein may lie the importance of genetic recombination. Each chromosome in a matching pair carries largely the same set of genes, which are now understood to be sequences of nucleotides coding for molecules (usually proteins) that give the organism form and function. Because of mutations accumulated over the history of the species, however, the two copies of a gene may differ in one or more nucleotides and may therefore dictate alternative traits. Part or all of one copy may even be missing or be interrupted by a stretch of extraneous DNA. Recombination means that a chromosome in a germ cell can carry any assortment of the mutations carried on the two parental chromosomes—not simply one set or the other. Limitless combinations of traits can be passed on to the offspring.

One Darwinian analysis holds that genetic recombination evolved for that very reason. In the face of environmental change, passing on varying combinations of traits should give the individual a selective advantage: it betters the chance that at least one descendant will have the precise set of characteristics needed to survive and multiply. Other evolutionists contend that under most conditions recombination entails a selective disadvantage. Since the individual has succeeded in reproducing, the argument runs, the combination of traits specified by the parental chromosomes by definition has proved adaptive. Provided the environment changes slowly, the individual benefits most by "freezing" its genetic legacy rather than reshuffling it.

Under that argument genetic recombination has persisted only because it serves some other purpose. For example, there is evidence that genetic recombination is one manifestation of a repair mechanism that enables a cell to reconstruct a damaged chromosome precisely by borrowing the missing information from a matching, intact chromosome.

Watson and Crick's description of DNA immediately suggested a way to account for the precision with which information is exchanged between recombining chromosomes. It reflects the fact that the nucleotide bases in the double-chain DNA molecule are arranged in complementary pairs, so that the order of bases on one chain specifies the order on the other chain. If, for example, a length of single-chain DNA is exposed on one chromosome during recombination, the segment can recognize the corresponding sequence on another chromosome through complementary base pairing.

How is that chain of DNA first exposed on one chromosome and brought into alignment with a homologous chromosome? What steps cut and rejoin the DNA chains of the interacting chromosomes to reconstruct one or two complete recombinant chromosomes, each one a new combination of the mutations carried by the interacting chromosomes? Those are the questions other workers and I have addressed in more complete molecular models of recombination.

Model Organisms

A. John Clark of the University of California at Berkeley laid the groundwork for future molecular models in 1965 when he embarked on a genetic analysis of recombination in the widely studied bacterium *Escherichia coli*. As do many other bacteria, *E. coli* engages in a kind of sexual behavior. Pairs of bacteria establish a temporary bridge, which enables the "male" to transfer its single chromosome to the "female." The two bacterial chromosomes then swap parts. Clark's work and that of other investigators led to the identification of genes whose protein products catalyze this act of recombination.

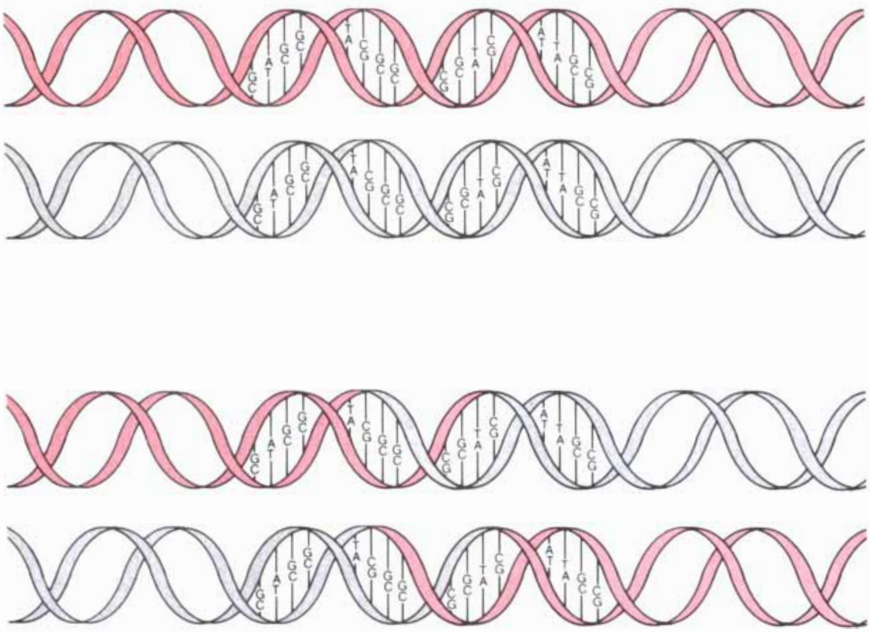
Mutations signaled the presence of the genes. In strains of *E. coli* carrying a mutation in the first of the genes to be identified, called *recA*, recombination was reduced to a thousandth of its normal level. Strains with a mutation in either of the other two genes, *recB* and *recC*, showed a hundredfold drop in recombination.

Investigators then searched for proteins that were present in wild-type (nonmutant) *E. coli* but missing from the mutant bacteria. The work led to the isolation of the *recA* protein and the *recBC* protein, which consists of subunits encoded by the *recB* and *recC* genes. That protein was later shown to contain a third component, contributed by the *recD* gene, and it is now known as the *recBCD* protein. Once the *E. coli* recombination proteins were in hand, their interaction with DNA molecules could be examined in search of clues to their role in genetic recombination.

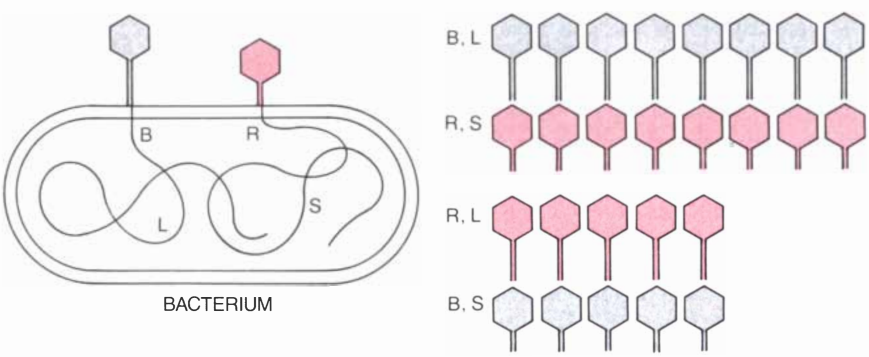
In the test tube the *recA* protein binds to single-chain DNA, which can be extracted from certain small viruses. The protein-coated DNA then invades intact, duplex DNA, causing its complementary chains to separate. The coated chain probes the duplex DNA until it finds a homologous sequence of nucleotides. It forms a new duplex with the complementary chain, which abandons its previous companion chain, and the *recA* protein leaves. The resulting structure is a "D loop," in which the displaced DNA chain loops around the new duplex [*see illustration on opposite page*].

In contrast to the *recA* protein, the *recBCD* protein shows an in vitro affinity for double-chain DNA. Under the right conditions the enzyme is visible in electron micrographs as it enters the DNA duplex at one end and travels along it between the complementary chains. The protein divides the chains as it advances and reunites them in its wake. It reassociates the chains more slowly than it separates them, however, with the result that growing loops of single-chain DNA develop on each side of the moving enzyme [*see illustration on opposite page*]. Under certain circumstances the *recBCD* protein not only separates the DNA chains but also cuts them.

The biochemical observations suggested a scenario for the initial steps of recombination in *E. coli*. The *recBCD* protein separates a length of single-chain DNA from one chromosome. With the help of the *recA* protein



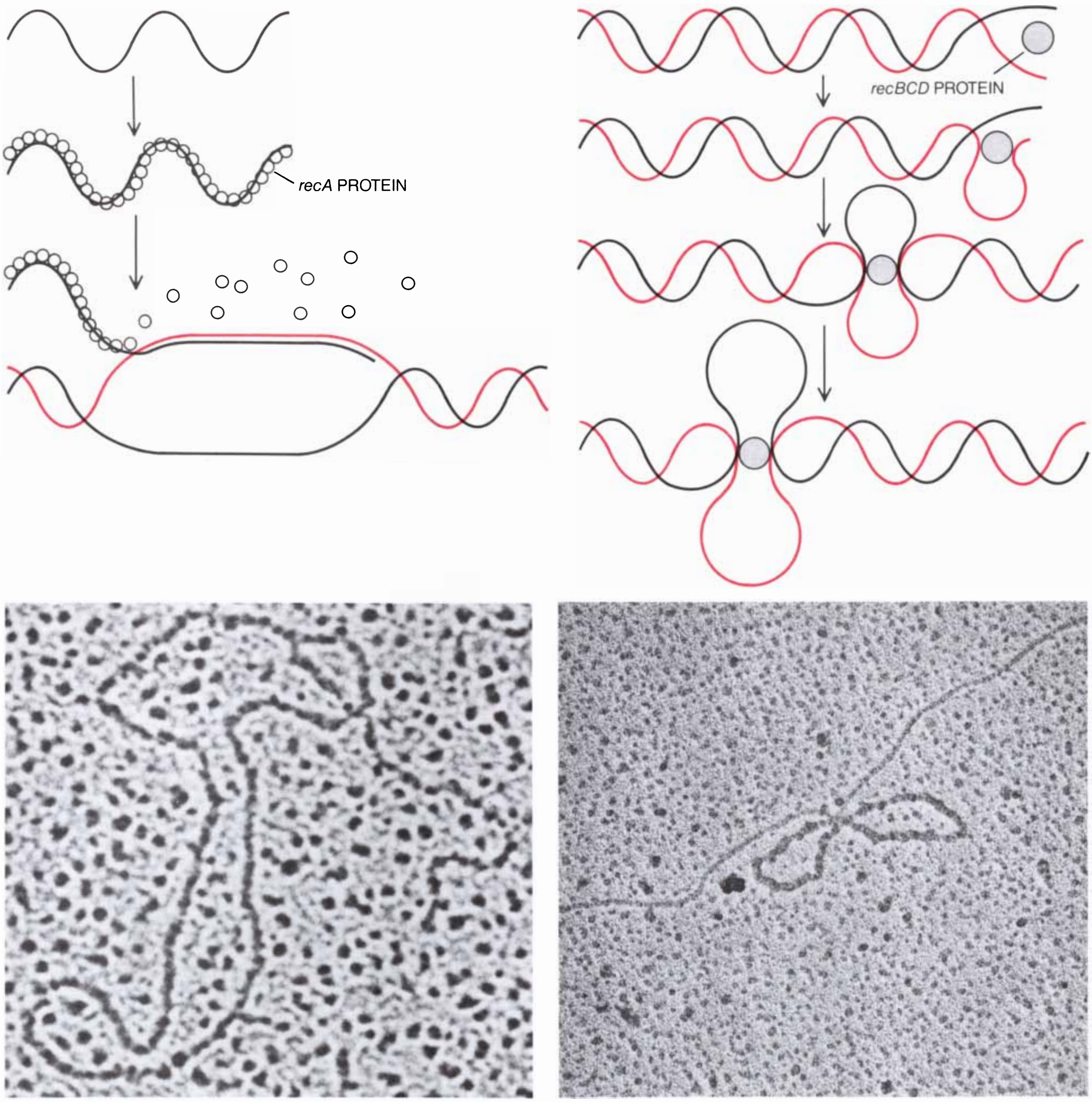
RECOMBINATION of two chromosomes can be understood as an exchange between two double-chain molecules of DNA. Recombination takes place at a region where the interacting chromosomes share a similar or identical sequence of nucleotide bases, the chemical subunits that encode information along the DNA double helix (*top*). The bases on each molecule form complementary pairs: an adenine (*A*) on one chain always pairs with a thymine (*T*) on the other, and a guanine (*G*) always pairs with a cytosine (*C*). Because of those matching rules, a length of single-chain DNA from one recombining chromosome can find the matching sequence on the other chromosome. The chromosomes are thereby aligned, with the result that an exchange of chromosome segments yields complete recombinant chromosomes (*bottom*). Even though the initial chromosomes are identical (or almost identical) at the recombination site, they may carry very different genetic information elsewhere, so that the process results in new assortments of traits.



EFFECT OF RECOMBINATION is fancifully shown for two particles of a bacteriophage (a virus that infects bacteria) and their progeny. The particles attach themselves to the bacterium and inject their chromosome (*left*). The two chromosomes bear different variants of imaginary genes for coat color (*B or R*) and tail length (*L or S*). Within the host cell the chromosomes replicate and recombine. In offspring released from the bacterium, recombination has reassorted the genes to yield all possible combinations (*right*).

the free chain then establishes a new duplex with the complementary sequence on one chain of another chromosome. Further steps, to which those early in vitro observations offered no clue, would then resolve this initial structure into new, recombined chromosomes. Other investigators and I have been able to supplement this picture by concentrating not so much on *E. coli* as on a virus that attacks it—phage lambda.

Like other bacteriophages, or viruses that infect bacteria, phage lambda contains a single chromosome, which it injects into a bacterial cell to initiate the infection. The virus then subverts the biochemical machinery of the bacterium to reproduce itself—to copy its DNA and make the protein capsids, or shells, in which the new chromosomes



RECOMBINATION ENZYMES from the bacterium *Escherichia coli* manipulate DNA in ways that may promote recombination. In the test tube the *recA* protein coats single chains of DNA (*top left*). A coated chain then insinuates itself into double-chain DNA and explores the duplex until it finds a complementary sequence of nucleotides. The single chain and the complementary chain pair to form a new duplex, and the *recA* protein is shed. The new duplex and the newly displaced chain make up a "D loop." The electron micrograph, provided by Charles M. Radding of Yale University, shows a D loop formed by the *recA* protein (*bottom left*). In vitro the *recBCD* protein, *E. coli*'s other recombination enzyme, enters double-chain DNA at one end and travels between the chains, separating them as it advances (*top right*). The enzyme reassociates the chains behind itself but does so more slowly than it parts them; as a result growing loops of single-chain DNA develop on each side of the traveling enzyme. Such loops are visible in an electron micrograph made by Andrew Taylor in the laboratory of Gerald R. Smith at the Fred Hutchinson Cancer Research Center (*bottom right*). The capacity of the *recBCD* enzyme to generate single chains of DNA and that of the *recA* protein to insert them into other duplexes support the idea that these proteins catalyze early steps in recombination of the bacterial chromosome.

are housed after the multiplying viruses rupture their host. During their brief sojourn in the host cell the viral chromosomes recombine. If two genetically distinguishable strains of phage share a host, some of the viral progeny end up with genes from both parent strains.

Phages are attractive subjects for studies of recombination because of the small size of their single chromosome, the physical simplicity of the infectious phage particles and the rapidity with which they go about the business of life. Focusing on phage lambda infections in *E. coli,* moreover, has made it possible to examine the effects on lambda's simple chromosome of two recombination systems: *E. coli's* and that of the virus itself, which may co-opt parts of the bacterial system but also contributes several elements of its own.

Phage lambda's recombination enzymes, like those of *E. coli,* have been isolated and examined for catalytic activity in vitro. One of the proteins, the product of the *red-alpha* gene, digests one of the two chains of DNA duplexes, exposing the other chain. Its function in viral recombination would seem to be analogous to the postulated role of the *recBCD* protein: providing single-chain DNA that can then go hunting for a homologous segment of another chromosome. The *recA* protein made by the host bacterium could coat the single-chain viral DNA and promote the hunt. A second viral gene product, the *red-beta* protein, might foster the matching between the single-chain DNA and another chromosome. In vitro studies indicate that the protein speeds precise base pairing between complementary sequences of DNA.

The Life of a Phage

The exact choreography of recombination in the chromosomes of phage lambda cannot be reconstructed from the behavior of individual recombination enzymes in the test tube. Instead the process must be traced in a living system. The experimental manipulations that have enabled other workers and my own group to do so take advantage of the reproductive cycle of the phage.

After the virus injects its chromosome (a segment of DNA 48,500 nucleotide pairs long) into the *E. coli* cell, the two ends of the linear molecule join. The genes arranged around the circle of DNA then direct the synthesis of viral proteins, taking advantage of the bacterial host's own machinery for protein synthesis. Several of the protein products are enzymes that, working in concert with enzymes encoded by host genes, enable the lambda chromosome to replicate itself. At first the DNA replicates in what is known as the theta mode. Theta replication produces new circular chromosomes directly: the replication enzymes operating on each chain work their way around the parent DNA in opposite directions, generating circular daughter chains. The replicating chromosome, as it is conventionally diagrammed, resembles the Greek letter theta (θ).

After several rounds of theta replication some of the chromosomes begin to reproduce in a more efficient mode known as sigma, or rolling-circle, replication. The circle extends a tail, like the lowercase Greek letter sigma (σ); at the base of the tail an ensemble of replication enzymes operates continuously, duplicating both chains of the parent DNA to make a string of viral chromosomes. Sigma replication may begin when a nick develops in the theta-replication form, initiating the growth of the tail. The viral proteins encoded by a gene called *gam* and by the *red* genes also figure in the transition to sigma replication. The *gam* protein inactivates the host *recBCD* protein, one of whose DNA-cutting activities interferes with rolling-circle replication. How the *red* proteins aid the shift is not known.

By the time sigma replication begins, other genes have directed the synthesis of proteins that make up the viral capsids and of enzymes that fill the capsids with DNA. The major actor in the packaging of the viral chromosomes is the enzyme Terminase; it is Terminase that cuts up the strings of chromosomes produced during sigma replication. A sequence of nucleotides called *cos* recurs in the chromosome string at positions corresponding to the joined ends in the parent, circular chromosome. Terminase recognizes *cos* and binds to the DNA to the right of the sequence, as the lambda chromosome is conventionally represented. The enzyme then cuts the *cos* sequence to its left.

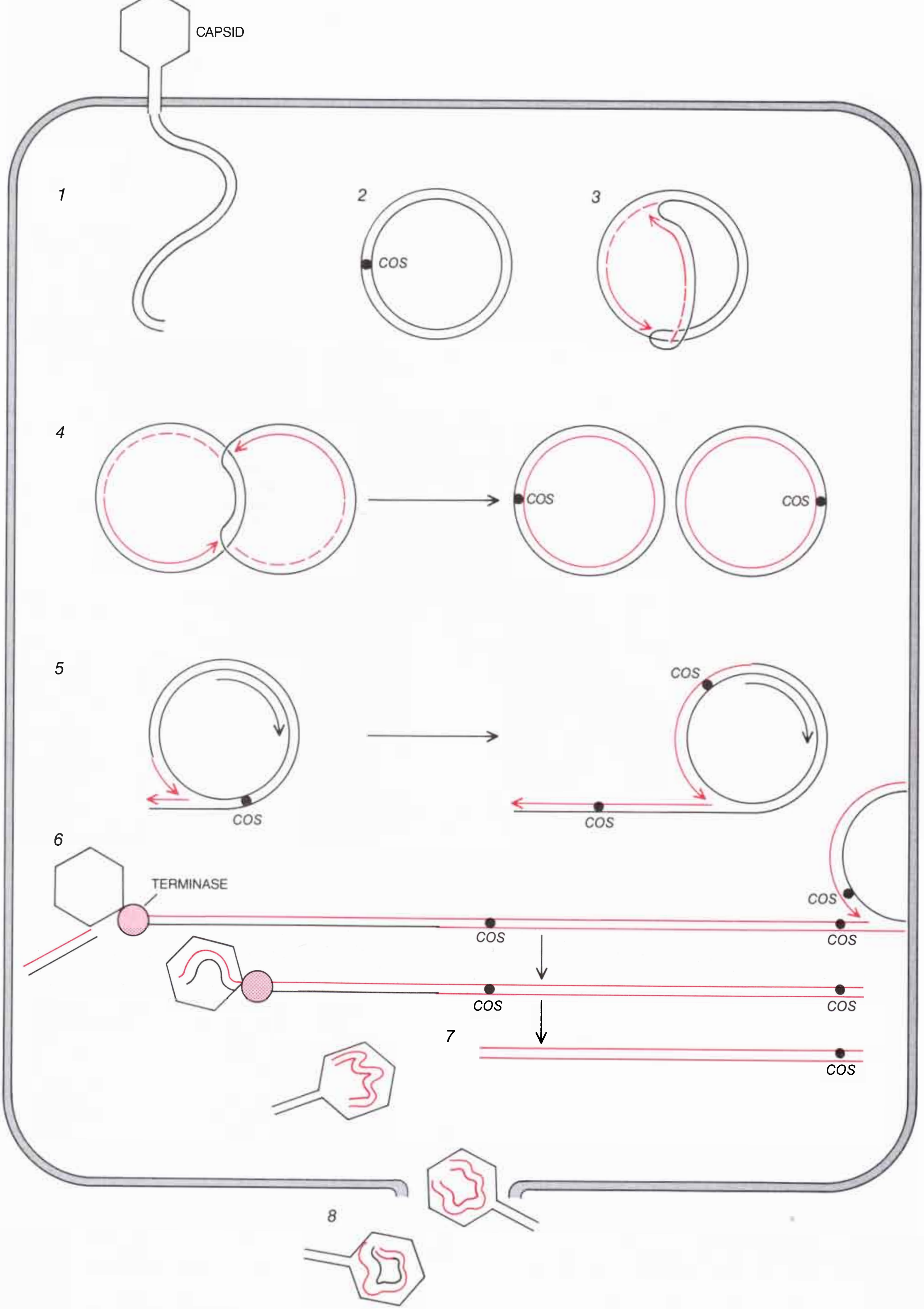
Without letting go of the cut end, Terminase binds to an empty phage capsid. It then moves to the right along the DNA, pulling the capsid with it and thereby filling it with DNA. When the advancing capsid-Terminase complex encounters a second, uncut *cos,* marking the end of one complete chromosome, Terminase makes a second cut and the packaging of the DNA into the capsid is completed. If the second *cos* has already been cut, the packaging will be aborted, even though most of the lambda chromosome has already been packed into the capsid. Only chromosomes that are embraced by a pair of uncut *cos* sequences are eligible for packaging.

Because the circular chromosomes created when the infecting DNA joins ends or made during theta replication contain only a single *cos* sequence, Terminase can cut the DNA but cannot package it. Through recombination such solitary chromosomes can become eligible for packaging, however. A simple example of recombination between circular chromosomes illustrates the principle. In linear chromosomes the breaking and swapping of DNA chains gives rise to two linear recombinants; the same kind of exchange between two circular chromosomes, however, produces a double-size circular molecule. In this dimeric (two-unit) molecule each of the new, recombinant chromosomes is now flanked by *cos* sequences. Either one of the chromosomes in the dimer (but not both of them) can now be packaged into a mature particle.

An Experimental System

The fraction of recombinant viruses in the particles released from an infected bacterium can be enhanced—and the analysis of the recombination systems of the host and the phage facilitated—if the viral chromosomes are not allowed to replicate. The object is to prevent them from undergoing sigma replication, which generates strings of chromosomes that can be

LIFE CYCLE OF PHAGE LAMBDA, a virus that attacks bacteria, is the setting for many studies of recombination. The virus particle first attaches itself to the membrane of its bacterial host, injects its chromosome and sheds its protein capsid (*1*). The chromosome's ends join to form a circle (*2*); at the joint a nucleotide sequence called *cos* is created. The DNA then begins to replicate itself, first in an inefficient mode called theta replication (for the shape of the structure formed partway through the process), which yields new circular chromosomes (*3, 4*). Later the replication shifts to the sigma mode (again for the resemblance of the structure to a Greek letter) (*5*). Sigma replication generates long strings of viral chromosomes. An enzyme called Terminase then packages individual chromosomes into new capsids (which have been synthesized with the help of the host). Terminase initiates the packaging of a viral chromosome when it cuts the string at a *cos* sequence (*6*). It then moves along the chromosome, filling a capsid with the DNA, until it reaches and cuts a second *cos* (*7*). Once a multitude of new virus particles have been assembled the host bursts and the virus disperses to start new infections (*8*).


CAPSID
1
2
COS
3
4
COS
COS
5
COS
COS
COS
6
TERMINASE
COS
COS
COS
COS
COS
7
COS
8

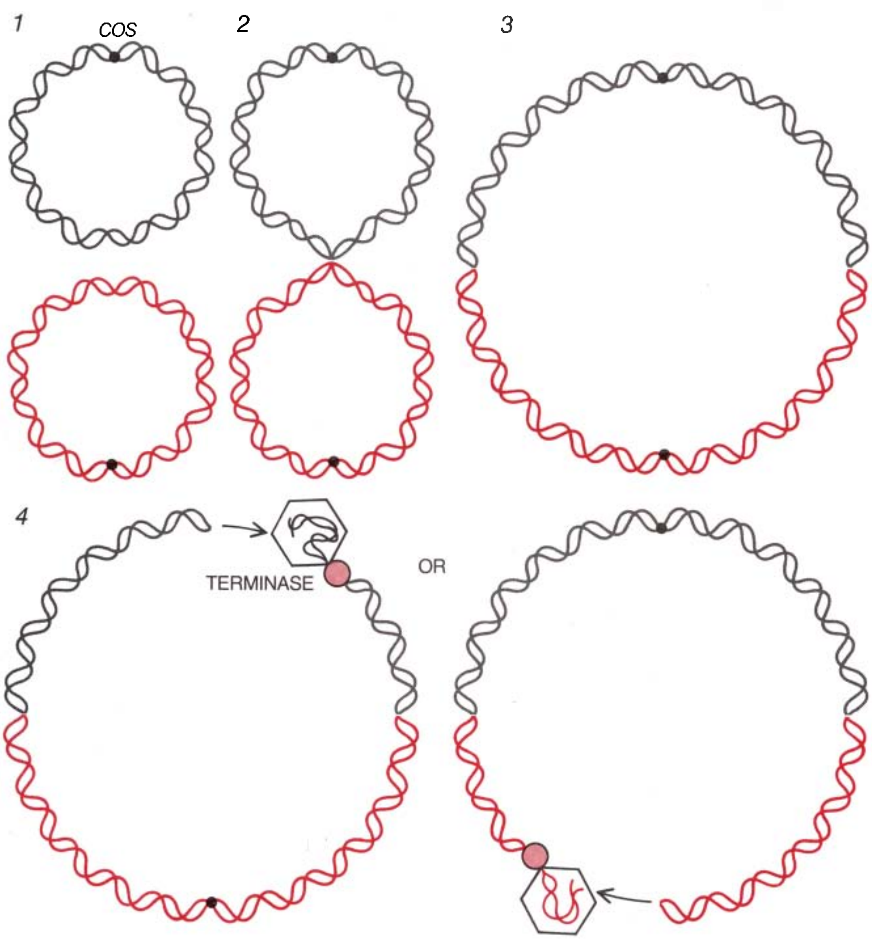
packaged directly. Because every step in the viral life cycle is carried out except for DNA replication, the circular chromosomes from the infecting virus can still be repackaged in new particles—but only if the chromosomes recombine.

Temperature-sensitive mutations in bacterial and viral genes that encode DNA-copying enzymes serve to block replication of the viral chromosomes. Such mutations do not manifest themselves at ordinary temperatures; otherwise the virus and its bacterial host could not be propagated. The experiments themselves can be done at higher temperatures, which activate the mutations and cause defects in the replication enzymes. The fact that the viral chromosomes do not replicate under those conditions can be verified by labeling the DNA of infecting particles and comparing it with the DNA of the particles that emerge from the host cell.

The standard technique for labeling phage DNA is to grow the virus in an environment rich in heavy, ordinarily rare isotopes of carbon and nitrogen. The phage particles incorporate the heavy isotopes into their DNA, thereby becoming denser than virus grown under ordinary conditions. By means of density-gradient centrifugation, a process that sorts virus particles according to density, the heavy isotopes can be detected in future generations of virus. In replication-blocked experiments the particles that emerge from bacteria infected with labeled phage have chromosomes made almost entirely of heavy isotopes, indicating that almost no new DNA was produced.

Kenneth D. McMilin, working in my laboratory, saw a way of combining density labeling with the replication-blocked system to study the recombination systems of the host bacterium and the virus. In the experimental system he devised, bacteria were infected with two kinds of lambda particles. One kind had been labeled with heavy isotopes and carried a genetic marker—an easily identifiable mutation—at the left end of the chromosome. The other kind contained ordinary, light isotopes and carried a mutation at the right end. The progeny from the genetically and isotopically mixed infection were sorted according to their density.

Among the viral progeny (virtually all of which must have undergone recombination) some of the particles were as dense as the labeled infecting particles and others were as light as the unlabeled ones. Those two groups presumably represented recombinations between two labeled and two unlabeled viruses respectively. Particles whose density fell between those extremes presumably had recombinant chromosomes containing some light and some heavy DNA. Coupled with the density information, the genetic markers on those chromosomes made it possible to determine where the recombination had been concentrated.



PACKAGING OF A CIRCULAR PHAGE CHROMOSOME into a capsid can take place only if the chromosome recombines. A circular chromosome contains only one *cos* sequence. Hence it offers only one DNA-cutting site to the enzyme Terminase, which must make both an initial and a final cut at intact *cos* sequences to package a chromosome. Here both recombining chromosomes are circular (*1*); the interchange (*2*) yields a double-size circular molecule (*3*). It contains two complete chromosomes flanked by *cos* sequences. Terminase can now encapsidate either chromosome (but not both of them) (*4*).

The Red System

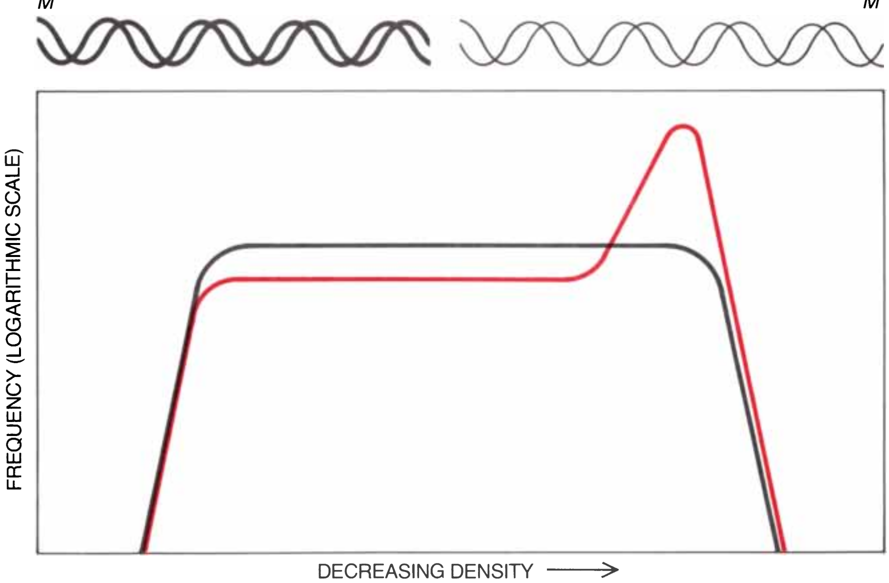
In the initial experiment a mutation in the *recA* gene of *E. coli* had inactivated the bacterial recombination system. Only the viral system, the Red system, could effect the recombinations needed to produce packageable chromosomes. The density of recombinant progeny that lacked both genetic markers, and therefore included the right part of a heavy chromosome and the left part of a light chromosome, was low. Those wild-type progeny evidently contained more light DNA than heavy DNA; hence recombination had been concentrated at the right end of the phage chromosome.

What feature of the chromosome end causes recombination to be concentrated there? The ends of the lambda chromosome are defined by *cos*, the nucleotide sequence at which Terminase can cut single circular chromosomes and the chromosome strings made during sigma replication. The techniques of gene cloning enabled Helios Murialdo at the University of Toronto and Ichizo Kobayashi in my laboratory to shift the nucleotides making up *cos* from the chromosome ends to the middle of the chromosome. In bacteria infected with the altered phage, Terminase presumably could cut the DNA only at the new site of *cos*, so that as the chromosomes were packaged the former left and right ends became the new middle. A genetic and isotopic analysis of progeny from the infection showed that recombination was now concentrated at the new right end of the viral chromosome; that is, at the new site of *cos*.

Thus *cos* is a "recombinator": a sequence of nucleotides that promotes genetic recombination in its neighborhood. The fact that *cos* stimulates recombination only at the right end of a chromosome (and not at the left end as well) suggested its mechanism of action. Kobayashi proposed that *cos* initiates an act of Red-catalyzed recombination by virtue of being the site where Terminase cuts a circular chromosome. Because Terminase remains bound to the left end of the chromosome, only the right end is free to engage another chromosome. The Red system acts on the free end.

One would therefore expect that the Red system would promote recombination wherever the two DNA chains of the lambda chromosome could be cut. Restriction enzymes, DNA-cutting compounds that are elaborated by bacteria as a weapon against invading viruses (and have become a familiar tool of laboratory genetic manipulations), made it possible to test that prediction. Each restriction enzyme recognizes and cleaves a specific short nucleotide sequence—and hence cleaves any chromosome containing it—unless that sequence has been chemically modified by another enzyme, which the bacterium produces in order to protect its own genetic material.

David Thaler, working in my laboratory, studied recombination in bacteria that made a particular restriction enzyme. One set of infecting virus bore an unmodified target sequence whose position on the phage chromosome was known. The other infecting particles also carried the sequence, but in modified and therefore uncuttable form. The modification ensured that some complete viral chromosomes, capable of being packaged into new phage particles, would emerge from the host. The host's recombination system was blocked, leaving only the Red system active. Thaler observed extensive recombination, and it was concentrated in the chromosome region bearing the cuttable sequence. Evidently the cut chromosome ends had sought out and recombined with modified, uncut chromosomes.



DENSITY DISTRIBUTION of phage particles released from infected bacteria yielded clues to the site of recombination on the viral chromosome. Two sets of virus particles initiated each infection. The DNA of one set carried a label that made it unusually dense; it also bore a mutation (*M*) at the left end of the chromosome (*top left*). The other set of particles contained ordinary DNA with a mutation at the right end of the chromosome (*top right*). The infections were conducted under conditions that blocked DNA replication but allowed recombinant chromosomes to be packaged and released. When the bacterial recombination system was inactivated, leaving only the recombination system of the virus, phage particles whose chromosomes bore neither mutation tended to be light (*colored curve*). The absence of mutations indicated that the DNA consisted of the right side of a dense chromosome and the left side of a light chromosome, and the low density suggested that the segment of dense DNA was short. Recombination must have been concentrated at the right end of the chromosome. When the bacterial recombination system was the active one, phage particles carrying neither mutation had a full range of densities (*black curve*). Recombination must have been evenly distributed along the viral chromosome.

The Role of Replication

How does the Red system behave during the normal reproductive cycle of phage lambda, when its DNA is replicating freely? Addressing this question by means of density labeling is difficult: unbridled DNA replication would result in so much new, light DNA that the density label of the infecting particles would be overwhelmed. When all the infecting particles contain heavy isotopes and only a small amount of replication is allowed, however, the experiment yields clear-cut results.

My colleagues and I infected cells with two types of density-labeled particles bearing contrasting sets of genetic markers. As before, the markers made it possible to determine where recombination had taken place in the chromosomes of the progeny. When the chromosomes consisted entirely of dense (and therefore unreplicated) DNA, recombination was concentrated at the right end, as it had been in replication-blocked trials. When they consisted partly or entirely of light (newly synthesized) DNA, recombination had taken place throughout the chromosome.

Presumably the Red system was acting on free ends exposed at random in the DNA sequence during replication. That notion gains some support from recent studies of Red-mediated recombination in the presence of a protein that binds to and occupies free ends of duplex DNA: in freely replicating virus the protein inhibits recombination throughout the chromosome. The simplest way to account for the randomly positioned free ends this result seems to require is to attribute them to the onset of sigma replication. The nick that initiates sigma replication is thought to occur at random on the circular chromosome, and so the tail that then starts to grow can begin anywhere in the DNA sequence.

The association of Red-mediated recombination with chromosome ends, together with the in vitro behavior of Red recombination enzymes, yields a clear picture of the initial steps in the process [*see illustration on next page*]. The protein product of the *red-alpha* gene partly digests one of a chromosome's two chains, beginning at the free end. The digestion exposes single-chain DNA, which invades a second, circular chromosome, establishing a D loop where it finds the homologous sequence. The phage's *red-beta* protein and the *recA* protein of the bacterial host may facilitate the invasion. Resolving the structure created by the single-strand invasion into a recombinant chromosome calls for further enzymatic steps, which are not completely understood.

The Rec System

My colleagues and I applied the same tools for assessing the site of recombination to the Rec system of *E. coli*. Bacteria that had an intact recombination system were infected with labeled and unlabeled phage that carried mutations in the *red* and *gam* genes, incapacitating the virus's own recombination system and making the virus unable to block the host system. The effect on the phage chromosome was intriguingly different from that seen when the Red system was the active one; in replication-blocked experiments recombination occurred with almost equal frequency all along the chromosome. It did not cluster near the *cos* sequence, and the pattern and frequency did not change when some DNA replication was allowed. Evidently the Rec system, unlike the Red system, has no special affinity for cut chromosome ends.

Another line of research unexpectedly shed light on the Rec system. David Henderson of the University of Edinburgh was studying lambda mutants lacking the *red* and *gam* genes. Because those genes are essential not only for recombination but also for sigma replication, such mutants can carry out only theta replication. He noted (as other workers had) that the mutant virus could not grow on bacteria carrying a mutation in their *recA* gene. In the absence of both the bacterial recombination system and the viral one, none of the original chromosomes or the circular progeny made by theta replication could be packaged and released as a new particle.

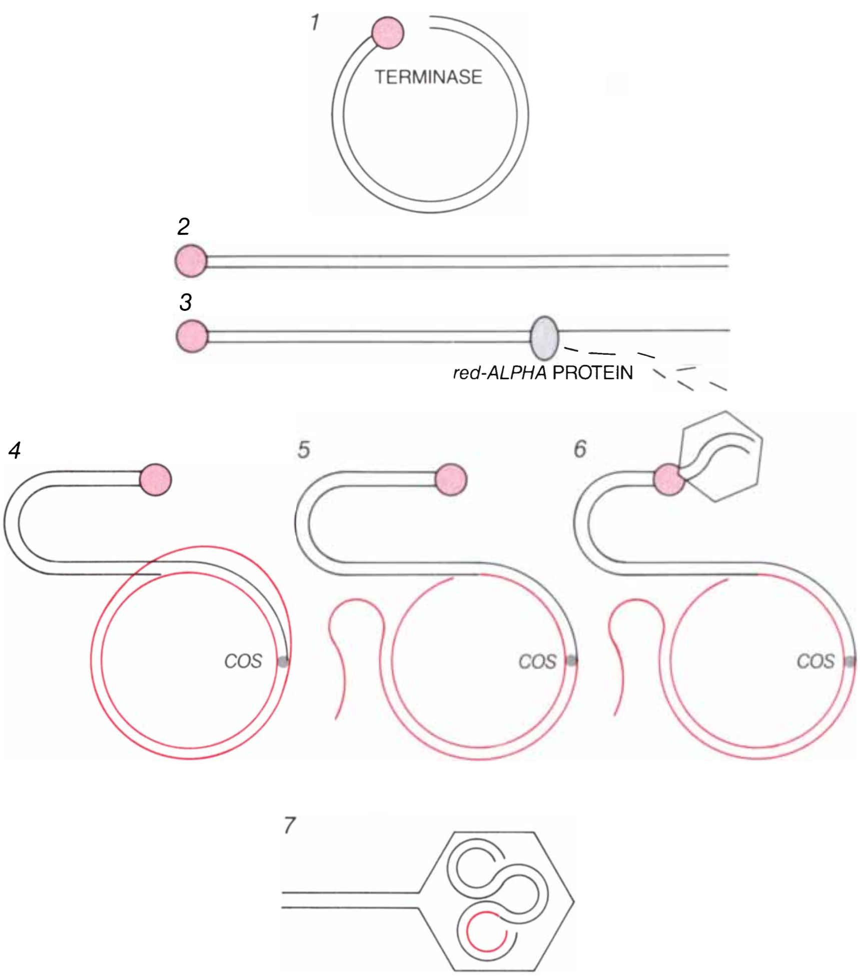
To his surprise, Henderson found that the mutant grew poorly even on bacteria with an intact Rec system. Although a functional Rec system should have made the circular chromosomes packageable by recombining them, few mature particles were produced. Apparently the Rec system worked inefficiently on the viral chromosome.

A secondary mutant arose spontaneously in Henderson's phage cultures and very soon outgrew the original, slow-growing mutant. The new phage was able to grow so well not because it had a new trick for becoming eligible for packaging but because the Rec system acted more efficiently on its DNA. The responsible mutation created a sequence of eight nucleotides that is now known as Chi.

My colleagues and I applied McMilin's method for determining the distribution of recombination sites to Chi-positive strains. We found that each of those mutants had an extraordinary rate of recombination near Chi. A mutation in the lambda chromosome evidently had led us to a sequence that acts as a recombinator in the Rec system of the bacterial host, just as *cos* does in the Red system.

How Chi achieves its effect may soon become clear. Any model of Chi-stimulated recombination must make sense of the many features of Chi that have been established since its discovery. Chi is known, for example, to play a crucial role in the Rec-catalyzed recombination of *E. coli*'s own chromosome, which is littered with Chi sequences. It has no effect on other systems, such as lambda's Red system. One would therefore suppose it interacts directly with one of the Rec recombination enzymes.

It also seems to exert its effects at some distance. Chi has to be present on only one of the two interacting chromosomes to stimulate recombination. On the other chromosome the corresponding site and thousands of nucleotides on each side of it can be missing, yet Chi-stimulated recombination takes place where the homology between the two chromosomes resumes. It is significant that such distant effects were at first seen only on



RECOMBINATION SYSTEM OF PHAGE LAMBDA is thought to act on a free end of double-chain DNA. Here the enzyme Terminase generates the free end by cutting a circular phage chromosome at the *cos* sequence (*1*); the enzyme remains bound to the left end (*2*). The protein product of the phage's *red-alpha* gene partly digests one chain of the chromosome, beginning at the free right end (*3*). The exposed single chain invades another chromosome, forming a new duplex and creating a D loop (*4*). Enzymes (not all of them known) cut and rejoin the chains in the D loop so that the initiating chromosome and the new duplex are continuous (*5*). Terminase then moves along the DNA, packaging it into an empty capsid (*6*). At the *cos* sequence of the invaded chromosome it cuts the DNA, yielding a new phage particle with a recombinant chromosome (*7*). This model accommodates the observed behavior of the *red-alpha* protein and the finding that genetic recombination catalyzed by the phage system is concentrated near double-chain ends.

one side of Chi: to the left of the sequence (as the lambda chromosome is conventionally represented).

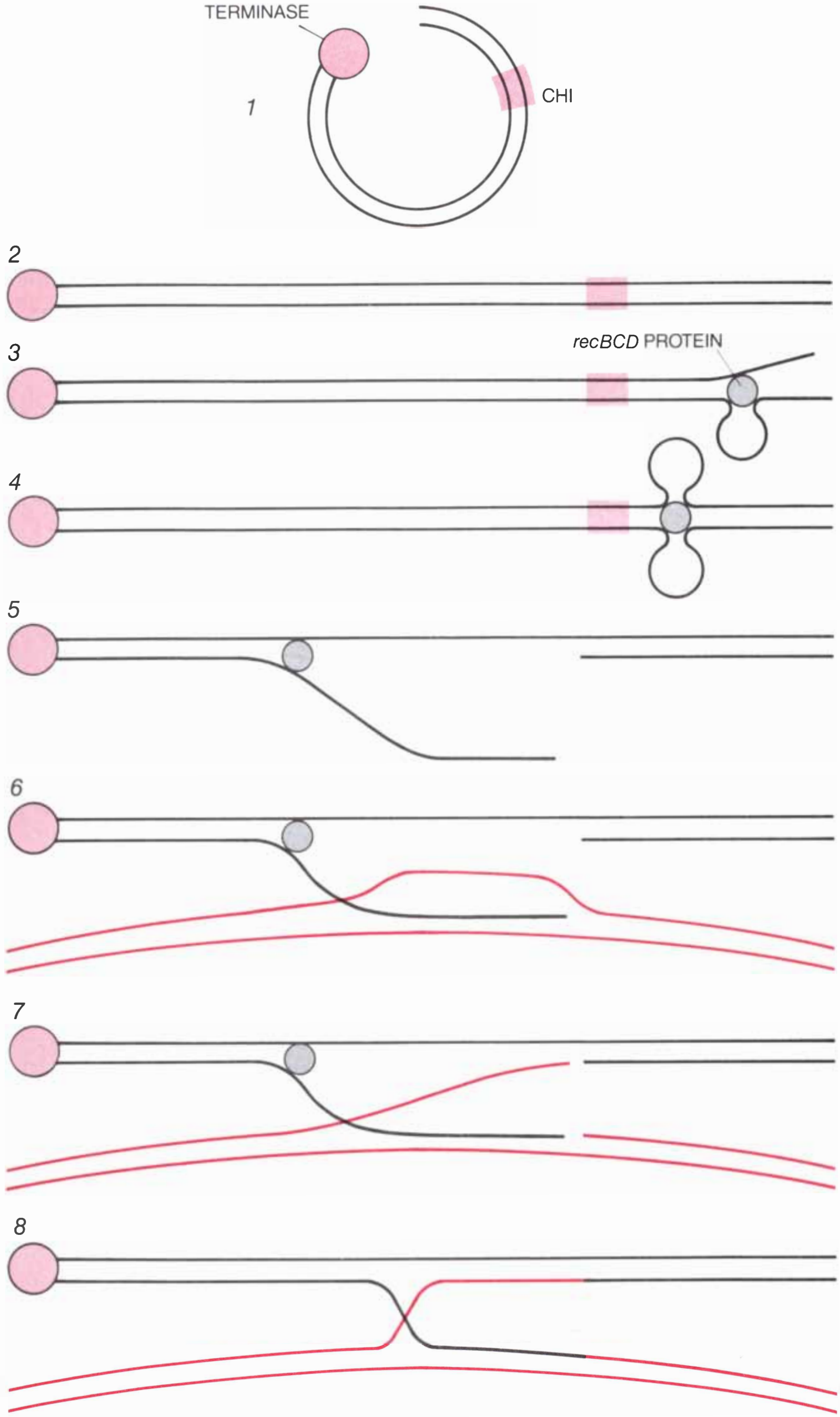
Chi's effects are not only directional but also related to its orientation in phage lambda's chromosome. When, by means of genetic-engineering techniques, the Chi sequence is inverted in the chromosome, it no longer functions. Strikingly, however, if an inverted *cos* sequence is introduced into a lambda chromosome bearing a silent, inverted Chi, the Chi sequence becomes active again. The relevant feature of *cos* turned out to be its role as a site for double-chain cutting by Terminase. When phage carrying a silent, inverted Chi infected a host that produced a restriction enzyme capable of cutting the lambda chromosome, Chi was again activated. Once it is activated by double-chain cutting, an inverted Chi sequence stimulates recombination to its right.

The best way to account for those varied observations is to propose that Chi interacts with a kind of traveling recombination machine. The machine—one enzyme or several—must bind to a free end of the chromosome and then move in one direction along the DNA until it encounters a Chi sequence in the correct orientation. The conventional orientation is the active one when the recombination machine begins its travels at the right end of the chromosome, which is the free end ordinarily generated when Terminase cuts a *cos* sequence. An inverted Chi will be active only if the recombination machine is traveling in the other direction, that is, from the left end of the chromosome. Terminase generates a free left end only if *cos,* which marks the site at which the enzyme cuts a circular chromosome, is itself inverted.

Once the traveling recombination machine has encountered a properly oriented Chi, the machine is much more likely to act. In keeping with the observation that Chi can act at a distance, the recombination event can occur after the machine has traveled beyond Chi: to the left of Chi if the sequence is oriented conventionally and to its right if the sequence is inverted.

The Holliday Intermediate

The traveling recombination machine of the Rec system has been not only described but also tentatively identified. Its properties recall the in vitro behavior of the *recBCD* enzyme. That enzyme also engages a free end of duplex DNA, travels along the duplex and generates single chains of DNA on the way. Furthermore, mu-



RECOMBINATION SYSTEM OF *E. COLI* is believed to depend on the interaction of the bacterium's *recBCD* protein and a DNA sequence known as Chi. The illustration shows a model of the interaction as it might take place in the chromosome of phage lambda, which infects *E. coli*. After Terminase cuts open a circular phage chromosome (*1*) to yield linear DNA (*2*), the *recBCD* enzyme enters the DNA duplex and travels along the molecule (*3*), spinning out loops of single-chain DNA until it reaches Chi (*4*). The encounter changes the enzyme's character, and it soon cuts one of the loops (*5*). With the help of the *recA* protein (not shown), the free chain invades another phage chromosome, forming a D loop (*6*). Further enzymatic steps cut the displaced chain in the D loop, freeing it to pair with the initiating chromosome (*7*). Newly juxtaposed chains join ends (*8*), forming a Holliday intermediate: a crossed-chain structure that can later be resolved into two recombinant chromosomes. Recent findings have undermined this model.

tations of *recBCD* have been identified that leave the Rec system active but unresponsive to the Chi sequence. They suggest a direct interaction between the *recBCD* protein and Chi.

Just how does the protein act as the recombination machine, and how does Chi stimulate it? Those questions can be posed more concretely in terms of the intermediate structure that is generated when recombination takes place away from a double-chain end, as it does in Rec-catalyzed recombination. In 1964 Robin Holliday, then at the John Innes Institute in Hertford, England, proposed that at some point in the course of recombination the process of DNA swapping will have affected only one chain in each chromosome. The chain will have broken and joined ends with its continuation on the other chromosome, forming a "crossed-strand structure," or Holliday intermediate.

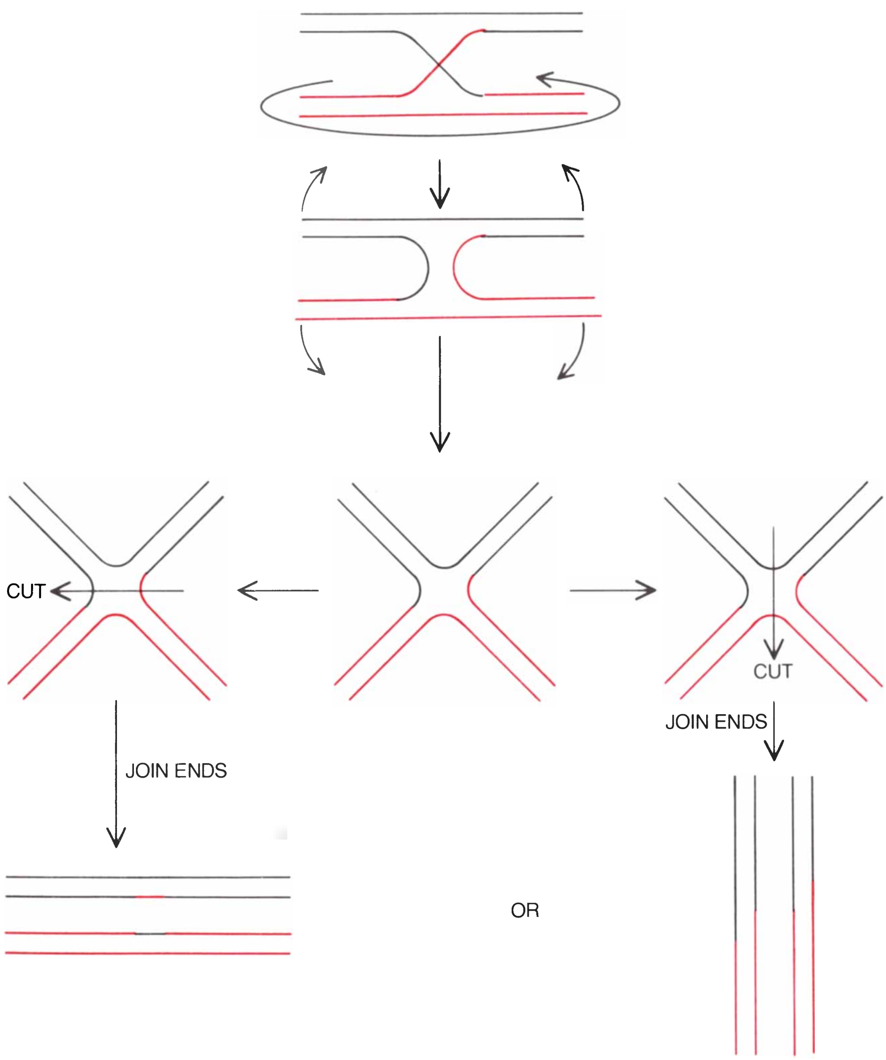
Resolving the intermediate into separate chromosomes calls for another event of chain breaking and rejoining. Because of the geometry of the intermediate structure [*see illustration below*], that event can take place either on the chains that were originally exchanged or on the complementary chains. In the former case the resolution yields "patch" recombinants: recombinant chromosomes that are essentially unchanged, having swapped only a short patch of single-chain DNA. In the latter case the complementary chains are also swapped, completing the exchange of chromosome segments and forming "crossover" recombinants.

The question thus becomes: Does the *recBCD* protein, activated by Chi, catalyze the formation of Holliday intermediates or their resolution? One observation, made by Gerald R. Smith and his colleagues at the Fred Hutchinson Cancer Center in Seattle, supports a model based on the first possibility. The Seattle group found that under certain conditions the protein, traveling along a DNA molecule in vitro, responds to a Chi sequence encountered in the proper orientation by cutting a particular chain: the one that during the synthesis of viral proteins is transcribed, or read, from right to left.

The model assumes the cut interrupts the reassociation of DNA that usually takes place behind the traveling enzyme. The single chain that now has a free end is spun out from the duplex. The chain becomes coated with *recA* protein and invades the homologous chromosome, establishing a D loop. The chain that is displaced in the D loop crosses to the original chromosome, and a step of chain cutting and rejoining gives rise to a Holliday intermediate [*see illustration on preceding page*]. Its resolution can yield either patch recombinants or crossover recombinants. This model for the action of Chi and the *recBCD* enzyme is so attractive that the latest edition of a standard text offers it as fact. A recent finding undermines it, however.

Susan Rosenberg, working in my laboratory, tested a specific prediction of the model. The model assumes that the same chromosome chain cut by the *recBCD* protein in vitro—the one that is read from right to left, or, for short, the leftward chain—is also cut in vivo and serves to initiate recombination. If a cut leftward chain invades a homologous chromosome and pairs with a complementary DNA sequence, it will displace that chromosome's leftward chain. Both crossed chains in the Holliday intermediate will therefore be leftward chains. Hence the model predicts that if the intermediate is resolved into patch recombinants, the patches will appear on the leftward chains of the chromosomes.

Rosenberg found just the opposite. Each DNA patch turned up on a rightward chain—the chain that remains uncut in vitro. That finding rules out the simple model based on the in vitro observation. Another version of the model also assumes that a cut leftward chain initiates recombination but describes a more complex scenario in



RESOLUTION OF A CROSSED-CHAIN STRUCTURE can have divergent results, according to the model proposed by Robin Holliday in 1964. The structure, formed early in some recombination processes (*see illustration on preceding page*), can also be represented as a "crossroads" configuration, in which four segments of double-chain DNA intersect at right angles. Cutting the intersection and joining the two single-chain ends on each side of the cut can produce patch recombinants: chromosomes that have exchanged no more than a short length of single-chain DNA (*left*). An alternative cutting and joining step can result in crossover recombinants, that is, full-fledged recombinant chromosomes (*right*).

which patches end up on lambda's rightward chain. The second model makes its own special predictions, one of which Rosenberg tested and also found to be false.

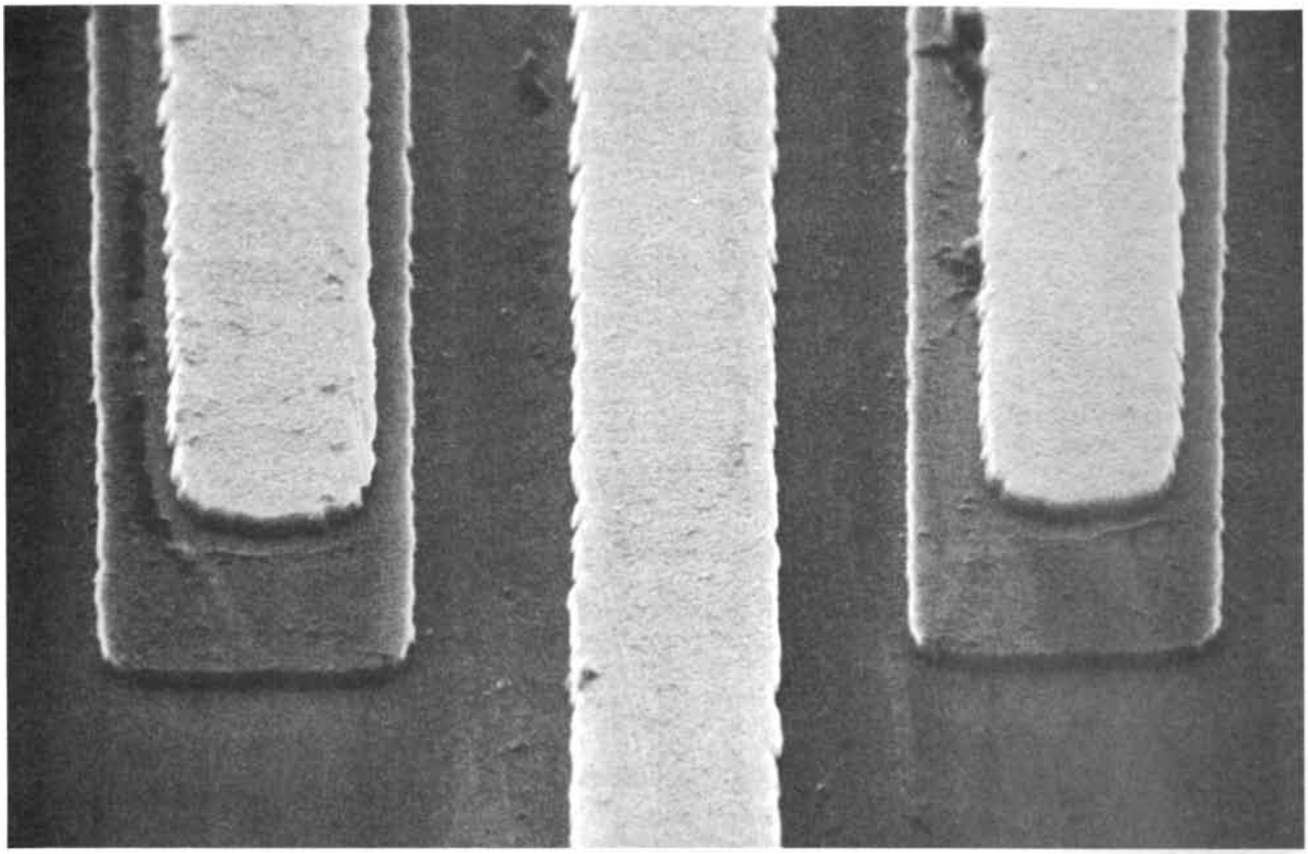
The demise of both models suggests that the chain cutting stimulated by Chi in vitro, even if it also occurs in vivo, might not serve to initiate recombination. Thus the initial question of whether Chi stimulates the formation or the resolution of Holliday intermediates has returned. Efforts to answer it and characterize the underlying interaction between Chi and the *recBCD* protein are spurring new research.

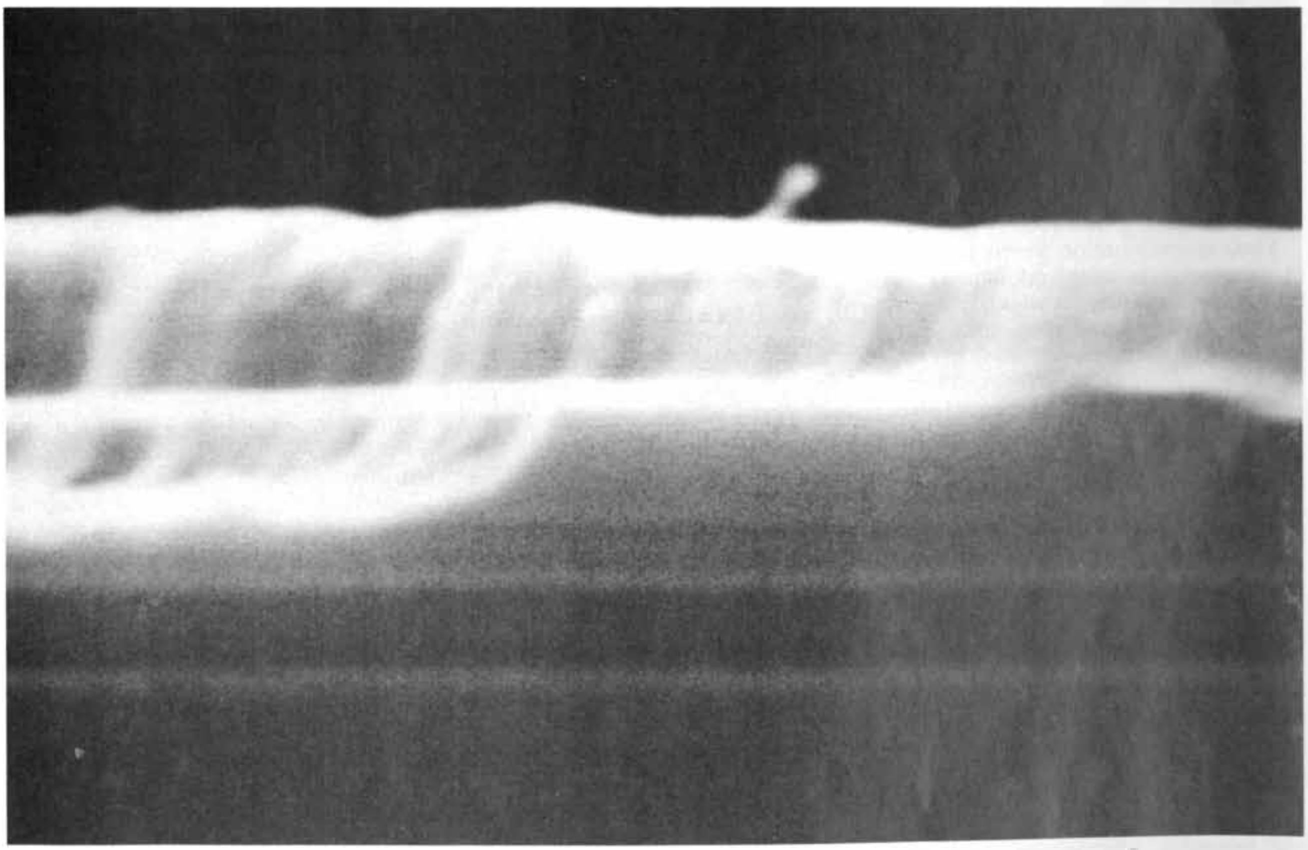
Many Styles of Recombination

Although details remain to be resolved, the overall picture is clear: recombination in phage lambda and *E. coli* follows two very different molecular pathways. In the virus's Red system the free end of the double-chain molecule is the active site; it is there recombination enzymes set to work to promote recombination. The system suits lambda's life-style, because the free ends arise at random in the chromosome sequence during sigma replication. In the bacterium, in contrast, the chromosome ends are fixed; to enable recombination to take place throughout its genetic complement it needs a different system. In the bacterium's Rec system a free chromosome end serves merely as the loading site for a recombination machine that catalyzes recombination at an arbitrary distance from the end.

Thus recombination seems to vary in its precise mechanisms depending on the genetic idiosyncrasies of the organism. Even a single organism may employ a variety of pathways. *E. coli,* for example, has an auxiliary system known as RecF, and certain fungi also seem to recombine their genetic material in several ways. Studies of the molecular underpinnings of recombination in more complex organisms may reveal further variants.

Such studies gain urgency from the prospect that natural recombination mechanisms might one day be harnessed to replace genes in multicellular organisms—a capacity that could be a boon to agriculture and to the treatment of human genetic disease. Inserting an altered gene in its proper place on a chromosome calls for a precision far beyond that of today's genetic-engineering techniques. The precise and beautiful mechanisms every organism employs to recombine its genetic endowment could well become the tools of a future biotechnology.





BALLISTIC-ELECTRON DEVICE, in which electrons can flow across a thin channel without being scattered by collisions, is seen from above (*top*) and from the side (*bottom*). The device consists of a number of thin layers of semiconductor, which represent such elements as the electron emitter (near the top of the device), the channel through which electrons flow and the region where electrons are collected. The device is magnified about 10,000 times in the top image and about 100,000 times in the bottom image. The images were made by Michael Rosenfield, Carmello Aliotta and Gordon Wilson of the IBM Thomas J. Watson Research Center.

Ballistic Electrons in Semiconductors

Devices in which electrons carry current without being scattered promise to be much faster than present-day components. They also allow close study of the electron's quantum-mechanical properties

by Mordehai Heiblum and Lester F. Eastman

In a television picture tube electrons move from the cathode to the picture screen ballistically: their motion follows smooth trajectories that are not interrupted by interactions or collisions with other particles. Because there are very few gas molecules in the tube, gas molecules do not scatter many electrons. In semiconducting devices such as transistors, on the other hand, the motion of the current-carrying electrons is constantly interrupted by scattering. Each electron is scattered frequently and in every direction, so that its motion resembles the random motion of a molecule in an ordinary gas or liquid. The average electron covers a distance in random motion much greater than the net distance it travels. The net flow of electrons in the direction of the electric field is consequently much slower than the actual speed of the electrons. If it were somehow possible to enable electrons in semiconductors to move without scattering—to move ballistically—then the speed of transistors could in principle be increased dramatically.

Such ballistic motion has now been achieved. Our work at the IBM Thomas J. Watson Research Center in Yorktown Heights and the work of investigators at other institutions has shown the conditions under which electrons can be made to move ballistically and has recently led to the first unambiguous demonstrations of ballistic motion in semiconducting devices.

Our interest in ballistic electrons is technological and scientific. Technologically, ballistic electrons might become the basis for a generation of electronic devices that are at least an order of magnitude faster than the devices available today. Scientifically, ballistic-electron devices make possible a new level of investigation into the quantum-mechanical properties of coherent electrons in semiconductors, because ballistic electrons in semiconductors, like ballistic electrons in a television tube, can be studied as freely moving quantum-mechanical wave packets; they exhibit wavelike effects, such as interference, that are usually observed in electromagnetic waves.

Electrons can move ballistically in a vacuum tube because the distance they are made to travel within the tube is shorter than their mean free path (the distance the average electron travels between collisions with the few gas molecules in the tube). In semiconductors it is impossible to eliminate scattering in the same simple way. Surprisingly, the density of the solid medium is not in itself an impediment to ballistic motion, although the atoms of a semiconductor are 10,000 times closer together than the gas molecules in a vacuum tube. The densely packed atoms in a semiconductor need not scatter electrons, because the atoms of a crystalline solid are arranged in a periodic lattice: they are spaced at equal distances in a repeating pattern. According to quantum mechanics, an electron can mathematically be treated in some ways as a wave; it is a consequence of this wavelike behavior that an electron is not scattered when it interacts with a perfect crystalline array of atoms arranged periodically. This surprising result, proved by Felix Bloch in 1928, is a basic concept of solid-state physics. (The electron does not, however, behave like a free electron in a vacuum but like a lighter particle with an "effective mass" that is smaller than the mass of the free electron and with a certain maximum velocity; the effective mass and the maximum velocity are dictated by the crystalline medium through which the electron moves.)

In any real crystal at any temperature above absolute zero, the perfect periodic order is only an approximation of the actual arrangement of atoms. For one thing, the atoms of the solid are always in thermal motion (the random vibrational motion due to heat energy provided by the surroundings). Thermal vibrations cause waves of compression and expansion to move through the crystal, and electrons can be scattered by these compression waves, since the waves destroy the perfect periodic arrangement of atoms. When an electron is scattered by a lattice vibration, it acquires energy from or loses energy to the crystal in fixed amounts, or quanta. Such a scattering event is called an inelastic collision.

Electrons in semiconductors can be scattered not only by lattice vibrations but also by the impurity ions that always exist in real materials. The impurity ions are distributed randomly rather than in a periodic arrangement, and so they scatter the electron waves. Collisions with ions do not change the electrons' energy much (because the ions are much more massive than the electrons), but they do change the electrons' direction of motion. Such collisions are called elastic collisions. Electrons are also scattered in many other ways, such as by inelastic collisions with other electrons and by faults and defects in the structure of the crystalline lattice. All these collisions shorten the mean free path of electrons in semiconductors.

In existing semiconducting devices the current of electrons flows over distances many times greater than the electrons' mean free path. To enable electrons to move ballistically, either the mean free path must be increased (by choosing the appropriate semiconducting materials and manufacturing them at very high purity) or the region through which current flows must be decreased considerably (by relying on sophisticated fabrication techniques).

The complex behavior of electrons moving along short paths is illustrated in a typical semiconducting device: the field-effect transistor. In a field-effect transistor a voltage difference causes electrons to flow from a region called the source through a thin region called the channel to a third region called the drain [*see illustration on opposite page*]. Near the channel is a metal contact called the gate. When a negative voltage is applied to the gate, it causes an electric field that repels electrons from part of the channel, creating what is effectively a narrower conductive channel with a higher resistance and thereby making it harder for current to flow from the source to the drain. The transistor thus can act as an amplifier: a small change in the gate voltage can effect a large change in the amount of current flowing through the device.

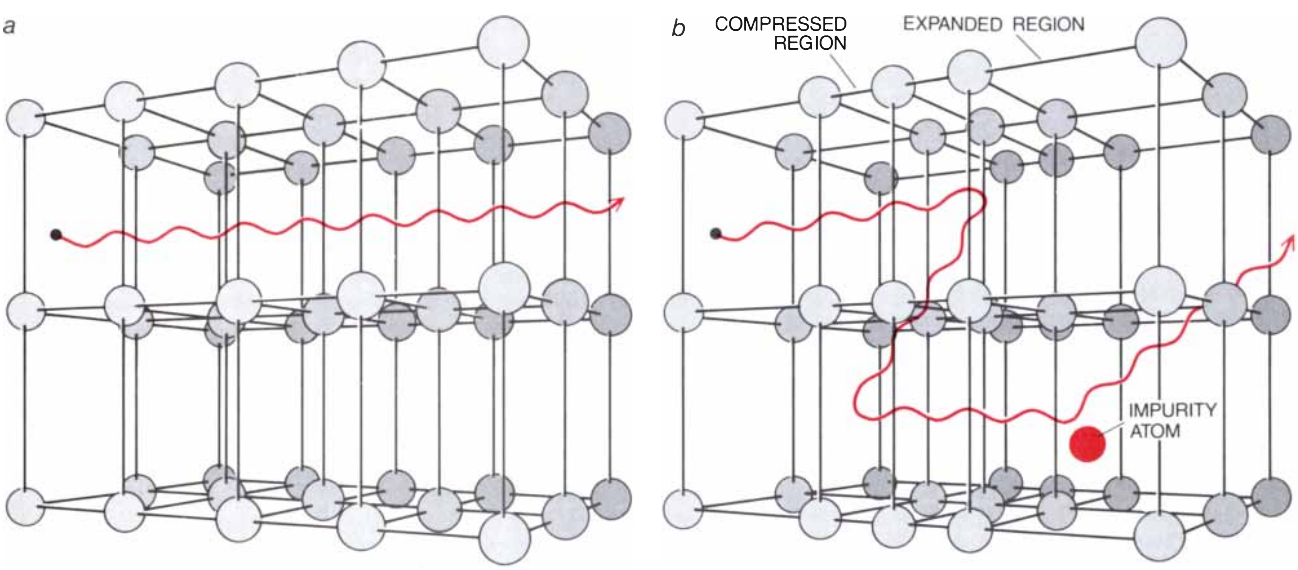
In terms of speed, the most significant dimension of a field-effect transistor is the length of the gate metal. The speed with which electrons traverse the channel that is under the gate and between the source and the drain determines the amount of time necessary for the device to respond to changes in the gate voltage. In a typical commercial field-effect transistor the channel is about a micrometer long, which is from 10 to 100 times the mean free path of the electrons (depending on the semiconducting material the device is made of), and so the electrons cannot move ballistically.

Precisely what happens to electrons as they move nonballistically from the source to the drain? Electrons leaving the source are at first accelerated by the voltage difference between the source and the drain. At a characteristic distance from the source (roughly the distance of the mean free path) the average velocity of the electrons reaches a maximum. At greater distances from the source the average velocity slowly declines to a stable level. This phenomenon is known as velocity overshoot. The decline in velocity is caused largely by inelastic collisions with lattice vibrations.

If the electrons could travel ballistically through the crystal, they would initially accelerate in a way similar to that of an electron under the influence of an electric field in free space (although the electrons in the semiconductor have a smaller effective mass): their velocity would be roughly proportional to the square root of the distance they had traversed. Even without collisions, however, the electrons would not accelerate forever. Because of the nature of the interaction between the electrons and the periodic lattice, they would eventually reach some maximum velocity. The maximum velocity is determined by the crystal structure and atomic makeup of the semiconductor, as well as by the direction in which the electron travels in relation to the crystalline lattice.

The average speed of the electron as it moves from the source to the drain is somewhere between the velocity with which it leaves the source and the maximum velocity it can achieve in the semiconductor. Clearly the electron would have a much higher average velocity (and the device could operate much faster) if the electron could be "launched" into the conducting channel at its maximum possible velocity. In the ballistic devices we have built we have managed to achieve just such an effect. In order to understand how the launching is done, and to understand the makeup of the ballistic devices themselves, it is first necessary to understand some of the properties of semiconducting materials.

The commonest semiconductor is silicon. Every atom in a crystal of silicon has four valence electrons, or electrons in the outermost shell (these take part in chemical reactions), which it shares with four neighboring atoms. The electrons act as a glue, bonding the crystal together. In certain so-called compound semiconductors, such as gallium arsenide, one atom (gallium) contributes only three valence electrons and the other (arsenic) contributes five, so that the average number of bonding electrons per atom is still four. The electrons cannot carry current when they are strongly localized between two nuclei; in order to be able to move freely about the crystal an electron must be given a certain amount of energy, which converts it



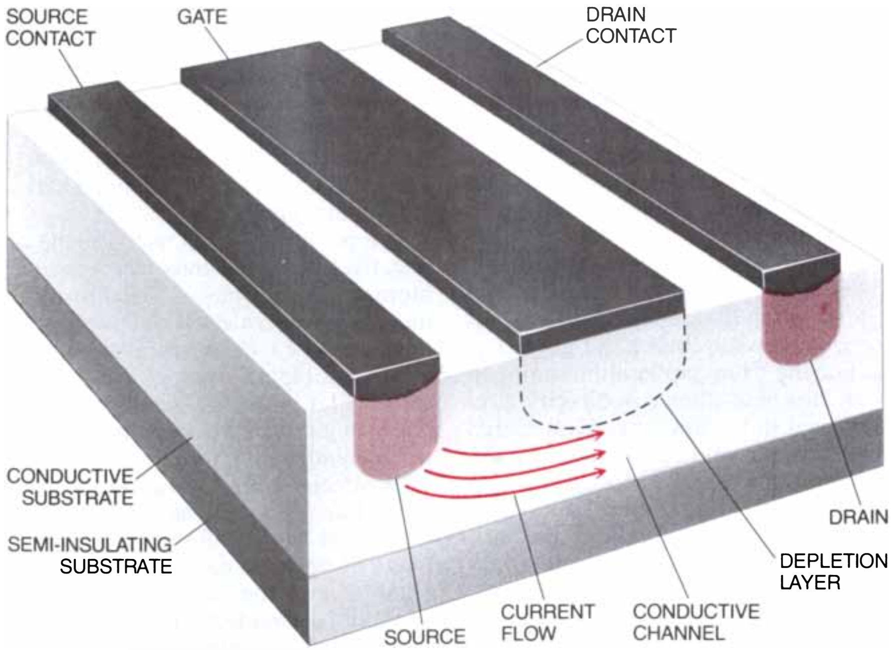
SCATTERING OF ELECTRONS occurs in semiconductors because of thermal vibrations (vibrations caused by heat) of the semiconductor's crystalline lattice and impurity atoms present in any real sample of semiconductor. The atoms of an ideal crystal (*a*) are arranged in a strictly periodic array. They are positively charged because they have given up electrons to bond the crystal together, and so they create an electric potential that varies periodically in space. According to quantum mechanics, electrons have certain wavelike characteristics. The wavelike electrons interact with the periodic potential in such a way that they pass through the crystal without being scattered at all. In a real crystal thermal vibrations cause waves of compression and expansion that disturb the periodic potential (*b*). Moving regions of irregular potential scatter electrons, as do impurity atoms buried within the crystal.

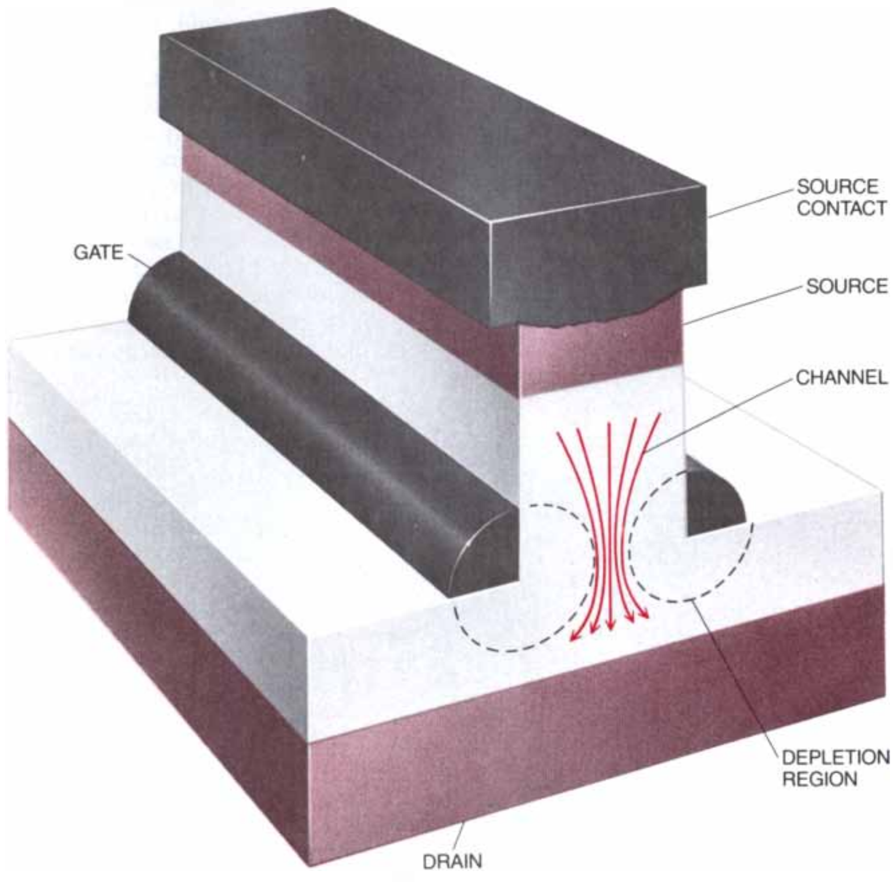
from a bonding (valence) electron into a conduction electron. The conduction electrons therefore have higher electric potential energy than the bonding electrons, just as a satellite high above the earth has higher gravitational potential energy than a satellite about to be launched. The difference between the minimum potential energy of the conduction electrons in a semiconductor and the maximum potential energy of the valence electrons is called the band gap, and its magnitude differs from one semiconductor to another.

In most of the commonly used semiconductors at room temperature few electrons have enough thermal energy to become conduction electrons (they would need to have a thermal energy greater than the energy gap), and so most pure semiconductors are very poor conductors. In order to make a semiconductor a better conductor it is possible to "dope" it by adding a relatively small amount of an impurity element. For example, it is possible to dope silicon by adding atoms of an element that has five valence electrons, such as phosphorus or arsenic. Since only four electrons per atom are needed for crystalline bonding, the fifth electron can be freed to conduct electricity by adding to it only a small amount of energy.

Pure gallium arsenide has a larger band gap than pure silicon, and so it is an even poorer conductor at room temperature. Nevertheless, it is possible to dope gallium arsenide with elements such as silicon to make it more conductive; the silicon atoms replace gallium atoms, releasing one of the silicon's four electrons to contribute to electrical conductivity. Of the available semiconductors, gallium arsenide is one of the most suitable for creating ballistic-electron devices, as one of us (Eastman) first pointed out in 1977. Electrons in gallium arsenide have a higher maximum velocity than those in silicon (in gallium arsenide the maximum velocity is about 10^8 centimeters per second and in silicon it is about 2×10^7 centimeters per second), and they are less likely to collide with lattice vibrations. The mean free path of an electron in moderately doped gallium arsenide is about 1,000 angstrom units (one angstrom is 10^{-10} meter), roughly 10 times the mean free path of electrons in silicon.

A semiconductor that is often used in conjunction with gallium arsenide is the alloy aluminum gallium arsenide. When aluminum atoms replace gallium atoms in the crystal lattice, no electrons are released because aluminum and gallium have the same number of valence electrons. Any fraction of gallium atoms can be replaced by alumi-





FIELD-EFFECT TRANSISTORS can be built in several configurations. In a standard configuration (*top*) electrons flow through a conductive channel from a region called the source to a region called the drain, both of which are made of highly conductive semiconductor. When a negative voltage is applied to a metal "gate" above the channel, electrons are repelled from a region of the channel called the depletion layer. The conductive channel is thereby made effectively narrower, and so less current flows. The transistor is an amplifier: a small change in gate voltage effects a large change in current. The speed with which the transistor responds to a change in voltage is directly related to the speed with which electrons can travel from the source to the drain. In a vertical field-effect transistor (*bottom*), a novel configuration, the various elements are grown as thin films on top of one another. It is therefore possible to make the electrons' path from the source to the drain very short (and the device's response very rapid) simply by growing an extremely thin film for the channel. The ease with which the conductive channel can be made extremely thin makes vertical field-effect transistors suitable for ballistic-electron devices.

num atoms. Because the lattice structure of aluminum gallium arsenide is similar to that of gallium arsenide, it is possible to grow a crystal of one semiconductor directly on the surface of the other in such a way that the crystals match up almost perfectly. Such a configuration is known as a heterojunction.

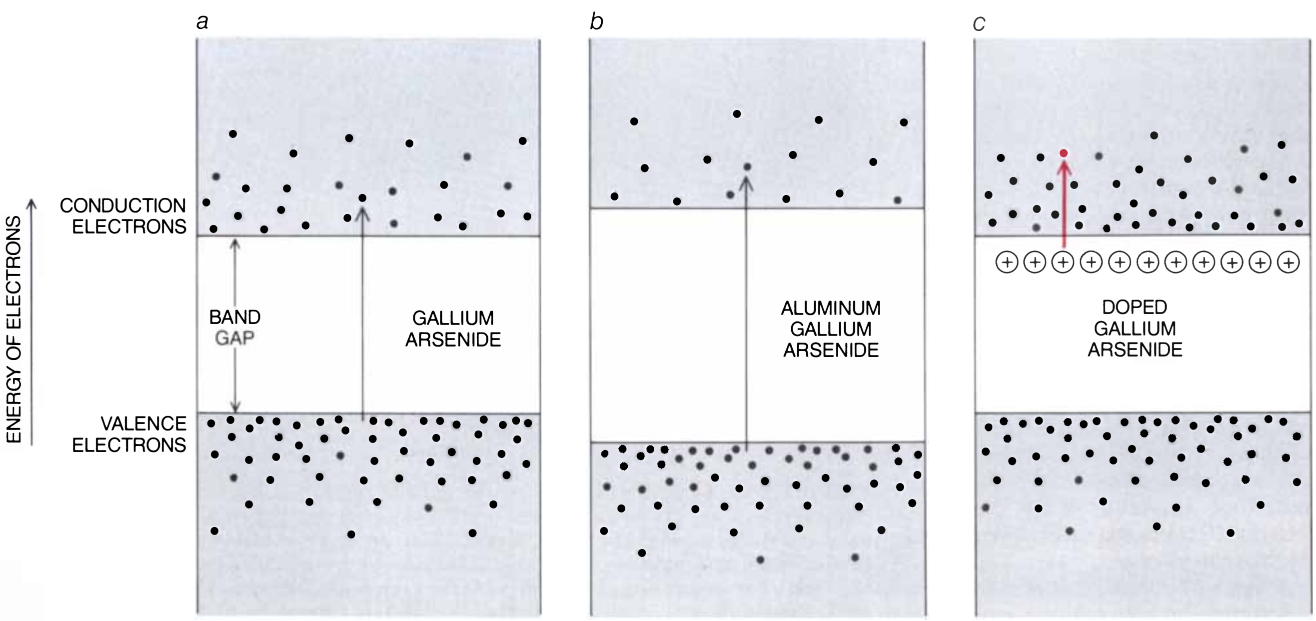
One of the most important differences between gallium arsenide and aluminum gallium arsenide is the difference between their band gaps. The size of the band gap in aluminum gallium arsenide alloys is directly proportional to the amount of aluminum they contain, and so aluminum gallium arsenide alloys containing high proportions of aluminum have wider band gaps than pure gallium arsenide or aluminum gallium arsenide alloys containing little aluminum. The conduction electrons on the aluminum gallium arsenide side of a heterojunction thus have higher potential energy than the electrons on the gallium arsenide side. When conduction electrons moving in the aluminum gallium arsenide pass into the gallium arsenide, their excess potential energy is abruptly converted into kinetic energy, and they are almost instantaneously accelerated to an extremely high velocity.

Such a heterojunction can be used as an injector of ballistic electrons. It represents one solution to the problem of accelerating electrons to a high velocity before they enter the channel of a transistor, where they might be made to travel ballistically. Two other kinds of injectors can launch electrons at higher or adjustable energies. They are called the planar-doped barrier (a variant of which is known as a camel barrier) and the tunnel barrier.

In a planar-doped barrier a plane of negatively charged ions (for example atoms of beryllium—which normally has only two valence electrons—that have an extra electron) is built into a thin sample of undoped gallium arsenide that is in turn sandwiched between regions of heavily doped gallium arsenide containing high densities of electrons [*see illustration on page 108*]. The atoms in the doping plane (the plane of beryllium) constitute a plane of negative electric charge. The plane of negative charge tends to repel conduction electrons, which must have a certain minimum energy in order to get past it from one doped region of gallium arsenide to the other.

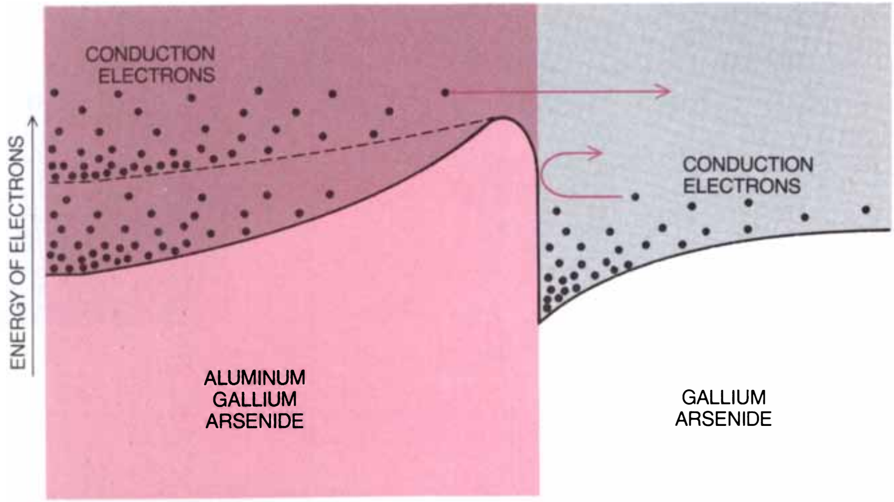
In terms of energy, the plane of beryllium represents a potential-energy peak. The potential energy rises from a low value in the heavily doped region at one end of the device, increases in the undoped region of gallium arsenide, reaches a peak at the plane of negative ions and decreases toward the other end of the undoped region, ending at a low value in the other heavily doped region. The height of the peak (the amount of energy required to cross it) is proportional to the number of negative ions, and it can be as high as the energy gap in the material. (In a camel barrier the beryllium is dispersed more widely than in a single plane, although the concentration is centered in a thin region, and so the energy peak is smoothed into a "hump.") By putting the plane of beryllium closer to one end of the undoped gallium arsenide than to the other, it is possible to create a very asymmetric barrier that has one long, gradually sloping side and one short, steep side. If a negative voltage is applied to the long side of the barrier, the electrons on that side acquire greater potential energy and move toward the plane of negative ions. It is as though the shallow slope of the hill has been raised and made even shallower. Eventually some of the electrons have enough thermal energy to pass over the plane of negative ions. As they descend the steep side of the potential-energy hill they are accelerated rapidly and injected at high speed into the nearby region of doped gallium arsenide, as long as the steep side is much narrower than the electrons' mean free path.

To make the other kind of ballistic-electron injector, a tunnel barrier, a thin sheet of undoped aluminum gallium arsenide is sandwiched between two regions of doped gallium arsenide that contain high densities of electrons. Normally the aluminum galli-



ENERGY DIAGRAMS of semiconductors explain many of their properties. The semiconducting crystal is held together by atoms sharing valence (outer) electrons. Electrons thus involved in bonding are localized between atoms and cannot move to carry current. They must be given a certain amount of energy to make them conduction electrons. The difference between the energy of valence electrons and that of conduction electrons, called the band gap, varies from one semiconductor to another. The band gap of gallium arsenide (*a*), for example, is smaller than that of aluminum gallium arsenide (*b*). A semiconductor can be made a better conductor by "doping" it: adding atoms that have more electrons than are needed for bonding and so give up electrons easily (*c*). Electrons from dopant atoms (*color*) need much less energy to become conduction electrons, leaving positively charged impurity atoms.

um arsenide would act as a barrier through which the electrons could not pass. If the barrier is less than about 100 angstroms thick, however, electrons on both sides have a certain probability of passing through it by means of a quantum-mechanical process known as tunneling. Even though the probability that any single electron will tunnel is very small, the large number of electrons in the doped gallium arsenide leads to a considerable number of tunneling electrons. If no voltage difference is applied to the two regions of doped gallium arsenide, the number of electrons tunneling in one direction is on the average exactly equal to the number tunneling in the other direction, and so no net current flows. If a voltage difference is established between the regions on each side of the barrier (that is, if one region of doped gallium arsenide is given higher potential energy than the other region), however, the number of electrons tunneling from the region with high potential energy far exceeds the number tunneling the other way, and current flows.



HETEROJUNCTION, an interface between two types of semiconductor (in this case gallium arsenide and aluminum gallium arsenide) grown together, is a way of injecting fast-moving electrons into one of the semiconductors. Conduction electrons on the aluminum gallium arsenide side have higher potential energy than those on the gallium arsenide side. (The curve in each semiconductor's energy diagram is caused by the proximity of the other semiconductor.) When a voltage is applied, the potential-energy function changes slightly (*broken line*) and electrons move from the aluminum gallium arsenide to the gallium arsenide. When they do, their total energy remains constant although their potential energy decreases. Their kinetic energy therefore increases: they move faster (*top, color*). Conversely, electrons on the gallium arsenide side have insufficient energy to enter the aluminum gallium arsenide and are turned back at the interface (*center, color*).

The probability that electrons will tunnel depends in part on the applied voltage and the thickness of the barrier. The total energy of the tunneling electrons does not change because tunneling is an elastic process, and so when they tunnel from a region in which electrons have high potential energy to a region of low potential energy, their kinetic energy increases: they are injected at high speed. In addition, the electrons that tunnel tend to be those that impinge on the barrier with the highest velocity and are moving in directions that are most perpendicular to the barrier, and so the injected electrons form a high-energy beam that is mostly aimed directly forward (perpendicular to the plane of the barrier). Moreover, the kinetic energy with which the electrons emerge depends linearly on the applied voltage, and so it is easy to adjust. The tunnel barrier is therefore particularly suitable for ballistic-electron devices.

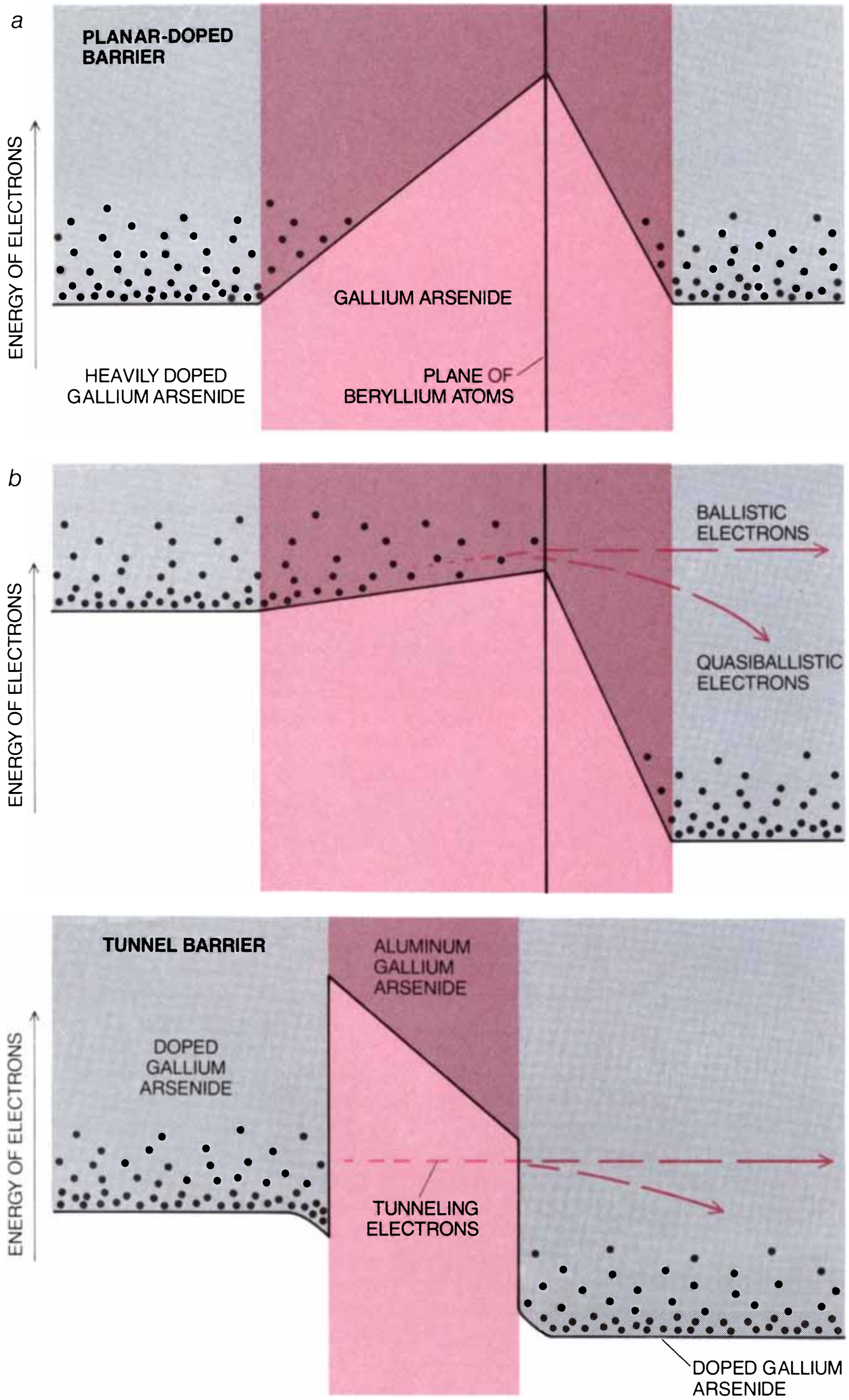
One novel device in which a ballistic-electron injector is particularly useful is known as a vertical transistor. In most commercially available transistors the source and drain regions are embedded in a horizontal sample of semiconductor, and electrons flow from one to the other in a stream parallel to the sample's surface. In a vertical transistor the individual elements of the device are actually layers of semiconductor grown on top of one another.

A vertical field-effect transistor can be made in the form of a thin vertical pedestal. At the bottom is an electrode; it lies directly under a layer of semiconductor that serves as the drain. One by one, layers that make up such elements as the channel and the source are grown on top of the drain region. A tunnel-barrier ballistic-electron injector, for example, can be built as the source simply by growing a layer of conductive gallium arsenide on top of a layer of aluminum gallium arsenide that has been grown on top of the channel (a layer of conductive gallium arsenide). The vertical field-effect transistor has the further advantage that the length of the conductive channel can be controlled very precisely: one simply varies the thickness of the layer of semiconductor that makes up the channel. In a horizontal transistor the channel can be shortened only by embedding the source, gate and drain very close together in the substrate, a process that is much harder to control precisely than the process of growing a thin layer of semiconductor. (The process of laying down the gate metal on the sides of a vertical pedestal is quite difficult, however.)

The first semiconducting hot-electron device (a device that relies on electrons that have had many collisions but have high energy and speed) was made in 1979 by John Shannon of the Philips research laboratory in England. In his device, which was made of doped silicon, the electrons were launched into the conductive channel by a camel-barrier injector. After they passed through the channel they were made to pass over yet another camel barrier before they reached the collector. The purpose of the second barrier was to prevent low-energy thermal electrons in the channel (which is usually called the base in hot-electron devices) from reaching the collector. The second barrier was made somewhat lower than the barrier of the electron injector, in order to allow some electrons to pass even if they had lost energy in collisions. When electrons were injected at sufficiently high energies, some were collected; they had lost a considerable amount of energy because the mean free path of electrons in silicon is short.

Shannon's transistor was followed by a gallium arsenide transistor fabricated by Roger Malik, Marc Hollis and Susan Palmateer at Cornell University in 1981. Their device had two planar-doped barriers made by beryllium doping: a high barrier acting as a hot-electron injector at the source (which launched electrons through the device at their highest possible velocity) and a lower barrier at the drain. Like Shannon, they observed hot-electron transport but did not look in particular for ballistic transport.

Until quite recently it has been exceedingly difficult to demonstrate experimentally that ballistic transport has indeed taken place in a device or to determine experimentally what frac-



BALLISTIC-ELECTRON INJECTORS of two types (called the planar-doped barrier and the tunnel barrier) launch fast electrons into a semiconductor. A planar-doped barrier is made by doping a sample of gallium arsenide with a plane of beryllium atoms (*top,* a). The entire assembly is sandwiched between two regions of gallium arsenide that have been doped to be rich with electrons. The beryllium atoms attract electrons, forming a plane of negative charge. The negative charge represents an energy "hill" that electrons do not cross. Applying the correct voltage to one of the electron-rich regions (*b*) raises the energy of the electrons in that region so that they cross the hill. When they reach the other side, their high potential energy is converted into kinetic energy. Some electrons experience a few collisions after crossing the hill; they are said to travel quasi-ballistically. A tunnel barrier (*bottom*) is made by sandwiching a thin layer of aluminum gallium arsenide between two regions of electron-rich, doped gallium arsenide. Electrons from the gallium arsenide cannot enter the aluminum gallium arsenide, but there is a small quantum-mechanical probability that they will tunnel through it. If the voltage difference between the two regions of gallium arsenide is adjusted so that the electrons in one region have higher potential energy than those in the other, those electrons will tunnel through the barrier. The velocity of the electrons can be tuned by changing the voltage.

tion of electrons that have traversed a device did so ballistically. In order to do so, a device would have to determine the distribution of energy (to measure inelastic scattering events) and directions of travel (to measure elastic scattering events) of electrons after they have completed their passage through the transistor and compare them with the initial energy and direction of the electrons as they enter the device.

Measuring the energy distribution of a beam of particles or radiation is known as spectroscopy. One way to make spectroscopic measurements of a beam of electrons is to place an "energy edge" in its path. An energy edge allows all electrons with energies greater than a certain value to pass through and turns back all electrons with energies lower than that value. By gradually changing the "height" of the edge (the minimum energy necessary to pass through it) and measuring the change in the number of electrons that pass, it is possible to find the relative concentrations in the beam of electrons with different energies.

A simple energy edge consists of a collector plate carrying a negative voltage, which repels electrons. Only electrons with energies high enough to overcome the repulsive force (that is, to pass over the potential barrier, whose height is equal to the applied voltage) can enter the plate to be measured as current. The height of the energy edge can be changed simply by changing the voltage applied to the collector plate.

How would one make a spectrometer that is sensitive to the direction in which the electrons are traveling as well as to their total energy? A simple example shows conceptually how it might be done. Suppose a bullet is fired directly at a thick, straight wall. The bullet will pass through the wall if it has an energy greater than some minimum value. If the bullet is deflected in flight, however, in such a way that it no longer flies in a direction exactly perpendicular to the wall, it will have to travel a greater distance within the wall; it will have to travel through what is effectively a thicker wall, and to do so it will need to have more energy. A thick, straight wall therefore measures something more specific than the total energy of the bullet: it measures the amount of the bullet's energy that is devoted to its motion perpendicular to the wall.

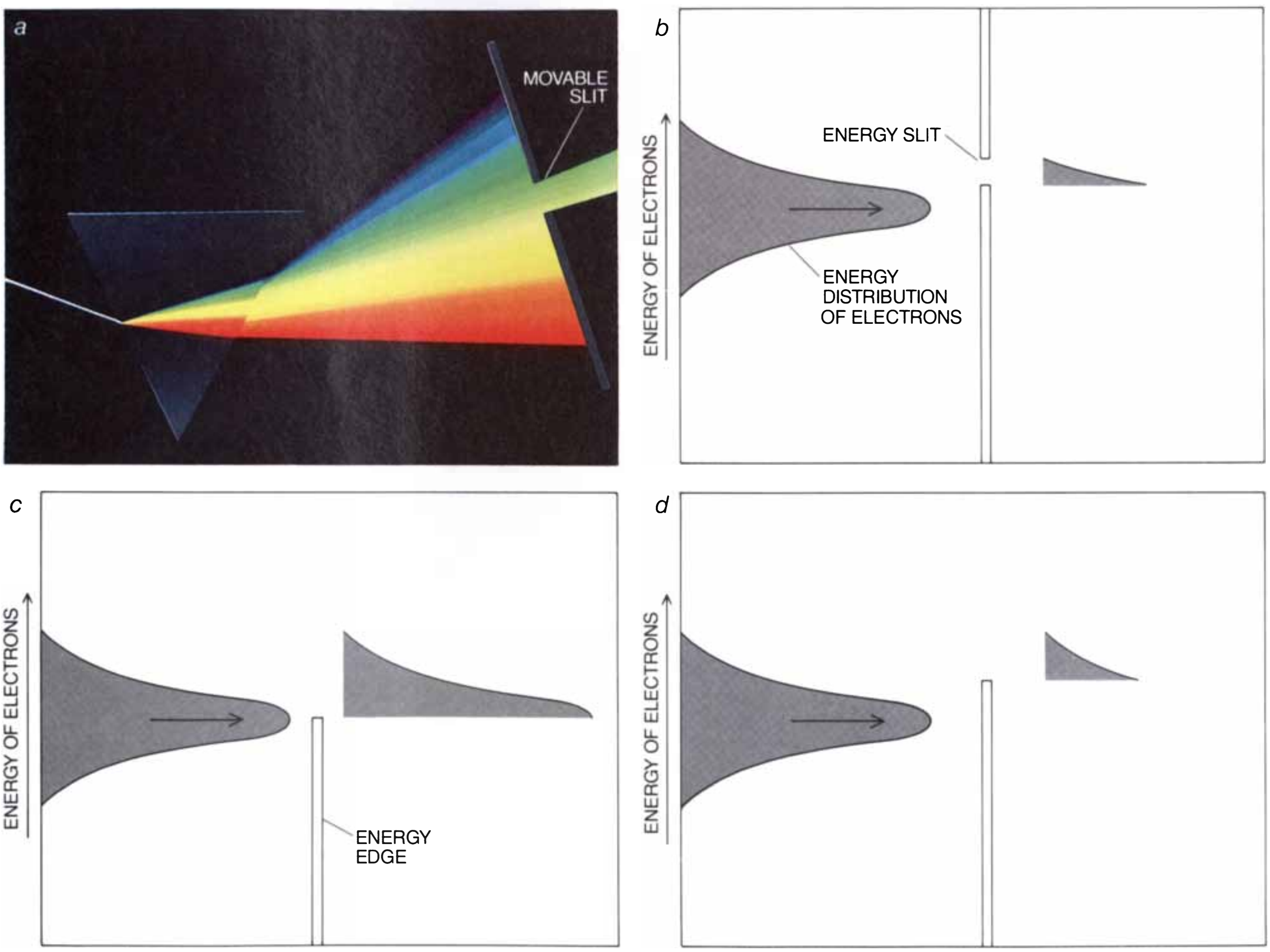
In a semiconducting material the bullets are replaced by electrons and the straight, thick wall is replaced by a thick potential-energy barrier (which could be a thicker version of the tunnel barrier that is in some hot-electron injectors). The height of the potential-energy barrier represents the minimum energy an electron traveling perpendicular to the barrier would need in order to cross it. If the barrier is thick, the tunneling probability is small and can be ignored. The energy distribution associated with the forward motion of the arriving electrons can be mapped by changing the height of the barrier and counting the number of electrons that make it through the barrier to contribute to current.

A device proposed in 1981 by one of us (Heiblum) is particularly well suited to this kind of spectroscopy because it relies on a well-collimated electron beam that has a well-known energy distribution. Doing spectroscopy in such a device enables the investigator to make a careful comparison of the initial and final conditions of the electron beam, thereby allowing a direct determination of any loss in energy or change in direction that might have occurred as the electrons passed through the device. The device is known as a tunneling hot-electron-transfer amplifier, or THETA.

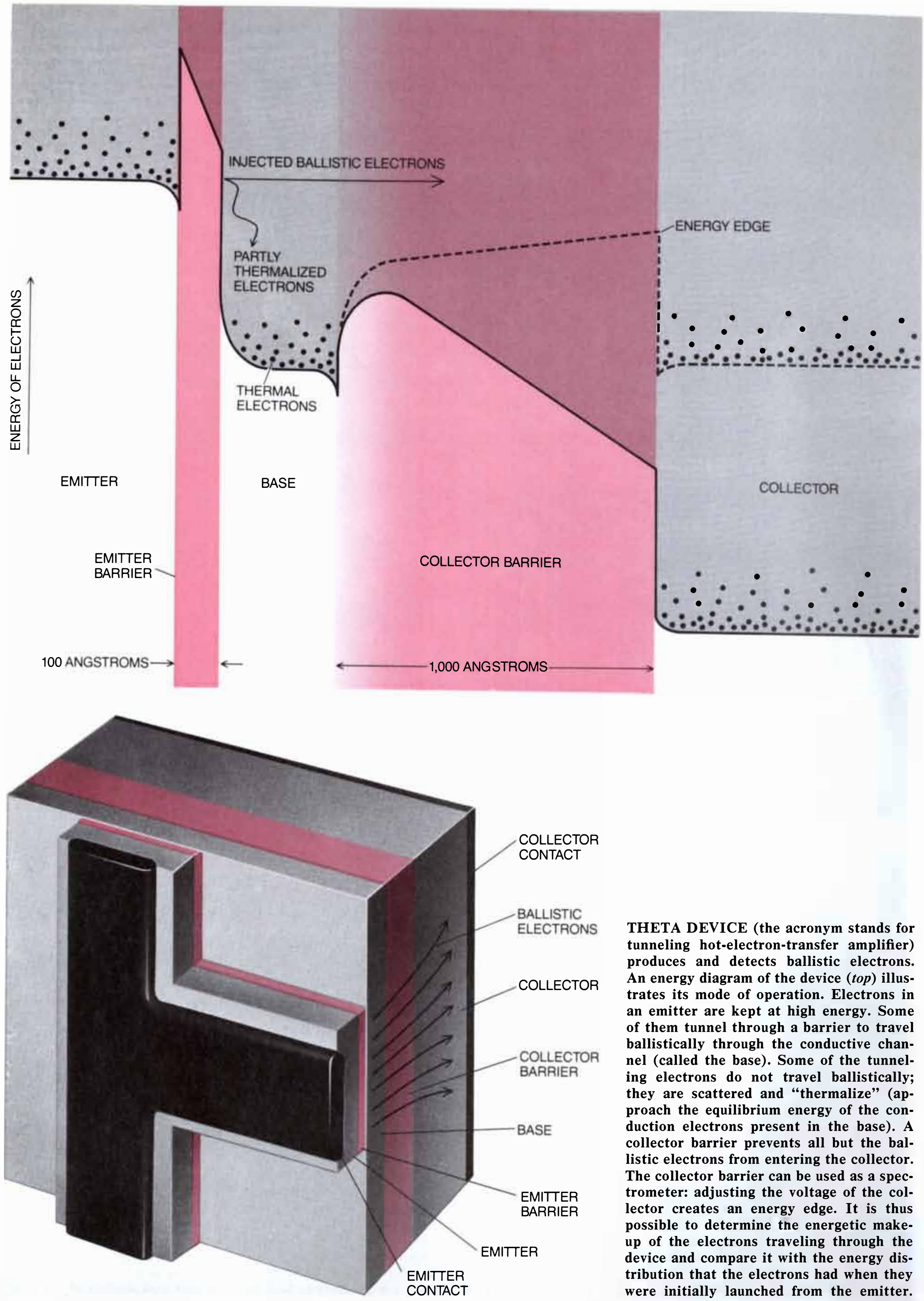
Most of the THETA devices fabricated so far have been made of gallium arsenide and aluminum gallium arsenide. They have been made by a group at the Fujitsu Laboratories in Japan and by Heiblum and his co-workers at the IBM Thomas J. Watson Research Center. In the THETA device a tunnel barrier injects electrons at very high velocity into a thin conductive layer called a base. The base is gallium arsenide that has been made conductive by being doped with silicon atoms, which provide the "cold" electrons, or electrons in thermal equilibrium with the lattice, that make it possible to apply a voltage difference across the tunneling barrier. The energies of the injected electrons are spread in a relatively narrow distribution about a central energy that is determined by the voltage applied to the tunneling injector.

The base is either thinner than or on the order of the length of the mean free path of the hot electrons, and so many of them traverse it ballistically. At the other end of the thin base is a thick barrier layer of aluminum gallium arsenide. The barrier prevents cold electrons from tunneling out of the base, but it can be surmounted by hot electrons that have traversed the base ballistically. On the other side of the barrier another layer of conductive gallium arsenide serves as a collector. In the collector the electrons lose most of their kinetic energy by various scattering mechanisms and are pulled toward the positive electrode of the battery [*see illustration on next page*].

To use the THETA device as an edge



SPECTROMETERS measure the distribution of energy in a beam of particles or radiation. The first spectrometer, devised by Isaac Newton, was a prism (*a*). It split white light into its component colors, each of which is made up of photons that have a certain energy. A movable slit allows only one color at a time to pass through and be observed. By moving the slit it is possible to determine the intensity of each color and hence the relative proportion of photons of each energy in the light. An analogous device for doing spectroscopy on a beam of electrons might be an "energy slit" that allowed electrons in only a certain range of energies to pass through (*b*). It is simpler to devise an "energy edge," which allows electrons above a given energy level to pass (*c*). By raising the edge (*d*) and noting the change in the number of electrons that pass, it is possible to find the energetic composition of the beam.



THETA DEVICE (the acronym stands for tunneling hot-electron-transfer amplifier) produces and detects ballistic electrons. An energy diagram of the device (*top*) illustrates its mode of operation. Electrons in an emitter are kept at high energy. Some of them tunnel through a barrier to travel ballistically through the conductive channel (called the base). Some of the tunneling electrons do not travel ballistically; they are scattered and "thermalize" (approach the equilibrium energy of the conduction electrons present in the base). A collector barrier prevents all but the ballistic electrons from entering the collector. The collector barrier can be used as a spectrometer: adjusting the voltage of the collector creates an energy edge. It is thus possible to determine the energetic makeup of the electrons traveling through the device and compare it with the energy distribution that the electrons had when they were initially launched from the emitter.

spectrometer, one injects hot electrons with a known energy distribution into the thin base while gradually raising the height of the thick potential barrier. The height of the barrier is raised by applying a negative voltage (which repels electrons) to the collector. The spectrometer's edge is the side of the thick layer of aluminum gallium arsenide that is closest to the collector, since this is the place where the potential-energy barrier is the highest and since only electrons with an energy greater than or equal to the barrier height are able to pass through the entire layer of aluminum gallium arsenide into the collector.

As the height of the barrier is increased, the current arriving at the collector decreases. The change in current that occurs with each change in the height of the barrier is proportional to the number of electrons whose energy (actually, whose energy associated with motion perpendicular to the barrier) is greater than the energy needed to surmount the barrier before its height was changed but less than the energy needed to surmount the barrier after its height has been changed. Hence by varying the height of the barrier it is possible to map the energy distribution of electrons that have traversed the device.

In several experiments of this type the number and spectra of collected electrons have shown extremely good agreement with the characteristics of the injected beam. The spectroscopic results are a direct proof that electrons traveled ballistically in the conducting layer of gallium arsenide and through the thick barrier of aluminum gallium arsenide. These experiments represent the first time such a narrow energy distribution has been seen in electrons that have passed through a semiconducting device and the first time the characteristics of the emerging electrons have shown such a close similarity to those of the injected electrons.

Although THETA devices, and other ballistic-electron devices, may eventually have applications in commercial semiconducting components, they have already shown their usefulness as scientific tools for investigating the behavior of electrons in semiconductors. The actual mechanisms underlying the passage of an electron through a semiconductor are not very well understood, and ballistic-electron devices give investigators an opportunity to change a few conditions within the semiconductor and gauge the effect of such changes on the passage of electrons.

For example, in THETA devices not all the injected electrons traverse the base ballistically. In a seeming paradox, however, if the base is made less conductive by decreasing the amount of doping atoms, the number of ballistic electrons increases. This result indicates that the hot electrons passing through the base interact with the cold electrons that make the base conductive (and perhaps also with the doping atoms). It is not yet clear what that interaction might be.

As another example, if one makes the base of the THETA device thicker, the energy distribution of the collected ballistic electrons does not change in shape although the total number of ballistic electrons decreases. This shows that each scattering event randomizes a single hot electron in such a way that it is plucked out completely from the distribution of ballistic electrons. In other words, there are no significant small-angle scattering events or inelastic scattering events that involve a very small loss of energy. Each collision is a "traumatic" one and makes a hot electron cold.

Another area of study in which ballistic-electron devices can be valuable experimental probes is the investigation of the fundamental quantum-mechanical properties of the electron. When electrons are scattered, particularly when they are scattered inelastically, they lose some "memory" of their past wavelike properties. In particular they lose information about the phase of the wave. Ballistic electrons, on the other hand, move unscattered through a solid medium and maintain their wavelike nature. For ballistic electrons it is possible to trace backward, in time and space, the motion of the electron wave. The relation between two ballistic electron waves remains constant as long as the electrons are ballistic.

The wavelike behavior of ballistic electrons in semiconductors can be understood by analogy to water waves and light waves. When a water wave strikes a solid wall, it sometimes is reflected: another wave is created, which moves back toward the source of the original wave. The forward-moving wave and the backward-moving wave interfere with one another. Where a crest of one wave meets a crest of the other, the total amplitude increases; this phenomenon is called constructive interference. Similarly, where the trough of one wave meets the peak of the other and vice versa, the total amplitude decreases; this phenomenon is called destructive interference. Light waves, when they pass from one medium to another (such as air and glass), are often reflected partially: some of the light passes through the boundary between the mediums and some is reflected in a wave that can interfere with the original ray. In a similar way, when ballistic electrons pass from gallium arsenide into aluminum gallium arsenide or from aluminum gallium arsenide into gallium arsenide, some of the electron wave passes through the barrier and some is reflected back toward the source of electrons.

Evidence for interference between the forward- and backward-traveling electron waves can be found in the THETA device. After the ballistic electrons have emerged into the thin base and traversed it, they encounter quantum-mechanical reflections at the interface of the base and the collector. Because the wavelength of an electron wave depends on its energy, the interference effects caused by the reflections depend on the energy of the injected electrons. For certain injection energies the reflected electron waves arrive back at the tunnel injector in such a way that they are exactly in phase with the waves emerging for the first time from the injector (that is, in such a way that they overlap exactly). The backward-moving wave then reinforces the forward-moving wave by interfering constructively with it and causes a greater amount of current to flow to the collector. At other energies the waves interfere destructively, and the backward wave partly cancels the forward one. The current to the collector is then diminished. These quantum effects result in easily observable local maxima and minima in the current that flows through the device as the energy of the injected electrons is varied.

It is surprising and in direct conflict with our day-to-day intuition that in very pure crystals at temperatures close to absolute zero electrons can theoretically be expected to travel great distances without being scattered by the dense matrix of atoms. In some ways it is even more incredible, however, that electrons can travel without scattering even in real devices, which are made of impure materials and operate at temperatures much greater than absolute zero.

The demonstration that electrons can be made to move ballistically is significant in several respects. As we have seen, the wavelike nature of ballistic electrons provides investigators with new tools for probing the behavior of electrons in solids and testing the quantum-mechanical view of the electron. In more practical terms, ballistic electrons in semiconductors make possible the design of new quantum devices that may be able to operate at speeds far exceeding the capabilities of the present generation of electronic devices.

Whales and Walruses as Tillers of the Sea Floor

As gray whales and Pacific walruses gather food from the bottom of the northeastern Bering Sea they produce pits and furrows to a degree that rivals the disturbances caused by geologic processes

by C. Hans Nelson and Kirk R. Johnson

In the late 1970's, while surveying the bottom of the northeastern Bering Sea for evidence of geologic hazards to possible offshore oil platforms and pipelines, one of us (Nelson) met with a surprise. In addition to the expected hazards, such as active faults, ice gouges and current erosion, there was evidence that the sea floor was marked by pits and furrows that could not be attributed to any known geologic process. Aware that many marine mammals live in or periodically visit the shallow waters (rarely more than 50 meters deep) between Siberia and Alaska, Nelson wondered if any of the animals might be the cause. Indeed they were. We have since discovered that California gray whales produce the pits; Pacific walruses produce the furrows. In the process the animals introduce significantly more sediment into the water of the northeastern Bering Sea than does the Yukon River, which annually discharges more than 60 million metric tons of sediment into the area.

Both the whales and the walruses modify the sea floor in the course of foraging on the continental shelf underlying the sea. The whales excavate huge patches of sediment from the Bering Shelf as they search for their preferred food, a bottom-dwelling shrimplike crustacean known as an ampeliscid amphipod [*see bottom illustration on page 116*]. Pacific walruses disturb the sediment as they forage for clams and some 60 other species of bottom-dwelling prey. In spite of the battlefield appearance of the sea floor, the feeding activities of the whales and walruses seem to be beneficial to the area, enhancing its productivity.

The walruses, which number about 200,000, are year-round inhabitants of the Bering and Chukchi seas. The whales, in contrast, are visitors. Every March some 16,000 of them leave their breeding grounds off Baja California and migrate more than 5,000 miles up the Pacific coast to feed. By the time they reach the northeastern Bering Sea, a principal feeding area, the winter ice has melted and the relatively calm summer waters teem with life. (The whales also feed elsewhere in the Bering Sea and in the nearby Chukchi Sea.) From May to November the whales indulge in a massive feast that provides the bulk of their yearly nourishment.

Nelson and his colleagues initially discovered the mysterious pits and furrows when they reviewed the data they had gathered from side-scan sonography. A side-scan sonograph is made when a ship drags a torpedo-shaped towfish, an underwater transducer. The towfish directs a fan of sound pulses at the sea floor and also receives return signals. These vary in strength depending on the acoustic reflectivity of the structures they encounter and the structures' distance from the towfish. The returning signals are relayed to the ship, where a recorder produces the sonograph, an image that somewhat resembles an aerial photograph taken of a land surface.

Circumstantial evidence immediately suggested that the gray whale might be causing the newly discovered sea-floor disturbances. (The realization that walruses were also involved came only later.) For one thing, the whales

CALIFORNIA GRAY WHALE emits a muddy plume from its mouth as it strains sediment for prey. This whale is representative of roughly 16,000 that feed in the northeastern Bering Sea, obtaining food by engulfing and expelling huge mouthfuls of sediment from the sea floor. As the whales search for bottom-dwelling prey, they create shallow, oval pits in the sea bottom. Some 200,000 Pacific walruses in the area also feed on bottom-dwelling animals; they form narrow furrows as they hunt for food.

had long been suspected of feeding on animals that lived in the sea bottom. As early as the 19th century, whalers reported seeing the mammals surface with plumes of muddy water streaming from their mouths. When the hunters cut the animals open, they found the stomach filled with large amounts of bottom-dwelling crustaceans that workers have since identified as ampeliscid amphipods.

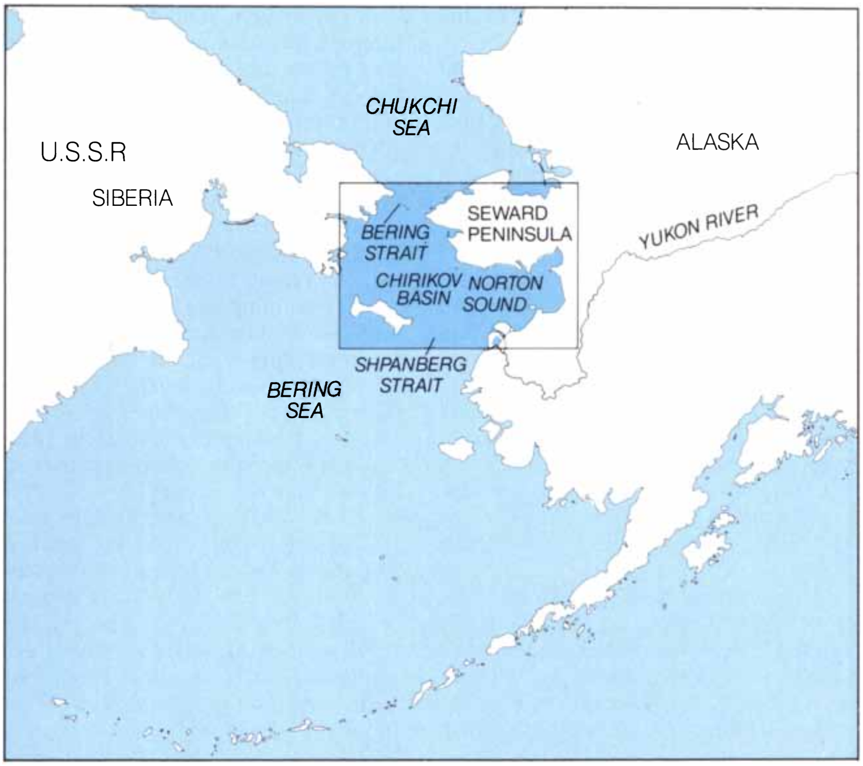
When a whale feeds at the bottom, it rolls on one side with its mouth parallel to the sea floor. By retracting its large tongue it creates suction, drawing a mouthful of fauna-rich sediment in through one side of its mouth. The whale then filters the sediment out the other side through the baleen: a series of fibrous plates that hang from each side of the upper jaw, forming a comblike structure. The hairlike interior of the baleen snares the prey, which the whale then swallows. Sometimes the whales expel streams of sediment near the sea bottom; at other times the animals release the sediment near the water's surface as they come up for air. (Although the whales clearly favor amphipods, they eat a variety of prey, including organisms that live in the water column.)

Two other findings suggested that the whales specifically caused the seafloor disturbances found in the Chirikov Basin, a 30-to-50-meter-deep region at the center of the northeastern Bering Shelf. In 1979 aerial observers who were tracking feeding gray whales on the basis of emitted mud plumes found that the whales were concentrated in the Chirikov Basin. Nelson had also evaluated sediment types in various parts of the Bering Shelf and had identified the fauna each supported. His data revealed that a sand sheet covering the Chirikov Basin was inhabited by the whales' favorite prey, the amphipods. Moreover, the extent of this sand sheet matched the area of the basin where the whales had been sighted.

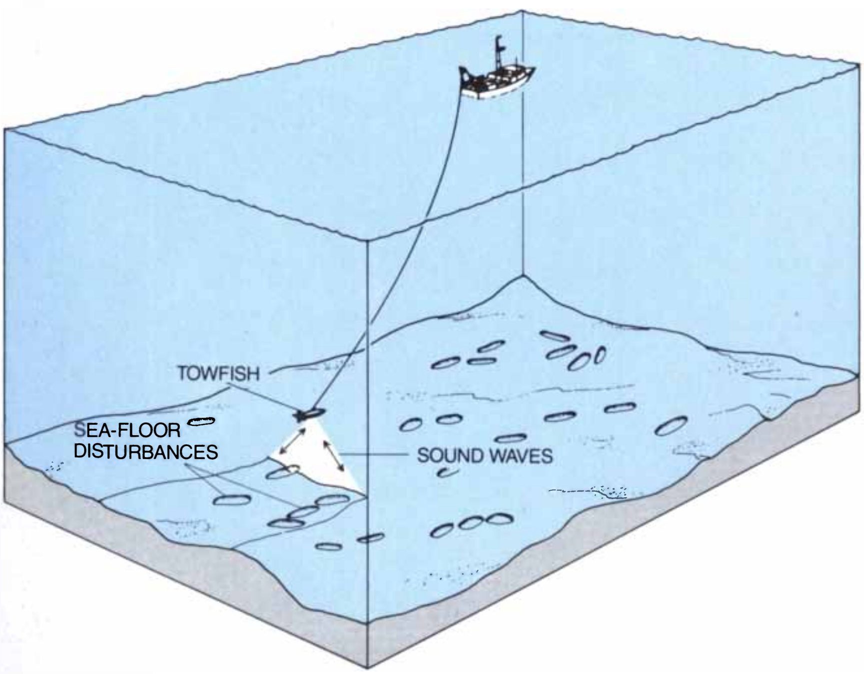
To determine the types of sediment and their associated fauna in the study area, Nelson had obtained sediment samples called "box cores." Each core covered a 20-by-30-centimeter area of the sea floor and penetrated as deep as 50 centimeters below the surface. In X-ray images of slabs cut from the samples, he could see the burrows created by bottom-dwelling animals and thereby determine the fauna the sediment supported.

The bottom of the Chirikov Basin turned out to be covered by as much as two meters of sand, deposited at the end of the last ice age (approximately 12,000 to 10,000 years ago). When the ice melted, it caused the sea level to rise, flooding areas that had once been land. The amphipods are the dominant burrowing animals found in this ancient sandy layer, where they can readily construct the mucus-lined tubes in which they live. The tubes bind the sand into a firm mat that resists erosion by currents.

On the northern and western margins of the basin a coarser mixture of glacially deposited gravel and sand predominates; to the east, in Norton Sound, mud prevails. Mud is a mixture of silt (which is finer than sand) and clay (which is finer still). The mud in the sound is several meters thick, mainly because the sediment dis-



NORTHEASTERN BERING SEA (*boxed*), at the apex of the Pacific Ocean between the Soviet Union and Alaska, is one of the principal feeding areas of California gray whales and Pacific walruses. The whales migrate there each summer from their breeding grounds off the coast of Baja California and concentrate their foraging in the Chirikov Basin. Walruses live in the north and feed in or near the basin for part of the year. The water in the northeastern Bering Sea is shallow (rarely deeper than 50 meters), allowing the bottom-feeding mammals to reach the surface for air without having to expend much energy.



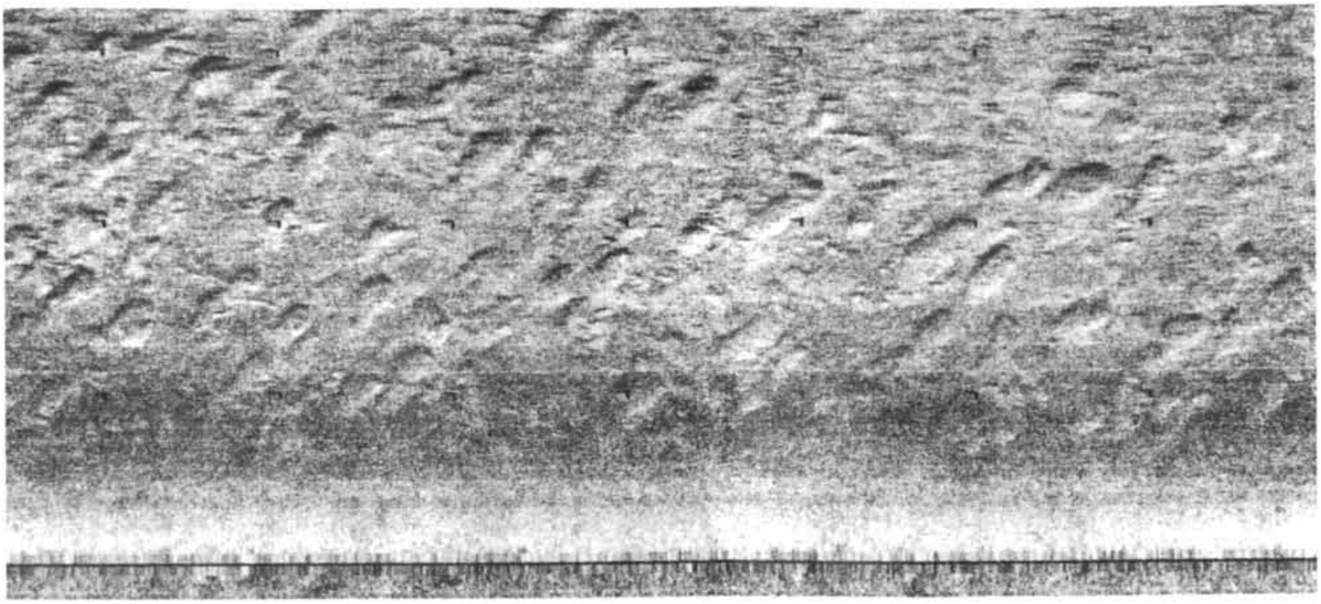
SIDE-SCAN SONOGRAPHY provided the first documentation that whales and walruses profoundly disturb the sea floor as they feed. A boat tows a torpedo-shaped transducer, known as a towfish, through the water. The towfish sends sound waves to the sea bottom and receives return signals, which vary in intensity depending on the shape and density of the structures they encounter. The return signals are then relayed to a shipboard recorder that translates them into a sonograph, an image that looks something like a photograph.

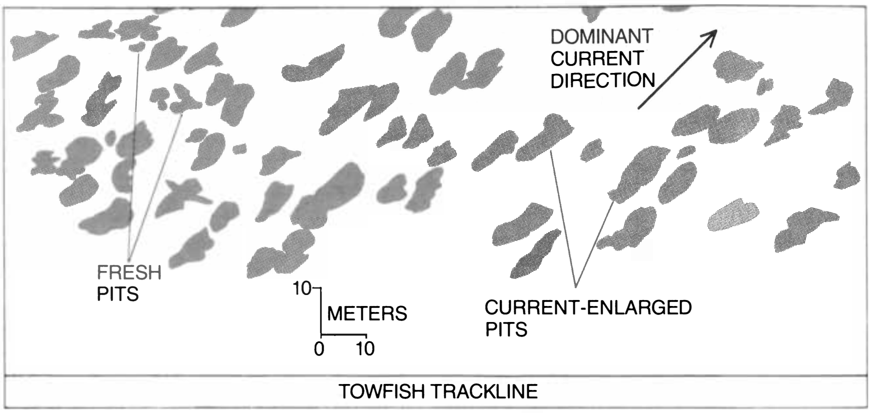
charged by the Yukon River tends to collect there. In a finding that became important when we eventually realized that walruses make furrows in the sea floor, Nelson and his colleagues noted that clams, the food favored by the walrus, dominate much of the gravelly and muddy areas surrounding the Chirikov Basin.

Convinced by the early 1980's that the Chirikov Basin was a major gray-whale feeding area, we attempted to confirm that the pits and furrows on Nelson's early sonographs were in precisely those regions inhabited by the amphipods and where the feeding whales had been sighted. We also had to measure the disturbances to determine whether they could have been produced by the feeding animals. If we found that the whales did indeed excavate the sea floor while eating, we then planned to determine the amount of food the area provided each year. We also hoped to determine the effect of the whales' feeding activities on the composition of sediment and on the food web in the northeastern part of the Bering Sea.

We began by mapping the exact location of all the pits and furrows presumably created by the animals. To do so we systematically examined all Nelson's sonographs as well as the sonographs made by other investigators; these data covered some 4,500 linear kilometers of the sea bottom. Nelson had produced his initial sonographs with 100-kilohertz (100,000 cycles per second) sound waves. The sonographs were quite adequate for identifying known, large-scale geologic features, but their distortion and low resolution made it difficult to determine the size and shape of smaller, unfamiliar sea-floor disturbances. (Small, fresh whale pits are among the smallest structures 100-kilohertz equipment can detect.)

Nelson and his colleagues had also made additional sonographs—covering the entire Chirikov Basin—with digital 105-kilohertz side-scan sonar. This system sends returning sound signals through a computer that adjusts for distortion. The resulting sonographs reflect the true shape and size of sea-floor features more accurately than nondigital side-scan sonographs. Help also came from Mary K. Nerini of the National Marine Mammal Laboratories in Seattle and Denis H. Thomson of LGL Ecological Research Associates, Inc., who were independently studying the feeding behavior of gray whales. At our suggestion they employed side-scan sonar generating 500-kilohertz sound waves and





SONOGRAPH and map of pits made by feeding whales show two kinds found on the floor of the Chirikov Basin: fresh, or newly formed, pits that are about the same size as the gray whale's mouth, and much larger pits that have apparently been enlarged by bottom currents. The influence of the currents is indicated by the fact that the enlarged pits tend to be oriented in the same direction as the flow of the dominant current (*arrow*). This sonograph was produced with 105-kilohertz equipment that converts return sound signals into digital electronic impulses, so that an onboard computer connected to the recorder can then eliminate some of the routine distortion found in many sonographic images.

produced high-resolution sonographs of small selected areas.

By correlating the side-scan data with direct observations made by divers working with Nerini and Thomson we were eventually able to identify not only the locations of the pits and furrows but also their shapes, sizes and probable origins. The pits, which were oval, were clustered within an area of 22,000 square kilometers, primarily in the central Chirikov Basin and just south of nearby St. Lawrence Island. Within the basin itself shallow pits marred as much as 18 percent of the bottom; these ranged in size from one to 10 meters in length, .5 to seven meters in width and .1 to .4 meter in depth. We knew of no geologic process that could carve such pits in the basin. Farther east, in Norton Sound, methane produced by the decomposition of buried organic matter forms circular gas craters, but the sand of the Chirikov Basin is too permeable for gas-crater formation.

The distribution of the oval pits corresponded to the distribution of feeding whales seen in aerial surveys. It also corresponded to the extent of the sand sheet and its resident amphipods in the Chirikov Basin. The whales certainly appeared to have made the pits.

To confirm further that the pits in the Chirikov Basin could be made by the gray whales, we measured more than 1,000 pits and compared their sizes to the size of the animal's mouth. A review of existing data on 240 gray whales revealed that the average gape length, or length of one side of the closed mouth, is 2.1 meters. On the average the pits were 2.5 meters long, 1.5 meters wide and 10 centimeters deep. They could easily have been made by a whale that kept its mouth parallel to the sea floor while sucking in sediment. Sometimes the pits were clustered in a way that suggested the animals often create a series of pits before they surface for air.

In addition to the 2.5-meter-long pits we found a number of larger pits averaging eight meters in length and four meters in width. Whales can conceivably make such pits if they move while taking in sediment or if they feed on the margins of old pits. It is also possible that currents—in particular the strong bottom currents associated with storms in the fall—enlarge freshly formed pits or groups of pits. According to our side-scan data, the large pits tend to have the same orientation as the predominant bottom current in the area. By ingesting amphipods and their surrounding tube mat, the whales presumably expose loose sand, which is more easily eroded by currents than the mat would be. (Eventually the pits stabilize as amphipods colonize them and rebuild the mats.)

The other major features—long, sinuous, narrow furrows—appeared most often within a 6,600-square-kilometer area around the margins of the Chirikov Basin and in the surrounding straits. Rapid bottom currents and relatively coarse sand are common there. The furrows average 47 meters in

PIT freshly formed on the sea floor by a gray whale is approximately 2.5 meters long, 1.5 meters wide and 10 centimeters deep. In order to capture bottom-dwelling prey, a whale rolls onto its side and sucks in sediment through one side of its mouth. Then the animal separates the prey by straining the sediment through its baleen: a series of fibrous plates that hang from the upper jaw. Generally animals larger than approximately four millimeters become trapped in the mouth, whereas many smaller animals flow back into the water. Larry Martin of LGL Ecological Research Associates, Inc., made the photograph.

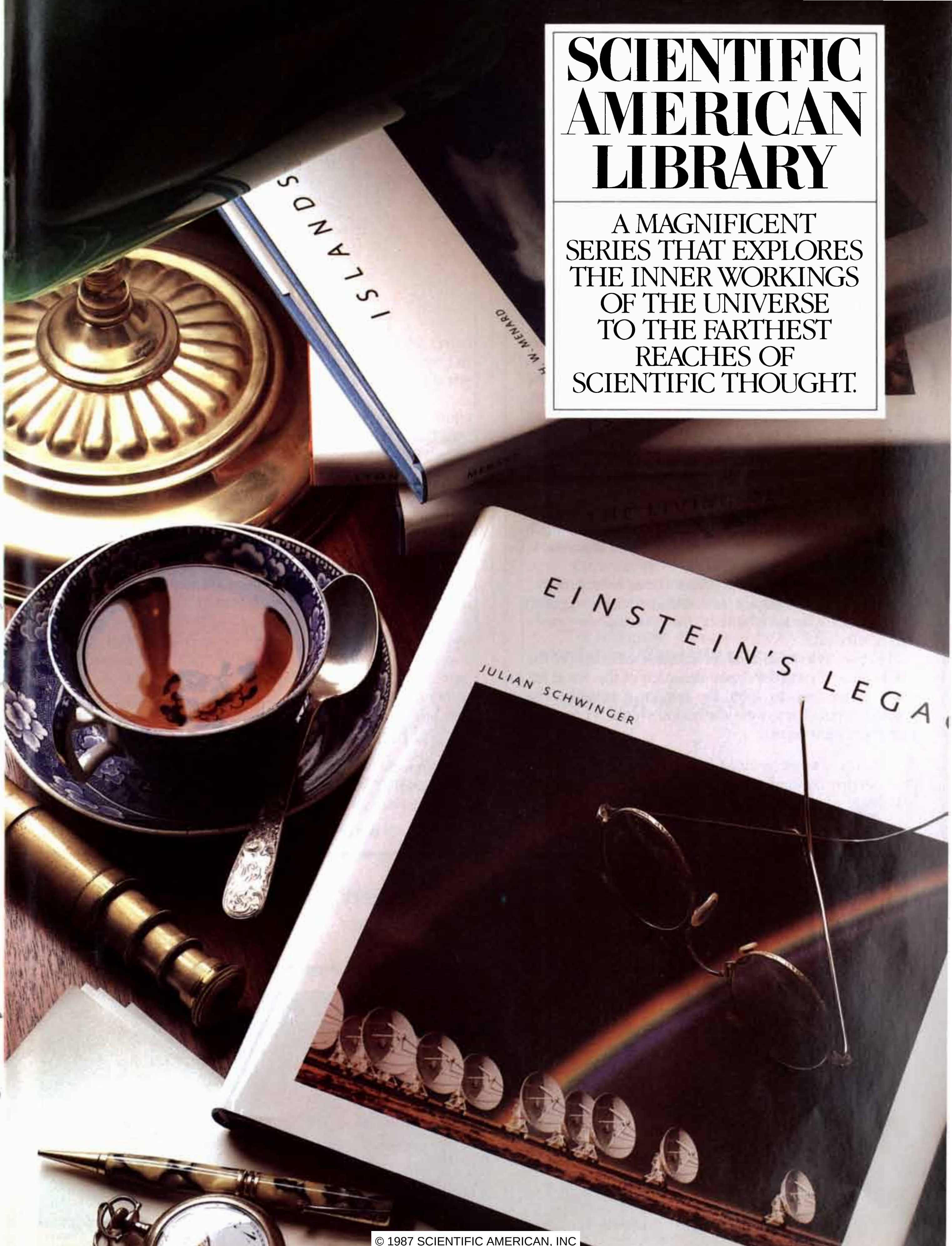
MAT of mucus-lined tubular burrows was built by ampeliscid amphipods: the bottom-dwelling, shrimplike crustaceans that are the gray whales' preferred prey. The network of tubes binds loose sediment into a mat capable of resisting erosion by strong currents. Feeding whales destroy parts of the mat, leaving behind pits that are vulnerable to erosion, but only temporarily. Because the juvenile amphipods are small, they often escape through the baleen without being swallowed. Since the amphipods thrive in a disturbed habitat, the spared animals rebuild the mat. By allowing juveniles to escape, the whales reseed the sea floor, preparing a new crop of amphipods for harvesting in the next season.

length and .4 meter in width, although some are as long as 150 meters. At first glance such furrows could be mistaken for gouges formed by keels of sea ice scraping along the bottom. Ice gouges, however, generally appear on shallower shelves (rarely deeper than 20 meters); they often are wider (more than a meter wide), more angular and arrayed in parallel groups.

There was no obvious way for the whales to have made these narrow furrows. On the other hand, observations by biologists suggested that walruses could readily produce them. In 1972 Samuel W. Stoker of the University of Alaska had sighted furrows from a submarine. Because he saw walruses feeding nearby, he suggested the tracks might be their feeding traces. Some 10 years later, while we were puzzling over the furrows, John S. Oliver of the Moss Landing Marine Laboratories showed that clam shells, which appeared to have been excavated and emptied recently, could often be found next to such furrows. Both observations confirmed the findings of Eskimos who, having found sand and gravel along with clams in the stomach of walruses, had concluded that the animals were bottom feeders.

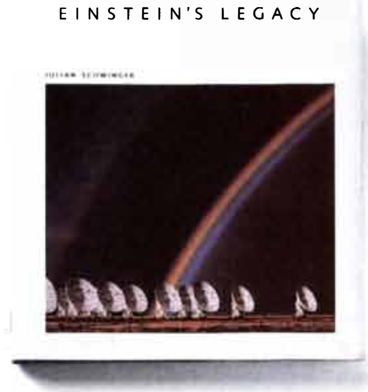
We too noted that the furrows generally were found in clam beds and that the width of the furrows was similar to the width of a walrus's snout. The latter finding was consistent with Oliver's suggestion that the walrus does not unearth clams with its tusks, as had been assumed. Rather it swims head down along the bottom, looking for prey or sensing them with its vibrissae, whiskerlike projections that cover its snout. It excavates the clams with its lips or with water jetted from its mouth, after which it clamps the prey between its lips and sucks the meat out of the shell. Mature walruses are estimated to need up to 85 kilograms of food a day.

Having determined the location and sizes of the whale pits and walrus furrows identified on the sonographs, we took advantage of a rare opportunity to measure the amount of food obtained by large marine mammals in a feeding area. Because the side-scan sonar undoubtedly missed many walrus furrows, we were not able to determine the amount of food those animals consume. (The sonar does not, for instance, detect narrow furrows that are oriented perpendicular to the direction of the towfish) We did, however, estimate the amount of food the whales gather annually from the Chirikov Basin.


SCIENTIFIC AMERICAN LIBRARY
A MAGNIFICENT SERIES THAT EXPLORES THE INNER WORKINGS OF THE UNIVERSE TO THE FARTHEST REACHES OF SCIENTIFIC THOUGHT.
ISLANDS
H. W. MENARD
EINSTEIN'S LEGA
JULIAN SCHWINGER

EINSTEIN'S LEGACY

Julian Schwinger



Nobel laureate Julian Schwinger tells the story of one of the twentieth century's greatest achievements—the theory of relativity—which is largely the work of one man: Albert Einstein.

The groundwork was laid in the seventeenth century by Isaac Newton, who unified motion on Earth and in the heavens in a framework of absolute space and time; then, in the nineteenth century by James Clerk Maxwell, who unified electricity, magnetism, and light. But it was left to the sixteen-year-old Einstein to glimpse, for the first time, that the theories of these two giants were incompatible.

Schwinger makes a lively narrative of Einstein's quest for the reconciliation of this conflict, a quest that led to the unification of matter and energy, and of space and time in his special and general theories of relativity.

The special theory has had its awesome confirmation in mankind's command of the nuclear force. Schwinger shows how the general theory, in turn, has stood up for 70 years to experiments drawn from the theory by Einstein, himself, and by his successors, employing ever more ingenious instrumentation and carrying the proofs ever farther beyond the decimal point.

In the last chapter, Schwinger describes space-age experiments made possible by the technologies that incorporate the theory itself. Their outcomes may not only secure further confirmation of Einstein's legacy, but face it with difficulties that lead on towards a still more comprehensive theory.

Julian Schwinger was awarded the Einstein Prize in 1951, the National Medal of Science in 1964, and the Nobel Prize for physics in 1965.

He is currently University Professor of the University of California, Los Angeles. He received his Ph.D. from Columbia University and has been on the faculty at Purdue University and Harvard University. Through the years, he has done theoretical work in various areas of both classical and quantum physics.

"I am happy to report, that his book makes for delightful and instructive reading. Technical demands on the reader do not exceed the most elementary algebra.

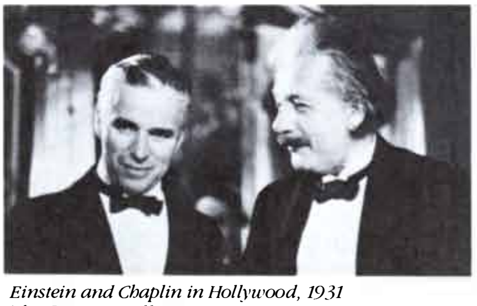
"It is particularly gratifying that the subject's intricacies, including glimpses of non-Euclidean and Riemannian geometries, and the theory's early successes, . . . are so well conveyed here in simple language. Altogether, this well printed and pleasingly illustrated book is an ideal gift for the curious non-expert." —NATURE

250 PAGES, 200 ILLUSTRATIONS

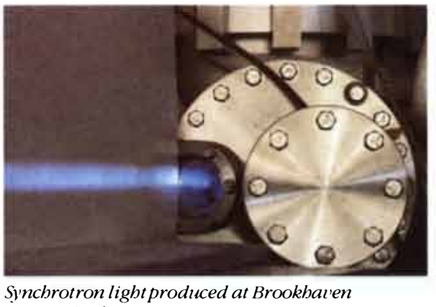
COVER PHOTO: The rainbow arches over the Very Large Array radio telescope of Soccoro, New Mexico. For nearly three generations, Einstein's legacy has withstood experimental test by instrumentation not yet invented when Einstein did his thinking. © Doug Johnson

TO GUARANTEE SWIFT DELIVERY OF YOUR FIRST VOLUME OF THE SCIENTIFIC AMERICAN LIBRARY, WE WILL SHIP *EINSTEIN'S LEGACY* VIA UPS.

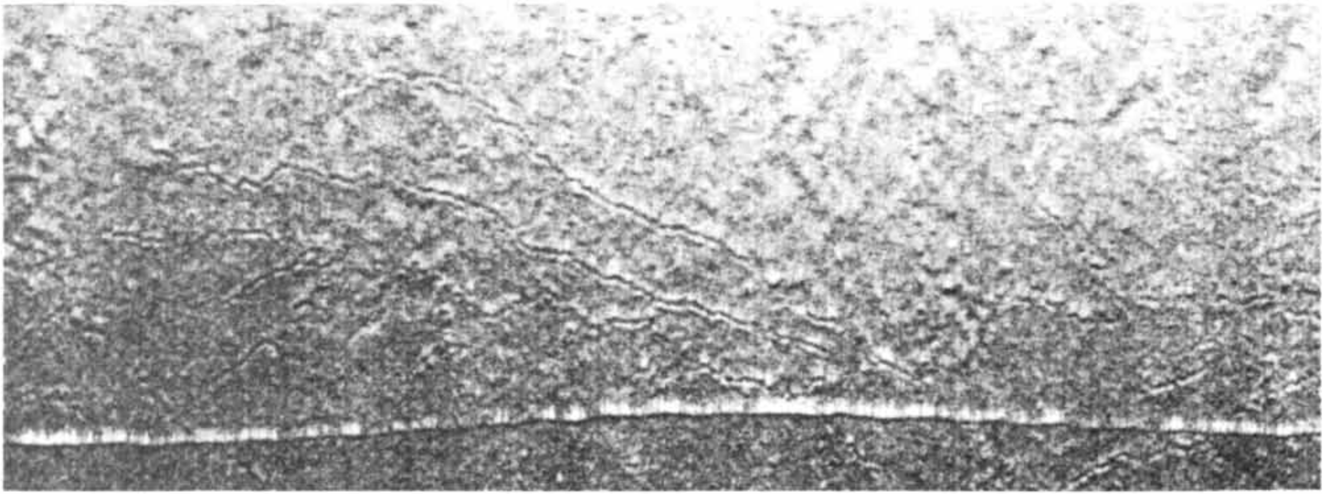
Radio telescopes
Stephen Northup/Black Star © 1980



Einstein and Chaplin in Hollywood, 1931
The Granger Collection



Synchrotron light produced at Brookhaven National Laboratory



FURROWS made by walruses as they forage for clams or other buried prey can be readily identified on a sonograph. In contrast to gouges made by ice keels dragging along the sea floor, walrus furrows are narrow and winding. Ice gouges tend to be wider, straighter and more angular. A walrus apparently forms furrows, which on the average are 47 meters long and .4 meter wide, with its snout. As the animal searches for food it is thought to have its head down and its snout in the sediment. It probably excavates clams by releasing a jet of water from its mouth. Then it sucks in the clam body, strewing the empty shells alongside the furrow.

We calculated that fresh (from one feeding season), unenlarged pits (ranging from two to four meters in length) disturbed 5.6 percent of the 22,000-square-kilometer whale-feeding area, or approximately 1,200 square kilometers. This is a conservative estimate because it ignores larger fresh pits made by whales in motion or feeding on pit margins.

Assuming on the basis of calculations by biologists that the area contained 171,000 kilograms of prey per square kilometer, we next estimated the total weight of prey consumed: approximately 205 million kilograms (171,000 kilograms multiplied by the 1,200 square kilometers). All that was needed then was to determine the food needs of the population. The gray whale population of 16,000 individuals needs approximately 1,100 kilograms of food per animal per day for an estimated 180 days (approximately 3.168 billion kilograms). We therefore concluded that the 205 million kilograms supplied by the Chirikov Basin area provides a minimum of 6.5 percent of the food needed by the animals; the maximum may reach 15 percent. The basin area, which represents just 2 percent of the gray whales' northern feeding range, is thus a particularly valuable territory for them.

Our side-scan data also enabled us to estimate the amount of sediment resuspended by the whales and led us to speculate on the impact of such disturbance on the ecology of the Bering Sea. If an area of 1,200 square kilometers (1.2 billion square meters) is disturbed and the average pit is 10 centimeters (.1 meter) deep, whale feeding resuspends at least 120 million cubic meters (172 million metric tons) of sediment per year. This is nearly three times the amount of suspended sediment the Yukon River discharges each year into the northeastern Bering Sea.

Whereas the Yukon River adds mud to the Bering Sea, the whale activity actually removes mud that has accumulated in the Chirikov Basin. After whales spew sediment into the water column, the heaviest particles tend to fall to the bottom again, but northerly currents carry clay and fine silt toward the Chukchi Sea. Such separation of the sediment prevents mud from covering the basin's ancient sand blanket and thereby maintains the area as an ideal amphipod habitat.

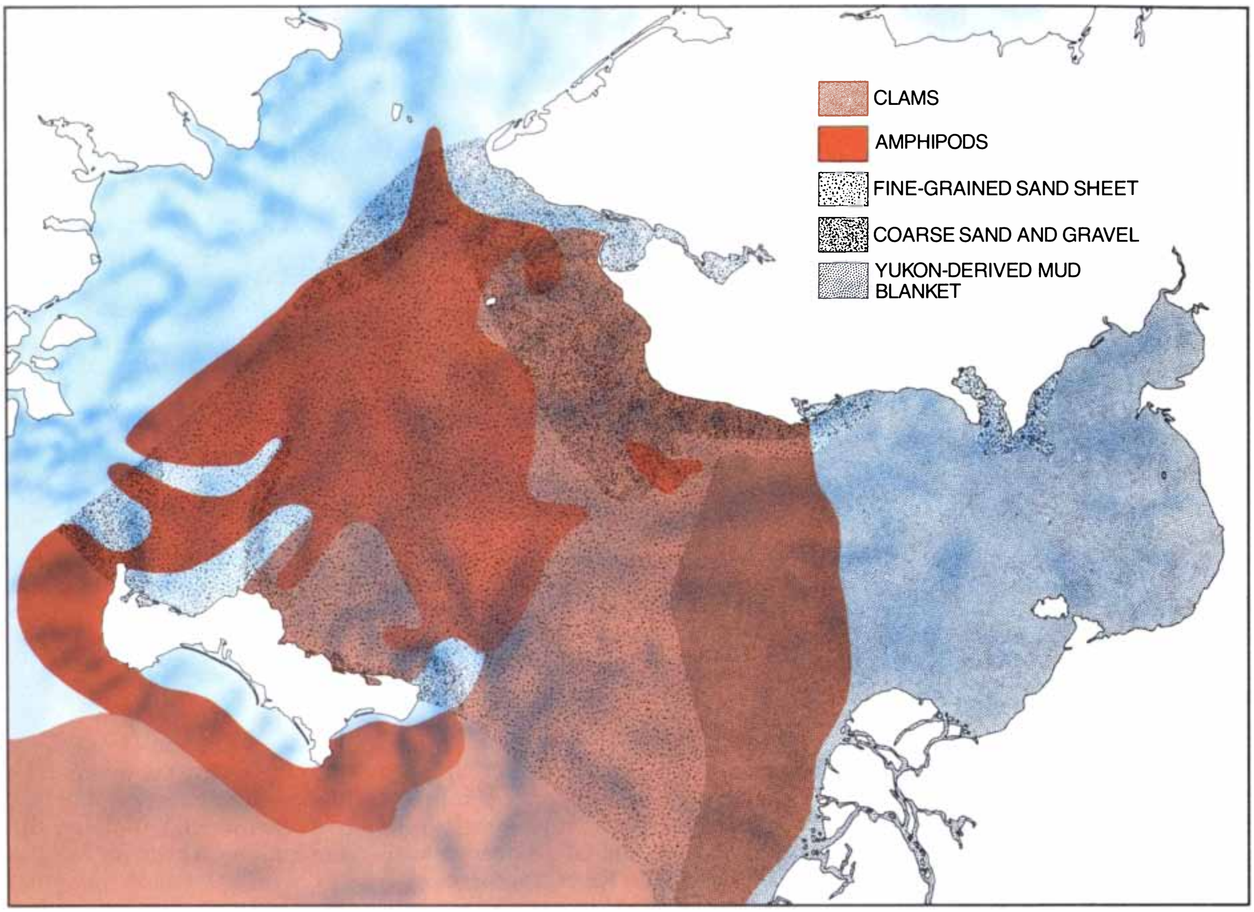
Separating mud from sand is only one of the apparently beneficial effects of the gray whale's feeding habits. In addition to keeping the sand in the Chirikov Basin "clean" for the amphipods, feeding whales inadvertently ensure themselves a healthy food supply for the next feeding season. The same baleen that retains adult amphipods seems to allow many smaller juveniles to escape into the water with the sediment; thus, much as a farmer casts seeds on newly plowed ground, the whales cast juvenile amphipods onto the sea floor. The amphipod is a pioneer species: it thrives in a disturbed habitat and quickly goes about reestablishing its tube mat.

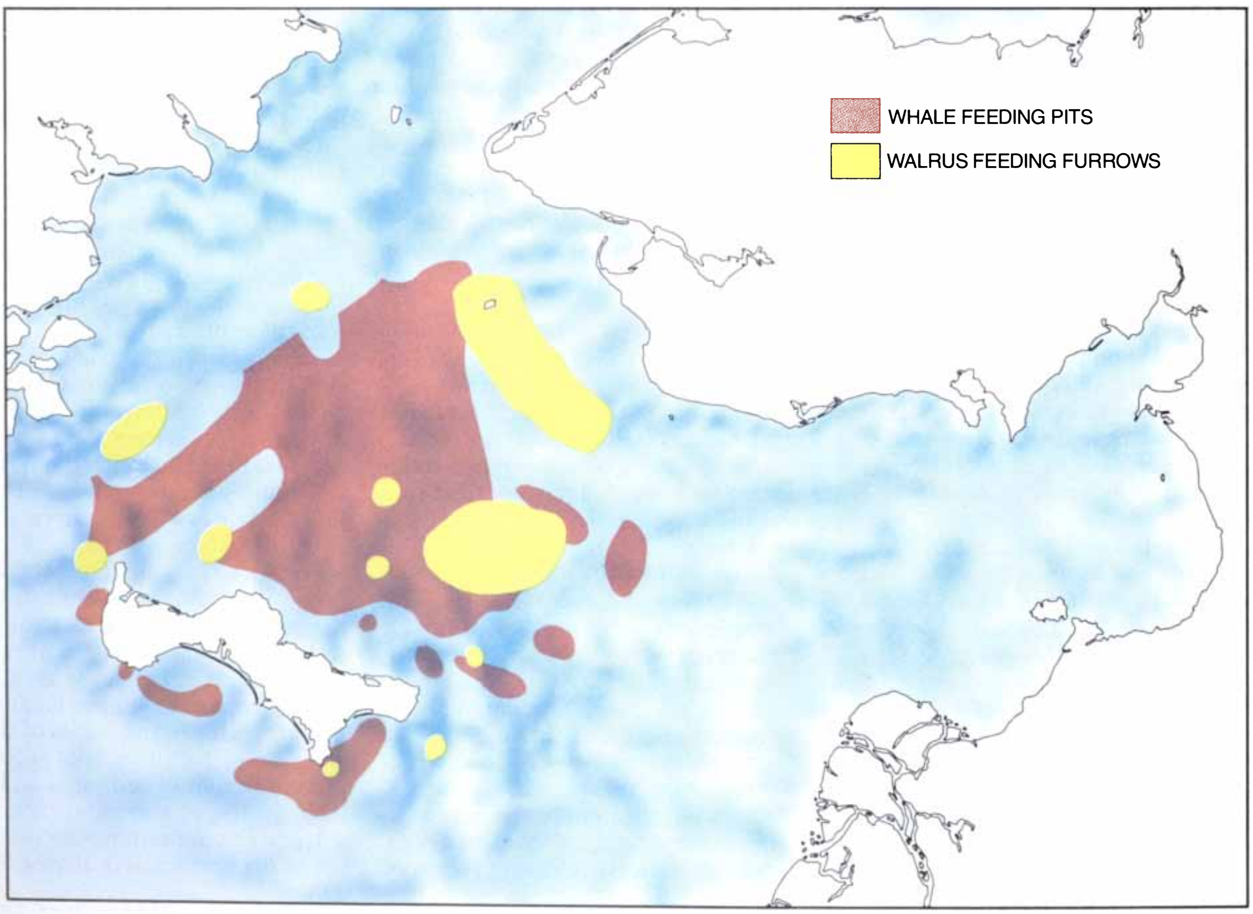
The whales promote the regeneration of the amphipod crop in another way as well. By releasing previously buried nutrients from the sediment into the water, the mammals enhance the crustaceans' food supply. Like the sun, the nutrients stimulate the growth of plankton, a primary food source for the newly "planted" amphipods. By the following year the whales find the tube mats replaced, and the cycle begins anew.

Although we could not measure the amount of sediment disturbed by walruses each year with similar accuracy, we did make a rough estimate. If the area's 200,000 walruses spend approximately 100 days in the northeastern Bering Sea, and if each animal digs at least two average-size furrows a day (47 meters times .4 meter times an estimated depth of .1 meter), then the walrus would yearly disturb about 75 million cubic meters (100 million metric tons) of sediment.

In contrast to the fine-grained fraction of the sediment disturbed by the whales, this sediment probably does not travel far. When sand and gravel are suspended in water, they tend to fall rapidly to the bottom instead of being carried away by currents. Moreover, the sediment disturbed by whales is resuspended well above the bottom and is sometimes even brought to the surface; the sediment disturbed by walruses is typically jetted immediately to the side of the newly created furrow. Compared with the feeding habits of the whales, the habits of the walruses seem to affect less profoundly the movement of sediment and hence the recycling of nutrients.

Rather than harming the balance of life in the northeastern Bering Sea, the


CLAMS
AMPHIPODS
FINE-GRAINED SAND SHEET
COARSE SAND AND GRAVEL
YUKON-DERIVED MUD BLANKET

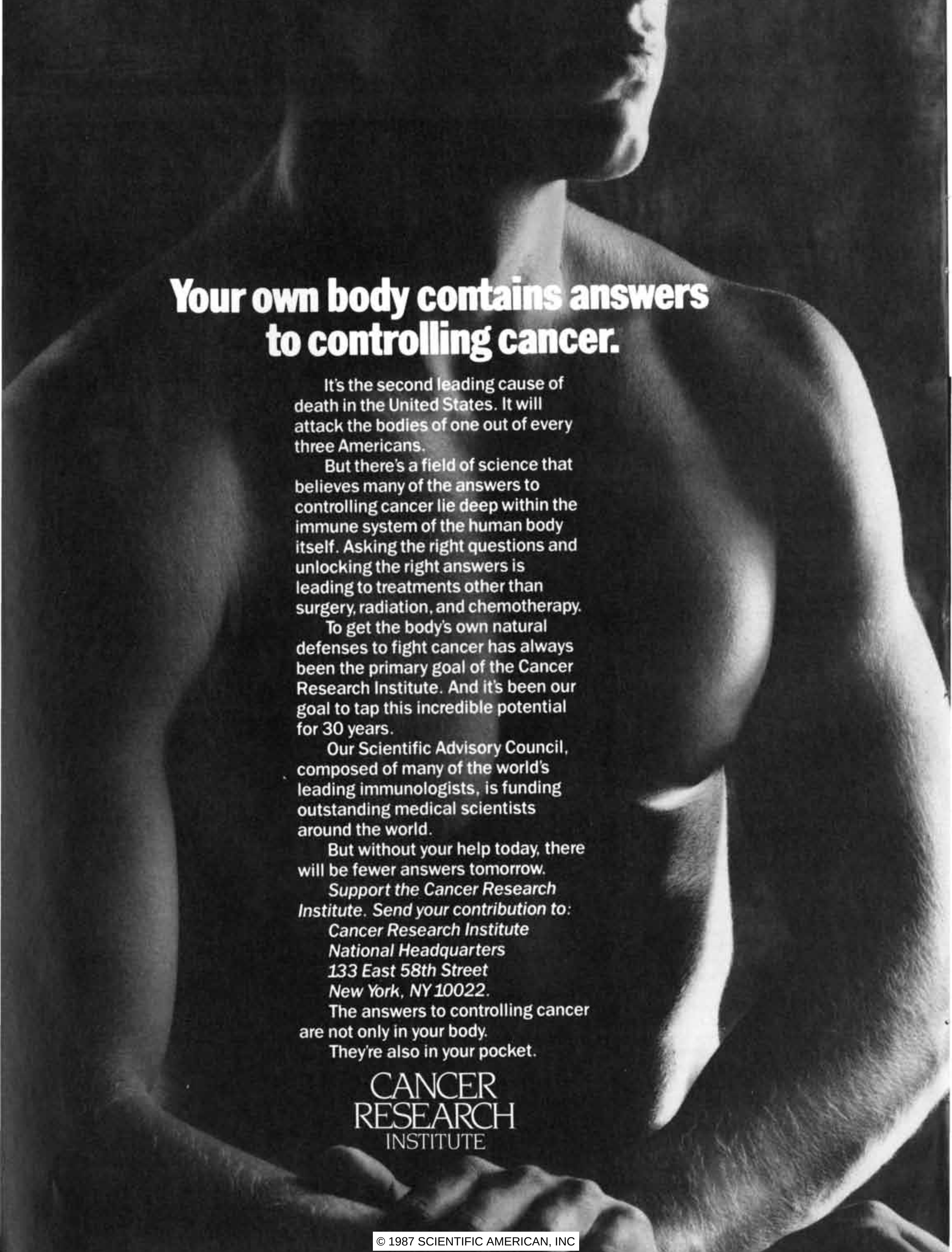

WHALE FEEDING PITS
WALRUS FEEDING FURROWS

profound sea-floor disturbance caused by feeding mammals is apparently an important part of the ecosystem. Certain activities of human beings, on the other hand, could interfere with this balance and particularly threaten the survival of the whales. If the Bering Sea is mined for sand or gravel, the ancient sediment that supports the amphipods—and therefore the whales—may be permanently destroyed. The sand layer covering the sea floor is very thin and modern geologic processes do not replace it; only another ice age could do that. An oil spill in the area could also be dangerous: amphipods appear to be extremely susceptible to oil contamination.

In the ancient past, natural phenomena must have repeatedly challenged the whales and walruses by decreasing their feeding areas. The animals seem able to adjust to such changes. During the Pleistocene epoch, which lasted from two million to 10,000 years ago, cycles of glaciation and melting produced drastic changes in the sea level. The gray whales and Pacific walruses probably developed their unusual and highly successful strategy of feeding on the sea floor at a time when the melting glaciers flooded the broad shelf that connects Alaska and Siberia, forming a shallow sea that supported abundant marine life. In times of glaciation the sea level dropped and the shelf became low-lying land, forcing the animals to search for food in a more limited area.

It is a testament to their ecological fitness that the whales and the walruses have survived repeated glacial periods and have markedly recovered from the intense hunting that severely threatened them some 50 years ago. The mammals' ability to prey on animals other than their preferred amphipods and clams and to "farm" the sea floor in times of plenty appear to have served them well.

MAPS identifying several characteristics of the northeastern Bering Sea helped to demonstrate that the pits and furrows detected on sonographs of the bottom were made by whales and walruses. A map showing the different kinds of sediment in the northeastern Bering Sea and the dominant fauna associated with them (*top*) indicated that fine-grained sand covers much of the Chirikov Basin and is dominated by the whales' favorite food: amphipods. The pits attributed to gray whales (*bottom*) are concentrated in the same area. A mixture of gravel and sand (*top*) predominates at the margins of the basin and is rich in clams, a favorite food of the walrus. The furrows attributed to walruses (*bottom*) are most often found in the margins of the basin. Walruses may feed in muddier areas but the authors have not found furrows there.


Your own body contains answers to controlling cancer.
It's the second leading cause of death in the United States. It will attack the bodies of one out of every three Americans.
But there's a field of science that believes many of the answers to controlling cancer lie deep within the immune system of the human body itself. Asking the right questions and unlocking the right answers is leading to treatments other than surgery, radiation, and chemotherapy.
To get the body's own natural defenses to fight cancer has always been the primary goal of the Cancer Research Institute. And it's been our goal to tap this incredible potential for 30 years.
Our Scientific Advisory Council, composed of many of the world's leading immunologists, is funding outstanding medical scientists around the world.
But without your help today, there will be fewer answers tomorrow.
Support the Cancer Research Institute. Send your contribution to:
Cancer Research Institute
National Headquarters
133 East 58th Street
New York, NY 10022.
The answers to controlling cancer are not only in your body.
They're also in your pocket.
CANCER RESEARCH INSTITUTE

Marriage, Motherhood and Research Performance in Science

Women publish less than men, but marriage and family obligations do not generally account for the gender difference. Married women with children publish as much as their single female colleagues do

by Jonathan R. Cole and Harriet Zuckerman

Studies of scientists' research performance, as gauged by their published productivity, find that women generally publish fewer papers throughout their careers than men matched for age, doctoral institution and field. Various explanations have been proposed to account for this disparity in scientific publication, ranging from systematic gender discrimination to biological differences, as yet undemonstrated, in scientific aptitude.

One frequent explanation holds that women, far more than men, bear the burdens of marriage and child care, and that this fact of social life best accounts for gender differences in scientific publication. Whether or not this is true, the belief that it is so affects women's career opportunities, their decisions and the way they are treated.

We decided (as part of a larger investigation of the careers of American men and women scientists) to test the counterclaim, made in earlier studies, that marriage and motherhood have no effect on women's research performance. We did so by assessing the dynamic relation of family life and womens' research throughout their careers—an approach not taken by earlier investigators, who just correlated the number of papers published with current marital and parental status. Our study draws on interviews with 120 scientists: 73 women and 47 men. We wanted to know whether scientists (both male and female) believe marriage and parenthood is incompatible with a scientific career in general, whether this had been the case for them in particular and what quantifiable effects (measured in number of publications) marriage and motherhood have actually had on the research performance of women scientists. Since men traditionally have not had primary responsibility for child care, we focused almost entirely on women, comparing publication rates for those who are married and those who are single, those who are mothers and those who are childless.

Publication counts are, to be sure, an imperfect indicator of scientists' contributions. Yet such counts are highly correlated with better measures, such as peer evaluation; moreover, the extent to which scientists publish is of major consequence to their careers. We therefore took the extent of publication as a rough but serviceable gauge of research performance. We recognized that women scientists are, in some sense, "survivors." By definition they have passed through the rigors of graduate training, have earned a doctoral degree and are employed in science. We did not seek to examine the impact of cultural expectations on women's chances of running this gauntlet; that would have called for investigating the processes by which women are winnowed out of scientific careers.

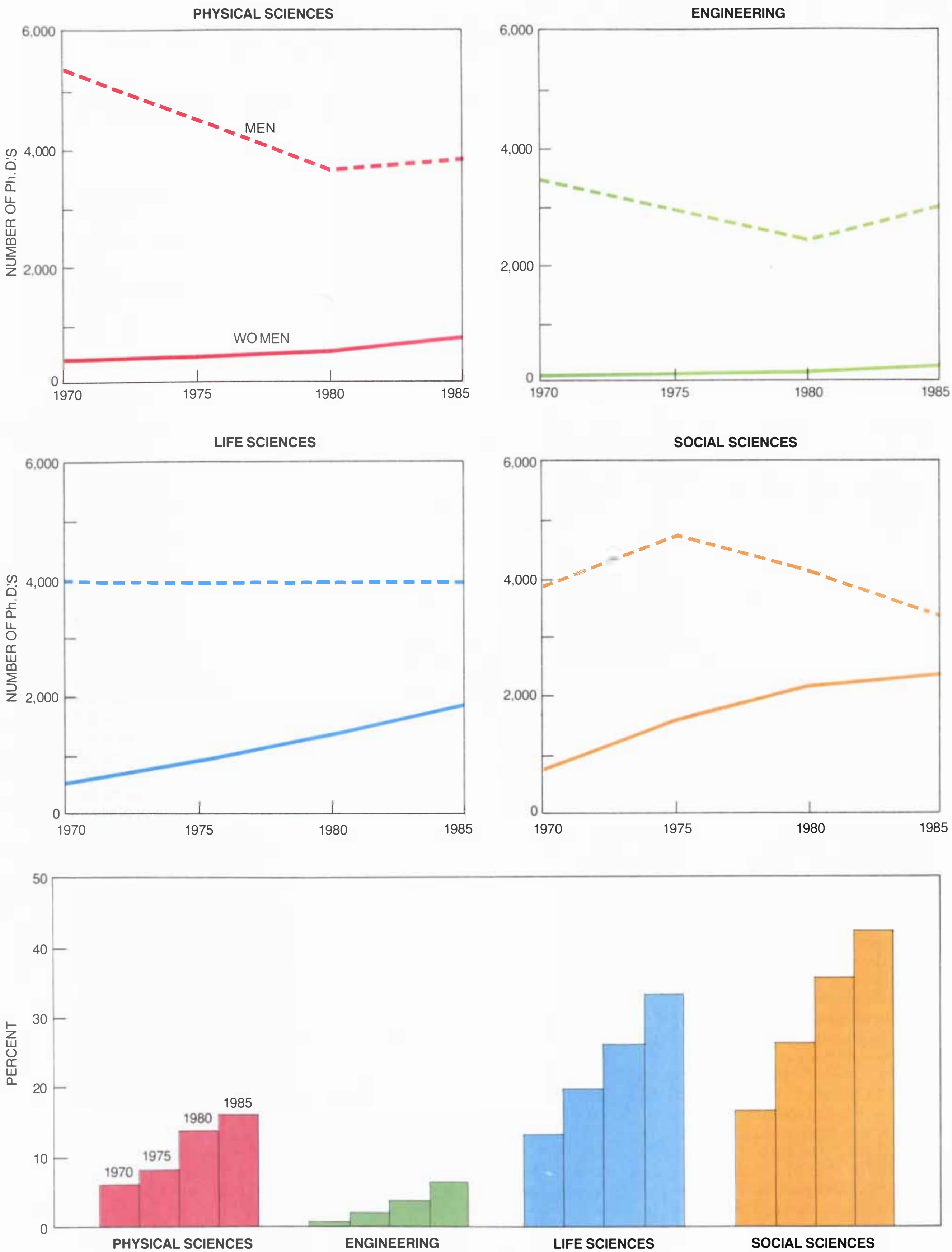
We chose subjects for the study by a stratified selection process that took into account gender, professional age, field of expertise and scientific standing. To compare the effects of marriage and motherhood on women who earned degrees in different historical periods, we divided the women into three age groups: those who received their doctorates between 1920 and 1959, before the advent of the women's movement, and during the decades from 1960 through 1969 and from 1970 through 1979, when the movement was getting under way and then becoming widespread. Eighty percent of the women were drawn from mathematics and the physical and biological sciences and the remainder from economics and psychology; the same 4 : 1 ratio was applied in each age group.

Scientists were further divided according to their peer recognition in relation to others of roughly the same professional age. The top tier of scientists (those we designated as "eminent") in the oldest group were members of the National Academy of Sciences or the American Academy of Arts and Sciences or were full professors in departments ranked in the top 10 in each field by national surveys of the quality of doctoral programs. In the intermediate age group, Guggenheim fellows or tenured professors in a top-10 department were classified as eminent. Younger scientists were considered eminent if they had held Guggenheim fellowships or were assistant or associate professors in a top-10 department.

Scientists not meeting those exacting criteria were designated as rank-and-file. They were randomly selected from lists of the faculty and research staff at accredited four-year colleges and universities in the same geographic regions as the eminent scientists. They were also matched to the eminent scientists for professional age and scientific field. Although these scientists were selected systematically, the criteria we applied, the small numbers of subjects and our exclusion of some important groups (such as scientists working in industry) mean this is decidedly not a true random sample of all U.S. scientists.

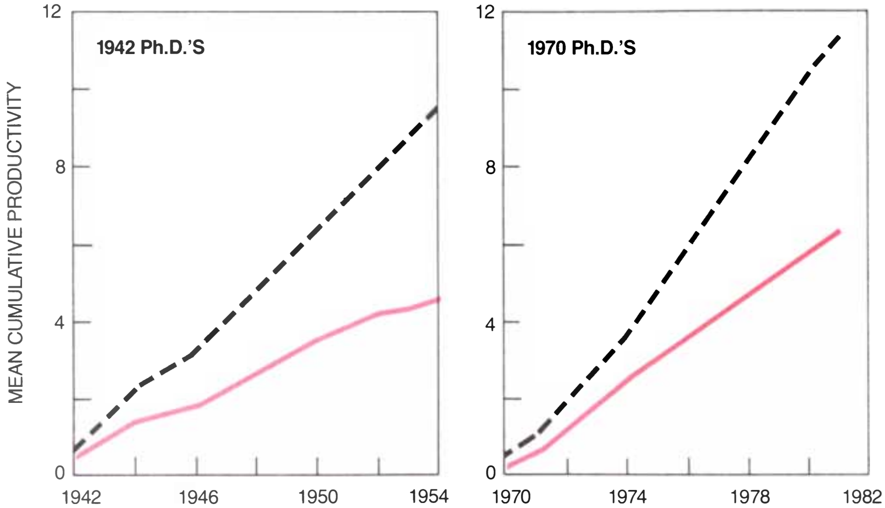
Our subjects were asked about their research and publication histories and for comments on graphs we prepared that showed the number of papers they had published each year along with important events in their career and personal life.

Both men and women scientists reported having come up against the be-



NUMBER OF Ph.D.'S awarded to women has increased since 1970 in spite of a general decline in the number awarded to men. The curves (*top*) show the annual number of U.S. doctoral degrees from 1970 through 1985. The percent awarded to women has risen sharply (*bottom*), but it remains low in the physical sciences and engineering. The data are from National Research Council studies.

lief that marriage and motherhood cannot be meshed with a demanding scientific career. Not surprisingly, the oldest group of scientists encountered this belief most often. Before (and even shortly after) World War II, the proper priorities for women were widely held to be marriage and motherhood first and science second, and good science was believed to be all-consuming. The notion that women could simultaneously be traditional wives, traditional mothers and productive scientists seemed to be patently absurd.



MEN SCIENTISTS publish more papers (*black curve*) than women scientists do (*color*), and the disparity increases with time. Data from large Ph.D. samples reported in earlier studies by the authors suggest that the disparity has lessened somewhat since the 1940's.

Such beliefs were articulated by a woman zoologist who recalled that "if one had children and a working husband, it was not part of the psychology to suppose that one's job was anything more than a secondary consideration." Many men and women scientists at the time shared these views. They believed most women could no longer be serious scientists once they were married. A distinguished woman biologist who is now in her seventies said her woman laboratory chief had been appalled that her protégée would marry: "She threw me out of the lab the minute she heard I was going to get married because that was treason against women." For one male chemist, marriage then meant that women scientists were "finished"; a male physicist said, "As soon as women got into domestic life, that was the end of it for all of them."

This climate of opinion meant that women determined to have serious research careers often did not marry. In the words of one biologist now in her seventies, "marrying was not considered the thing to do [for women scientists]. In science, you're dedicated. You go into a shroud, you don't wear normal clothes...you shouldn't get married; you shouldn't have children."

Not all women scientists accepted those views, of course, and some did marry, have children and continue to work. Yet until the end of World War II at the earliest, women scientists were few in number, and fewer than half of them were married. Married women with children were nearly invisible in American science. Such women were viewed as violating the prevailing family norms.

Although social attitudes concerning the roles of wives and mothers have changed significantly, even the youngest women scientists report that many people still consider marriage and motherhood to be incompatible with a scientific career. One young chemist says, "When I got pregnant, I was written off as a serious scientist...by lots of people." When those occupying positions of power and authority act in accordance with these beliefs, they severely limit the opportunities and careers available to married women.

To assess the actual impact of marriage and motherhood on women in science we needed answers to four questions: Are married women as a group less productive researchers than single women? Among married women, do those who have children publish fewer papers than those who are childless? Is there a drop in women's published research performance after childbirth? Does the number of children a woman scientist has affect her research performance?

The publication and career histories of eminent scientists provided the first clues that marriage and children do not generally affect scientific productivity. On the average these eminent married women (and eminent women are just as likely to marry and have children as their rank-and-file counterparts) publish slightly more over their careers—not less—than eminent single women: an average of 3.0 papers per year compared with 2.2. Among the eminent married women, those with children publish 2.9 papers annually and childless women publish 3.3. Moreover, during the three-year periods preceding and following the birth of first children the annual published productivity of these women does not fall but actually rises from 1.5 to 2.7 papers. Finally, the rate of publication of these scientists is unrelated to the number of children they have.

These statistical findings are plainly counterintuitive, and yet they are consistent with earlier cross-sectional studies. How well do they correspond with the subjective reports of women in the interviews? Do women think marrying and having children really is unrelated to the amount of research they publish and, if they do, how is one to account for that belief?

The publication histories of two older eminent women scientists, one with four children and the other with three, illustrate one general pattern [*see illustration on next page*]. Such scientists in general published less when they were young and had young children; there is a marked upward trend in the number of papers published after the first decade of their careers. There are also year-to-year fluctuations—peaks and valleys—within the general upward trend. All these women do, to be sure, acknowledge that children take up a great deal of time. They are "a definite time commitment. That means you are doing less with other things"—but not less scientific research. The research continues.

How can it continue? These eminent women emphasize, first, that thinking about science goes on at home as well as at work. It does not end when they close the doors to their laboratories. "When the kids were small...I had ideas when I was washing the dishes and nursing the babies. Scientifically speaking, I did my best work during the period when the kids were coming." Second, if they have a scientist husband (and that is the typical situation), they talk about their research during "so-called off-time." Third, professional obligations other than research are far more limited for younger scientists than they are for older ones. "I spent more time doing science then than I do now.... The calls on my time [then] were my job and my kids. Now it's so many other things." Fourth, lower rates of publication in the early years are not necessarily attributable to the demands of motherhood but rather are characteristic of the beginning phase of a developing research program. As one physical scientist observed, "In the first years...we were building those enormous instruments. Developing the theory and experiment [took up] a lot of time, so there weren't that many papers." Another, commenting on the downturns in her own publication graph, said: "You are very busy in the valley."

According to these eminent scientists, marriage and motherhood did not reduce their published productivity. Should one believe their retrospective accounts? Perhaps their perceptions were not correct. After all, inspection of the graphs of older eminent married women with children shows that they indeed were less productive scientifically when they were young and had young children. Did having young children in truth affect their rate of publication, at least in the short run?

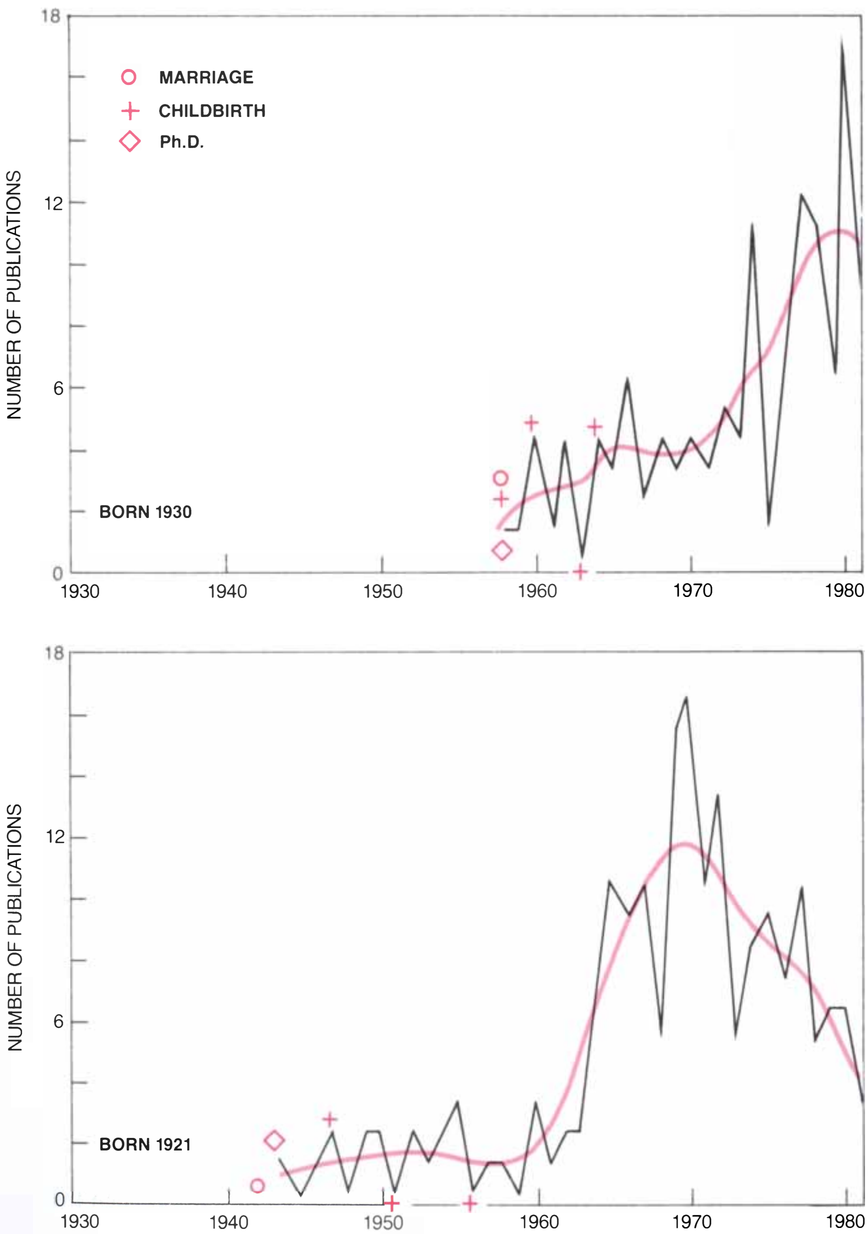
We thought the publication patterns of two groups of scientists who should not be affected by marriage or parenthood—eminent single women and eminent men whose wives took responsibility for looking after children—might further illuminate these first counterintuitive results. Do the presumably unencumbered scientists publish at a more rapid rate in the early years than women who had young children? The answer appears to be no [*see illustration on opposite page*]. Single women and married men are just as likely to show a low level of publication in the first decade of their careers. They are also as likely to show oscillations and an overall rising slope of publication with time. The fact that the early publication patterns of these two groups do not differ much from those of married women who have children lends credibility to the married women's observations.

Another question comes up, however. The eminent older women say that marriage and childbearing did not reduce their scientific productivity. If that is so, why do their publication rates increase as they pass beyond the child-care age—particularly in view of the additional distractions and responsibilities they say come with professional maturity and a higher degree of recognition? Part of the answer is that the opportunities for collaborative research increase as one's career progresses. Beginning scientists do most or all of the benchwork themselves; more established ones often assume major administrative roles and oversee the work that goes on in their laboratories. Their publication records reflect the resulting upsurge in collaborative research.

We should mention that the publication record of some eminent women scientists does not exhibit the typical rising slope with time [*see illustration on page 124*]. These women too say they think marriage and motherhood had little bearing on their scientific productivity.

Our data seem to indicate, then, that older eminent women with children generally publish as much early in their careers as their unmarried counterparts do. Could we have made a critical error, however, in compiling and interpreting the data? Could it be that women who have children and yet remain scientifically productive are "self-selected," that is, are simply more talented scientists than those who choose to remain childless?

Although strict comparisons of sci-



EMINENT WOMEN SCIENTISTS, during the first three decades of their career, typically show a general upward trend in the number of papers published annually. The publication histories of two married women, one with four children (*top*) and the other with three children (*bottom*), conform to this pattern; they show no lasting negative effects of marriage or children. The jagged black curve indicates the number of papers published in each year; the smooth curve (*light color*) is calculated to indicate the general trend.

entific ability cannot be made, we can compare the publication rates of older eminent scientists who did and did not have children, focusing on the years before motherhood. To this end we matched the two groups of women roughly by their birth dates. The publication rate during the three years before women with children had their first child was compared with "equivalent years" in the life span of women without children. We found similar early histories: approximately 1.3 papers annually for women who subsequently had children and 1.6 for those who did not. In other words, older eminent women who eventually had children published inconsequentially fewer papers initially than women who never had children.

More important, might we be mistaken in concentrating on the histories of eminent women scientists instead of on those who might be more likely to experience the debilitating effects of marriage and motherhood on publication rates? The eminent women, after all, are very successful scientists; if marriage and motherhood had taken a toll in their case, such women would presumably not have been able to achieve the recognition they actually did achieve. Do the publication histories of other women scientists, those we have designated as rank-and-file, testify to the negative impact of marriage and child care?

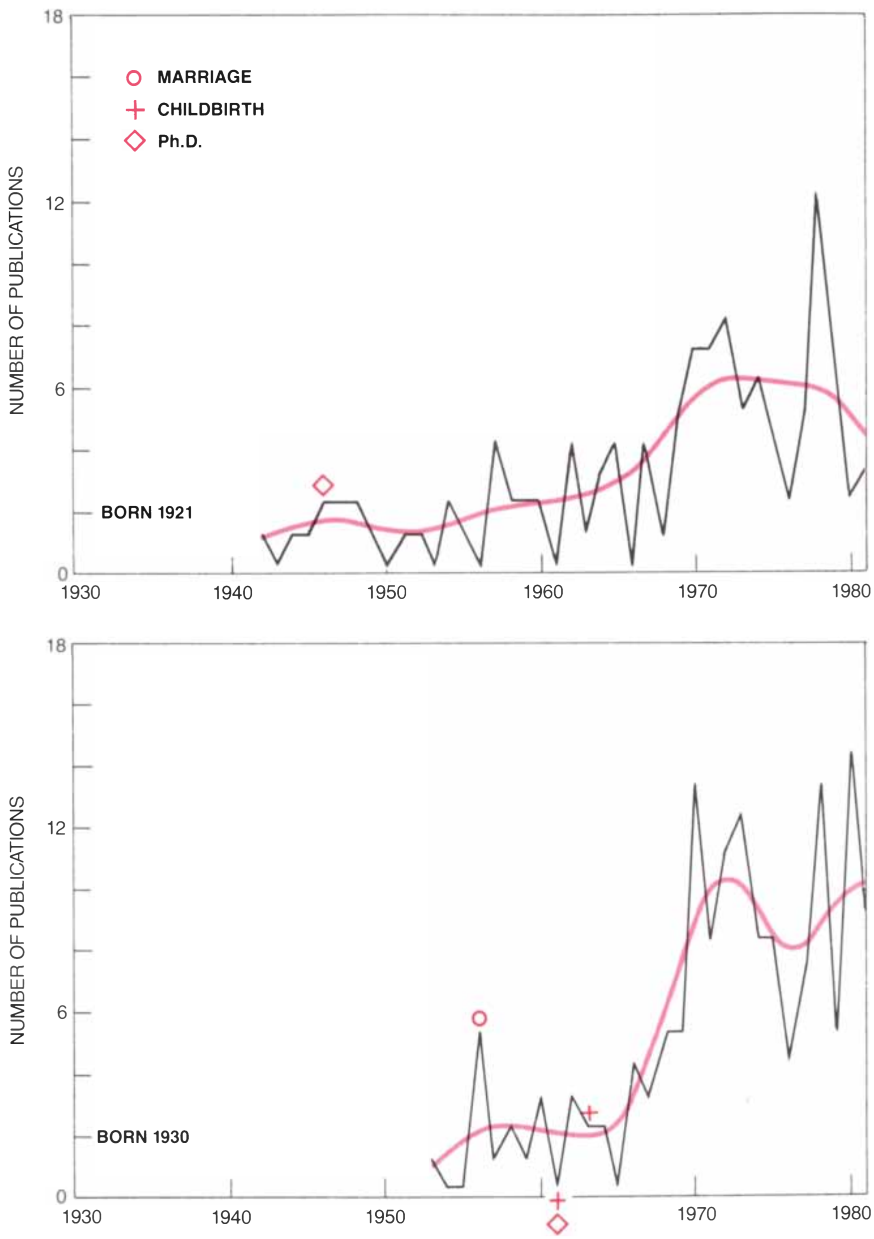
Rank-and-file men and women scientists do, of course, in general publish less than their eminent colleagues. Within the rank and file in our sample, married women did publish slightly fewer papers than single ones (an average of 1.1 a year compared with 1.7). But married women with children publish no fewer papers than married women without children; both groups average about one paper per year. As in the case of eminent women, their publication rates did not decline after children were born. Rank-and-file women averaged well under one paper per year (.2) in the three-year period prior to the birth of their first child and they averaged just under one paper per year (.8) in the three years following the birth.

Much the same impression is conveyed by these scientists' own testimony: having children did not significantly affect their records of research and publication. As one relatively unproductive behavioral scientist observed, "It didn't occur to me to stop working when I had [a child].... In fact, right after she was born I wrote one paper and started to work on the next one,...so if anything...I seemed to work better (which is to say more efficiently albeit under greater pressure)."

Her account is consistent with accounts of other women, such as a biochemist who asserted that her publication rate was not affected by family obligations. "It's just fortuitous.... The kinetics of me as a parent and me as a researcher don't have a direct relationship.... One has not interfered with the other." Contrary to expectation, then, such women are no more likely than older eminent ones to say that marriage and family responsibilities account for their rate of publication, and the statistical data we have in hand bear out what they say.

Would we find a similar pattern in data for younger women? Because marriage and childbearing usually come early in a woman's career, the records for younger scientists should show their effects on published productivity, at least in the short term.

A behavioral scientist who is now a full professor in a high-ranking department, and who had just had a child, suggested that motherhood was unrelated to her pace of publication. "Hav-

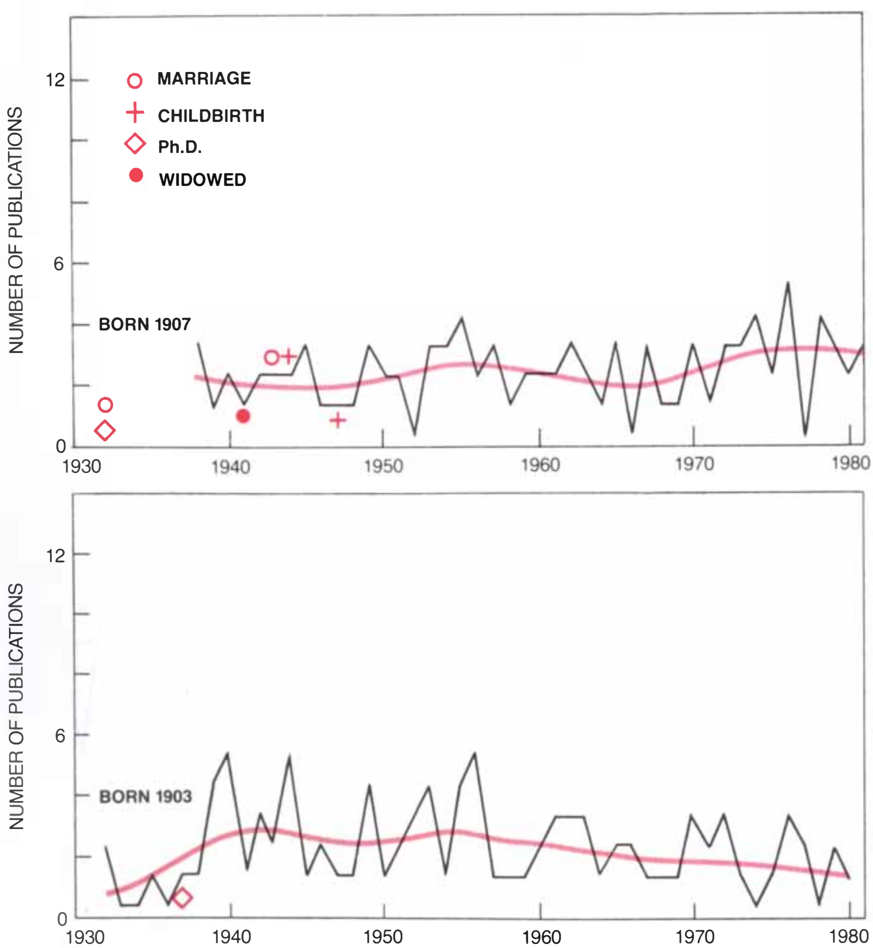


LOWER RATES OF PUBLICATION in the early part of a career are characteristic of both married men and single women. The publication profile of a distinguished woman biologist (*top*) who never married shows the same pattern of oscillations and an overall increase as the graphs of women who married and had children. The same pattern can be seen in the profile of an eminent male chemist (*bottom*). He published at a much slower pace when his children were young, although his domestic responsibilities were minimal.

ing a child is draining in many ways, but not in terms of having affected my work, especially when I look at how much I've...done this year. No, it's really movies, social life and things like that [that go].... I feel chronically slow and behind and this year I'm blaming it on the baby, but I realize that has nothing to do with it." In the past it was "too many graduate students,...[a] grant reviewing committee, [an] editorship [and now a] baby. So I have always had some kind of baby to blame it on."

Perhaps the limiting case is a woman who has an endowed chair in a major department. She has been married four times and divorced three times, and she has had four children by three different husbands. If marriage and motherhood should bring a scientist's career to a halt, they should have done so here. Yet her published output has risen throughout her complex history. Ironically, the largest dip in her pattern came in 1979, one of the few years in which she did not get married, have a child or get divorced [*see illustration on opposite page*]. Asked about her publication pattern, she replied, "Suddenly you're ready to report on three different projects and therefore [the papers] roll off the presses.... The ups and downs [have] nothing to do with the rest of my life."

Even so, pregnancy and its aftermath did interfere temporarily with research in the case of three of the 37 women in our sample who had children. A woman biologist told us, "I was one of those women who said I'm so well organized, I'll just drop a child and that will be it—but I didn't realize that hormones could do such a job on a person." Asked to explain why her publication rate declined only temporarily and not by much, she said, "I was lucky because by then I had people in my lab. They were working [and] productive.... But I found that for a whole year my mind wasn't functioning."



SOME SCIENTISTS publish at a rather constant rate throughout their career. These are profiles for two eminent women: one who married twice and had two children (*top*) and another who never married (*bottom*). Although the annual number of publications fluctuates, the mean number in five-year intervals remains much the same for these scientists. This pattern appears as often among married women as among unmarried ones.

These longitudinal data indicate that marriage and children are not inimical to the published productivity of women in the aggregate. Although few people would question the fact that marriage and motherhood impose formidable responsibilities, apparently many women scientists can manage a career and family obligations simultaneously. How do they do it? And how, when conflicts arise between home and career, do women scientists deal with them?

The answer can be found in part in how women scientists manage their "status set," that is, the array of social positions each of them occupies (such as professor, laboratory director, wife, mother and citizen). We focused on three interconnected aspects of status sets: size (the number of positions held simultaneously), congruence (the extent to which various obligations are consistent rather than in conflict) and the timing of the addition and deletion of status obligations.

At the extreme, several women, convinced that marriage is incompatible with scientific work, opted not to marry, in effect limiting their status sets. Yet three-fourths of the women in our sample did marry—a proportion that appears to be typical of women scientists in general now. For the majority of them the fundamental question about marriage was one of timing. As a young economist said, it "would also be a big career disadvantage for me [to get married]. Once I have tenure and am more settled down in a university, it would be a little bit easier."

Two-thirds of these married women had children. Timing their arrival, many women say, helps one to maintain a research career. A renowned physical scientist delayed having her first child for nine years after marrying in order to prove herself as a professional. Many younger women said they were delaying motherhood until they receive tenure. Doubting that she could have a child and maintain the level of performance needed for tenure, a young biochemist noted, "My ideal scenario is to get a tenured position and then have a child or two."

All told, eminent and rank-and-file women had about the same number of children: an average of two, with none exceeding four. Our data show that annual rates of publication are virtually the same for women who have one child and for those with two or more.

There are aspects of marriage and motherhood other than timing that can make for congruent status sets. Close to four-fifths of the married women we interviewed were married to scientists (again, a proportion typical of married women scientists in general). Such "assortative," or selective, mating apparently gives these women (and men too) a variety of benefits, including ready understanding of their professional obligations and way of life. A molecular biologist observed that her husband could hardly be upset when she came home late because "he knew that no matter how well I had planned something, experiments do get delayed. I think that's made it a lot easier." Women scientists married to scientists publish, on the average, 40 percent more than women who are married to men in other occupations. The difference in publication rate may result from self-selection, congruence of values or the flexibility of academic schedules.

Women scientists can also achieve congruence of status obligations by compartmentalizing their lives, but they report that compartmentalizing is not always feasible. In fact, many find it harder to keep their mind off work when they are at home than it is to keep their mind off children when they are at work. Moreover, every woman with children emphasized that she relied on some form of child care or household help—necessary arrangements, but fragile at best. The illness of a spouse, child or housekeeper can throw the entire scheme awry.

In view of these difficulties, is it possible that women and men scientists manage to continue their research only by neglecting their spouses and children? Our study was not designed to answer that question. This much we do know: The divorce rate for both women and men is unrelated to published productivity.

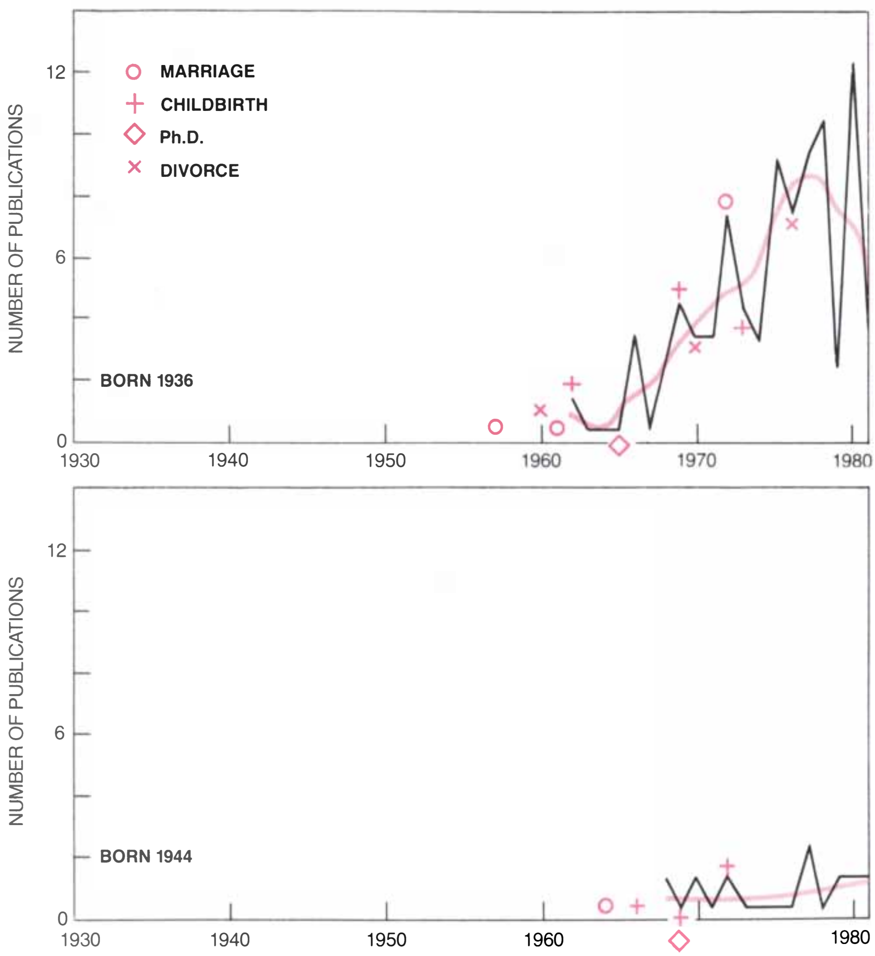
Married women scientists with children do pay a price to remain scientifically productive. They report having had to eliminate almost everything but work and family, particularly when their children were young. As an eminent psychologist observed, what goes first is "discretionary time. I think I can only work effectively... 50 hours a week.... If I didn't have children, I'd probably read more novels...or go to more movies."

Loss of discretionary time not only affects leisure pursuits but also sometimes has serious consequences for women's research and careers, even if it has no significant effect on their rate of publication. Women scientists who adhere to rigid family schedules say they have lost the flexibility to stay late in the laboratory to work on an interesting problem; they report not feeling part of "the club," not having time for informal discussions with colleagues.

Other investigators have shown that only some 12 percent of women scientists stop work after getting their Ph.D. Surely some of them do so because of the intense conflicts that arise between science and parenthood. One woman who had left a promising research career for an administrative job said in a supplementary interview, "I was only in the lab [for] the hours that the children were in school.... I was working with really bright people who ground out the publications at a rate I couldn't keep up with.... It was too frustrating." A small subset of women, then, find that science and motherhood do not mix and alter their careers to give more time to their families.

Our study shows, however, that for most of these women science and motherhood do mix. Women scientists who marry and have families publish as many papers per year, on the average, as single women. Managing the simultaneous demands of research careers, marriage and motherhood is not easy; it requires organization and an elaborate set of personal adaptations.

The results of this study should not be interpreted as meaning that marriage and children have no effect on the careers of women scientists. They do, but they generally do not take their toll on women's research performance. How then can the persistent disparity in rate of publication between men and women scientists be explained? Why do men publish substantially more papers over the course of their careers than women with comparable backgrounds? The difference is evidently not explained by marriage and motherhood. It remains a puzzle requiring further comparative inquiry into the research careers of men and women scientists.



PUBLISHED PRODUCTIVITY is not related to family obligations, as is indicated by the publication histories of two women with strikingly different career patterns. One (*top*) is an eminent behavioral scientist whose productivity rose steadily in spite of the fact that she has been married four times (most recently in 1982) and divorced three times and has four children. The other (*bottom*) is an associate professor of chemistry and has three children. She attributes her low productivity to constraints other than domestic obligations. Her publication history is typical of many rank-and-file women scientists.

The Computer as a Musical Instrument

If a computer generates the right sequence of numbers, any sound—including some never before heard—can be produced. Because of its versatility, such digital sound synthesis has found a place in music

by Max V. Mathews and John R. Pierce

The French composer and conductor Pierre Boulez startled American concertgoers last year when he toured the country with an ensemble that included a computer and an array of electronic sound-modifying devices. Boulez' works for orchestra and computer-generated sound integrate digital electronic equipment in an active, sound-producing role that goes far beyond the more established role of such equipment in the recording and playback of symphonic music. Although computers are not yet a part of every symphony orchestra, digital sound synthesizers are already serving as alternatives to traditional instruments in the production of sound tracks for film and television, and they are quickly becoming the instruments of choice in popular music.

Behind all these efforts to incorporate the digital synthesis of complex sounds into music lies some pioneering work on the computer processing of sound, with which we were privileged to be involved, that began in the 1950's at the Bell Telephone Laboratories. We were originally drawn to the computer as a sound-analyzing and sound-producing device while investigating the factors that contribute to the efficient transmission of speech through telephone lines. It soon became clear to us that the quality of sound is of great importance not only in speech but also in music, and we enthusiastically began to study the production of musical sounds.

Although our first attempts to produce musical sounds from a computer were disappointing, the electronic instruments and computer programs that eventually evolved from these initial attempts are now sophisticated enough to have significant impact in several music-related areas. First, the instruments are of considerable commercial importance to the music-recording industry since they can efficiently generate music or sound effects that can be readily synchronized with action depicted in a motion picture or on television. Second, they provide a virtually limitless universe of sounds through which composers and performers can express their thoughts and feelings. Finally, and perhaps most important, they can deepen understanding of the particular patterns of sound called "music."

Reduced to its essential physical nature, sound is no more than a pressure fluctuation in the air. As such it can be expressed graphically by means of a waveform: a plot of how the ambient air pressure varies as a function of time. How sound is perceived, that is, whether it is pleasant or unpleasant, depends specifically on the way the various features of the pressure fluctuations are translated into nerve impulses in the ear and on the way the nerve impulses are then interpreted subjectively by the brain.

Sounds that are heard as having a definite pitch have waveforms that exhibit a nearly periodic variation in pressure. The pitch of a sound corresponds directly to the variation's frequency of repetition. For example, a pressure variation that repeats itself 440 times per second is perceived by a musician as a tone with a definite pitch: A above middle C.

Most sounds one hears originate from the vibrations induced in common, everyday objects: human vocal cords through which air is exhaled, violin strings that are bowed and automobiles that collide—to name a few. Sounds can also be generated from the vibrations set up in a loudspeaker by varying the voltage of its electrical input. Indeed, as any audiophile knows, an excellent reproduction of a given sound can be elicited from a good loudspeaker if an accurate voltage analogue of the sound's pressure function is applied to the speaker.

A mathematical verity known as the sampling theorem states that any waveform made up of multiple components of various frequencies can be exactly described by a sequence of numbers that give the value of the waveform's amplitude at a rate determined by the waveform's bandwidth, or range of component frequencies (conventionally expressed in hertz, or cycles per second). Specifically, the rate at which the numbers must be generated is equal to twice the bandwidth of the waveform. It is this theorem, proved by Claude E. Shannon of Bell Laboratories in 1948, that underlies all digital recording, processing and generation of sound.

The sampling theorem implies that one second of a sound whose bandwidth is 20,000 hertz (which spans the range of frequencies audible to the human ear) can be exactly recorded if 40,000 numbers, called samples, corresponding to the evenly spaced, instantaneous values of the sound wave's pressure amplitude (or the voltage analogue thereof) are collected during that one second. Conversely, if the appropriate 40,000 sample values per second can be summoned, any perceptible sound could conceivably be produced in all its acoustic intricacy. The compact disc, on which about 40,000 samples per second of sound are encoded and stored as points of varying reflectance, is an example of such a storage-and-retrieval system.

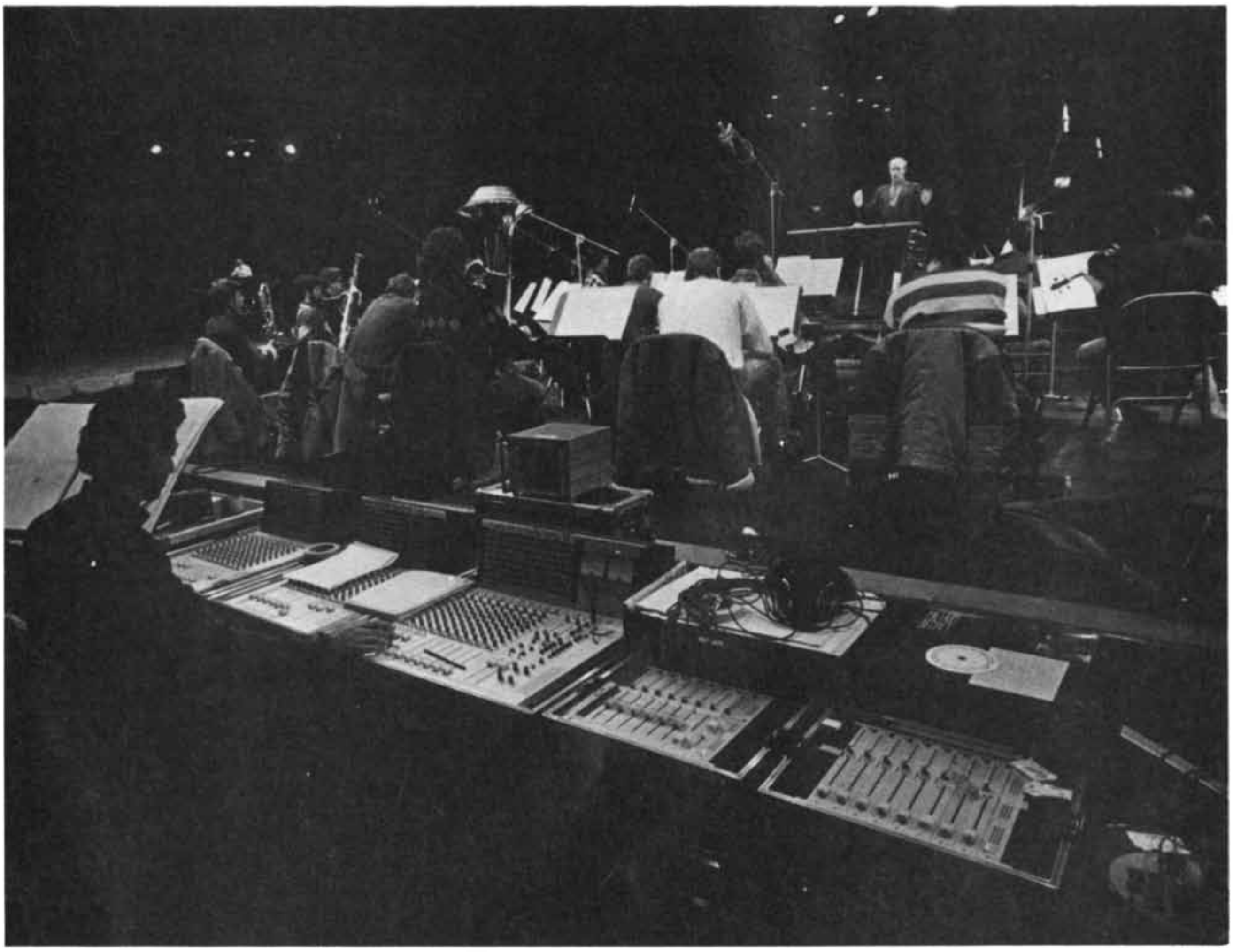
Another way to store and retrieve such huge quantities of numbers for sound-generation purposes is afforded by digital microprocessors, such as those found in computers. Converting numbers in a computer into voltages—an essential step in digital sound processing—can be accomplished easily by means of analog-to-digital converters, which translate an electrical signal into a sequence of numbers proportional to the signal's voltage, and by digital-to-analog converters, which perform the reverse process. The succession of discrete voltage pulses that a digital-to-analog converter produces from a sequence of amplitude samples is generally "smoothed" into a continuous waveform by a special filter before the electrical signal is amplified and broadcast through a loudspeaker system.

For practical reasons both computers and compact discs express amplitude samples in terms of binary numbers. The current standard is to specify a sample by means of 16 bits, or binary digits. This allows the amplitude of a sound waveform to be divided into 65,536 discrete levels. (In practice, half of these binary numbers are used to represent positive amplitude samples and half are used to represent negative amplitude samples.) This range of amplitude levels is not enough to achieve a reproduction of either rock or symphonic concerts that is absolutely free of perceptible noise, but the resultant sound fidelity is certainly very much better than analog recordings on standard phonograph discs or magnetic tapes.

Our initial efforts to apply the sampling theorem to generate sound by means of a computer were met with an unfortunate surprise: we seemed to be able to synthesize only unmusical sounds rather than beautiful musical sounds. Those early efforts consisted primarily in converting patterns of numbers representing simple waveforms (such as sinusoid or sawtooth) into sounds, and the sounds we thus produced tended to be either bland or buzzy and "electronic."

The problem was not that the computer we worked with was inherently limited in the sound waves it could produce (although there were some practical constraints in terms of money and time). The problem was rather that no one knew what constituted a waveform the human ear would perceive as a beautiful musical sound. A sound's characteristic quality is called its timbre, or tone "color." Although we were able to generate sounds of a given pitch and loudness from the computer, sounds of a pleasing timbre proved to be difficult to generate.

The modest amount of available literature on the physics of sounds of traditional instruments was not of much help. It turned out to be not only incomplete but in many cases blatantly wrong. For example, although the literature emphasized the description of a musical tone in terms of its steady state, or the middle part of its waveform, it soon became clear to us that the waveform's beginning (called the attack) and end (called the decay) were



DIGITAL SOUND PROCESSING plays an active, integral part along with the instrumentalists of the Ensemble InterContemporain in a rehearsal performance of Pierre Boulez' composition *Répons*. Technicians (*foreground*) control the real-time, or instantaneous, processing of sound from an electronic work station, taking their cues from Boulez himself (*background*), who is conducting.

more important. The contour of a waveform, drawn by connecting crest to crest and trough to trough, is called its envelope. An envelope that has an abrupt, steep attack followed by a gently sloped decay results in a sound of plucked or struck timbre, regardless of the waveform's steady-state fluctuations within its envelope.

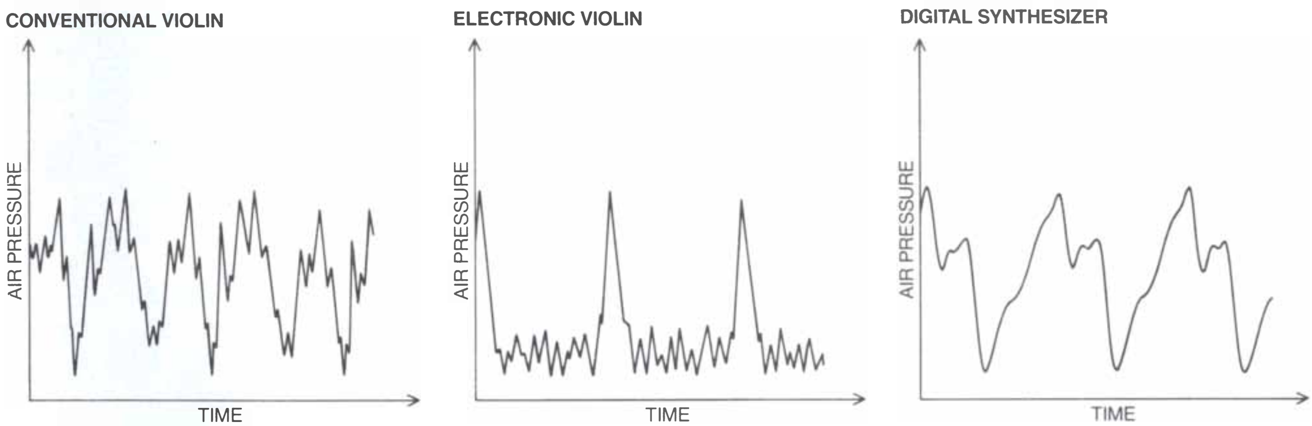
It was also thought that knowledge of the relative amplitudes of the various frequency components of a musical tone's steady state was sufficient to characterize the tone's timbre. A frequency spectrum of a musical tone shows that its waveform contains not just one component at the frequency corresponding to the pitch (called the fundamental frequency) but several components whose frequencies are generally, although not always, whole-number multiples of the fundamental frequency. These components are called partial tones, or partials for short, and can be numerically ranked according to their frequency. Hence the fundamental frequency is the first partial, the component with the next-higher frequency is the second partial and so on.

The traditional view of timbre, for instance, held that the clarinet's sound was completely determined by a spectrum in which the frequencies of all partials were odd-number multiples of the fundamental frequency. Although this property of the clarinet's frequency spectrum is to a certain extent true, we found that it is not enough to characterize the instrument's timbre.

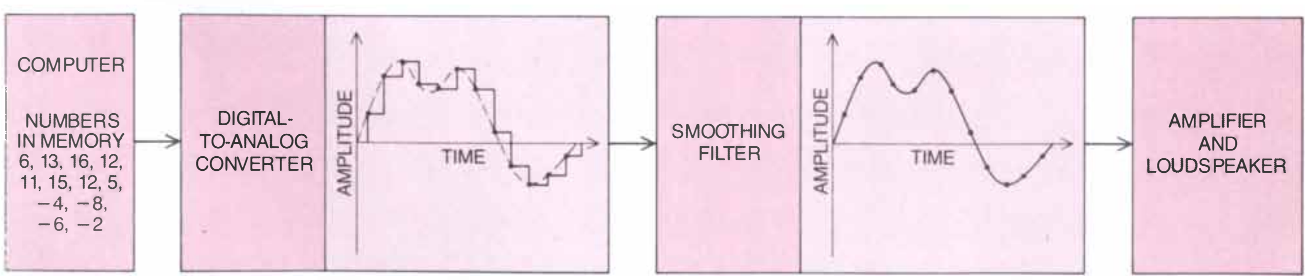
It has taken much detailed study to understand the timbral aspects of sound and to dispel some of the antiquated, misleading generalities. Fortunately the computer itself proved to be a powerful tool for studying musical timbres. A striking demonstration of how to analyze and synthesize a sound of good timbre is the early work done at Bell Laboratories in 1965 by the French composer and physicist Jean-Claude Risset.

Risset sought to synthesize good brass-instrument tones based on what he had read in the existing literature about the relative amplitudes of a trumpet's partials. He programmed a computer to generate numbers that would correspond to the amplitude samples of what he thought would be a trumpetlike waveform. When he converted the numbers into sound, however, he found that the resultant sound was not at all like that of a trumpet.

Risset then recorded real trumpet tones and analyzed their spectra with a computer. He found that a trumpet spectrum changed during the playing of the tone and that the high-frequency partials had much larger amplitudes in the middle of the tone than in the attack and decay parts of the tone. By synthesizing a sound whose high-frequency partials build up slowly during the sound's attack and reach their maximum amplitude during the tone's steady state, he succeeded in producing tones the average listener could not distinguish from recorded trumpet tones.



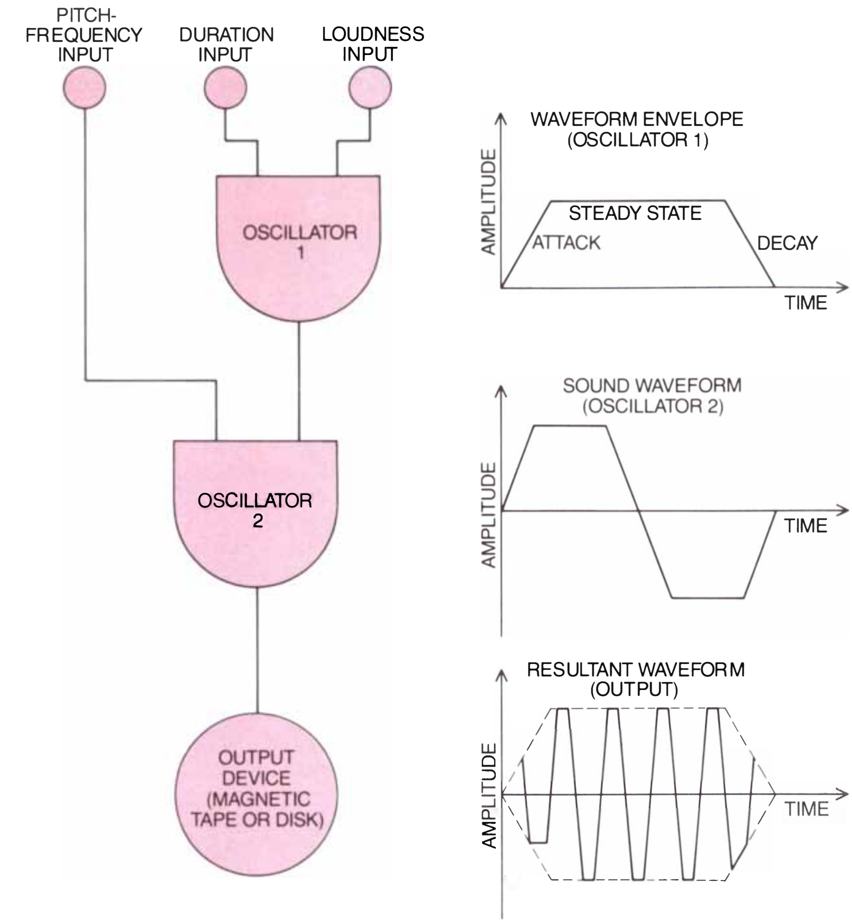
WAVEFORMS show the way a particular sound causes the ambient air pressure to fluctuate. The sound from a conventional violin (*left*) is characterized by a complex periodic waveform. An electronic violin (*middle*) built by one of the authors (Mathews) converts the motion of a bowed metal string into an electrical signal that is then filtered to yield a simple violinlike waveform. A popular digital sound synthesizer (*right*) can mimic the waveform of an actual violin sound more closely than the electronic violin can.



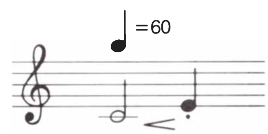
SAMPLING THEOREM, which posits that any waveform made up of components of various frequencies can be exactly described by a sequence of numbers, underlies all digital processing of sound. The numbers, called samples, are proportional to the instantaneous amplitude of the waveform, and the minimum sampling rate is twice the waveform's bandwidth (the range of component frequencies). Samples stored in a computer's memory can be converted into voltages proportional to each sample's value. The discrete voltages can then be "smoothed" into a continuous signal that is amplified and sent to a loudspeaker, where it is converted into sound. Sound synthesis involves programming a computer to produce the appropriate sequence of numbers for a given sound.

In brief, Risset discovered that the steady-state frequency spectrum is not an adequate description of a timbre. In order to synthesize a musical sound of a particular timbre one must know how the sound's frequency spectrum changes when a tone is played, that is, how the various partials build up at the beginning of the tone and how they die away at the end. Moreover, he found that different partials followed different time courses and that these were critically important to the listener. Hence a steady-state frequency spectrum shaped by a single envelope rarely suffices to reproduce the sound of conventional instruments.

Risset's work followed a methodology known as analysis by synthesis, which has led to great progress in understanding the timbre of sounds from traditional instruments. The analysis-by-synthesis methodology begins with an analysis of a sound—often by means of a computer, which can break down the sound into individual frequency components and determine each component's envelope. Such analysis usually gives a plethora of detail, from which the experimenter must select those features he believes are important in producing the sound's characteristic timbre. The investigator then formulates a hypothesis that gives a simple, physical description of the sound and tests the hypothesis by synthesizing a sound based on it. The hypothesis is evaluated by comparing the synthesized sound with the original sound. If the two are indistinguishable to the listener, the experimenter has succeeded in finding those details that effectively characterize a particular class of sounds.



STARTING TIME (SECONDS)	DURATION (SECONDS)	LOUDNESS (ARBITRARY UNITS)	PITCH FREQUENCY (HERTZ)
0.0	1.5	1	262
2.0	.25	5	325



COMPUTER "INSTRUMENT" is constructed from so-called unit generators in the sound-synthesis program Music V, written by one of the authors (Mathews). Unit generators are subprograms whose numerical inputs and outputs can be interconnected. The most important unit generator is the oscillator. Every time an oscillator is cycled it generates a series of numbers that correspond to a preselected waveform. The output waveform's amplitude and the frequency of the waveform-generating cycles are determined by the oscillator's two inputs. The amplitude input of a pitch-determining oscillator often is the output of another oscillator that controls the sound's envelope. The envelope determines the sound's attack (how quickly it builds up), its steady state (its middle part) and its decay (how quickly it fades away). An instrument thus constructed is "played" by means of note lists (*bottom left*): computer instructions that specify essentially the same information that a note on a musical staff (*bottom right*) conveys to the performer.

In order to synthesize sounds, whether for analysis or for musical purposes, one must have computer software that can efficiently generate the sequence of binary numbers representing the successive samples of a waveform. The software has to be simple to work with and yet sophisticated enough to allow the incorporation of virtually any sound-timbre hypothesis into a waveform.

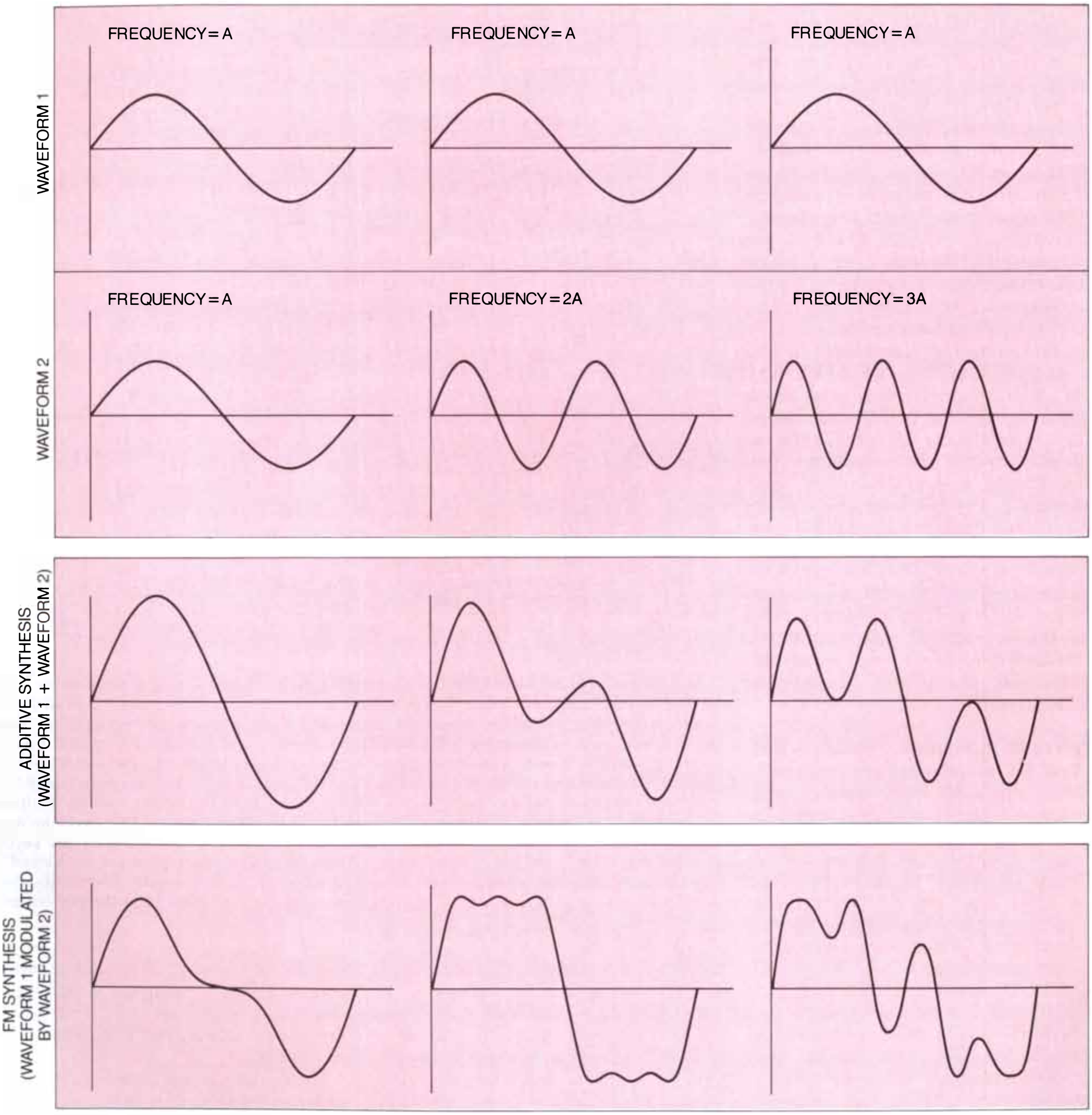
During the late 1950's and early 1960's one of us (Mathews) wrote such software. One of these computer programs, Music V, contains the basic concepts that are still to be found in the software employed in computer-music centers throughout the world. These include a number of simple program building blocks, called unit generators, to give the musician flexibility in creating unique instrumental sounds; stored tables of numbers to generate efficiently certain waveforms on command, and lists of notes (instead of the conventional music notation) to specify what is to be played.

A composer can construct his own musical "instruments" by connecting unit generators in a variety of ways. In some respects unit generators are computer simulations of the electronic devices in early analog synthesizers, which the user interconnected with patch cords. Whereas the modular devices in an analog synthesizer manipulate electric voltages, the unit generators of Music V manipulate numbers. The most important unit generators are the oscillator, the adder and the multiplier.

Every time an oscillator is cycled it generates a series of numbers that corresponds to a waveform drawn from a catalogue of possible waveforms stored in sampled form in the computer memory. Such waveforms can take on a variety of shapes: sinusoidal, square and sawtooth, among others. Stored waveforms greatly increase the efficiency of an oscillator because the computer need only look up samples of the waveform rather than having to calculate them. In addition the stored waveform determines some aspects of timbre.

Two input variables control the oscillator: one determines the amplitude of its output and the other determines the frequency of waveform-generating cycles. Both inputs can change with time, allowing a rise or fall in amplitude and frequency. The amplitude input of a sound-producing oscillator is often the output of another oscillator that functions as an envelope generator to control the sound's attack and decay. As we mentioned above, attack and decay functions also have an important influence on the timbre of a computer-generated sound.

The adder and multiplier unit generators, as their names imply, respectively calculate the sum and the product of two input numbers. The adder can sum the outputs of two sinusoidal oscillators that have been "tuned" to be the instrument's partials. It can also add to the input that specifies the frequency of an oscillator a small, regularly changing number. In this way a vibrato effect can be simulated. The



TWO COMMON METHODS of generating complex musical waveforms are additive synthesis and frequency-modulation (FM) synthesis. In additive synthesis waveforms are summed to produce a musical waveform. The frequencies of the summed waveforms are generally whole-number multiples of the pitch-determining frequency, corresponding to a musical sound's multiple frequency components, called partials. In FM synthesis the frequency of a waveform, called the carrier, is varied according to the amplitude of another waveform, called the modulator. The resulting waveforms tend to be more complicated and produce richer tones. Still more complex tones can be achieved by varying the relative amplitudes and phases of the carrier and modulator during the course of a tone. Since FM synthesis requires fewer waveforms to produce richer musical tones, it is more popular than additive synthesis.

multiplier can be used in several ways: as a volume control (by multiplying a given factor with an oscillator's amplitude-setting input) or as a pitch transposer (by multiplying a given factor with an oscillator's frequency-setting input). Unit generators that supply random numbers are also frequently used to make both noiselike sounds and small random fluctuations in the pitch and amplitude of generated tones in order to make them sound livelier and less machinelike.

The composer "creates" an instrument at the start of a Music V program by selecting a set of unit generators and specifying the interconnections among all their numerical inputs and outputs. At least one output must go to a digital-to-analog converter, which converts the binary samples into a form that can then be played through a loudspeaker. Designing an instrument is a highly creative function that is available to a composer of computer music; in contrast, a composer who relies on traditional instruments to interpret his work can expect only traditional sounds. But how does one "compose" a piece of music using such a program? It is done by means of note lists.

A note list is a computer instruction that specifies essentially the same information a note on the staff of a musical score conveys to the performer. It specifies when a note is to be played, its duration, what instrument it is to be played on, its pitch and its loudness. In addition special information that regulates the timbre of the instrument is often included. Note lists do not look like notes on musical staffs, of course; they are letters and numbers that the computer interprets as input for the sample-producing "instruments" made up of unit generators.

Several general strategies can be employed to synthesize complex sounds. Risset's work, for example, made use of a technique called additive synthesis or summation of partials. In this technique the individual partials of a given sound are synthesized separately, allowing each to have its own independent frequency and envelope, before they are added together to achieve a synthesized version of the sound. Hence by means of additive synthesis the slightly inharmonic partials (partials whose frequencies are not whole-number multiples of the fundamental frequency) of the piano or the prominent inharmonic partials of bells and drums can be closely duplicated. Additive synthesis is the most general way of synthesizing timbres.

Although additive synthesis is powerful, it is also expensive and slow. Timbres have many partials, and if each one is generated separately, a great deal of computation is required. Also, if each partial follows a different course, a great deal of control information may be needed to generate their envelopes. Many musicians have therefore sought short cuts that could generate timbres comparable to those generated by additive synthesis, but with less toil. One of the notable short cuts is frequency-modulated synthesis, invented by John M. Chowning of Stanford University. It is the technique applied in the most popular digital synthesizers today.

Frequency modulation (FM) is usually thought of as a radio-communications technique for transmitting information by modulating, or varying, the frequency of a high-frequency signal (called the carrier) with a low-frequency information signal (called the modulator). Chowning's technique relies on carriers and modulators that have either identical frequencies or frequencies of the same order of magnitude. Frequency relations between carrier and modulator of this kind are avoided in FM radio transmissions because they would fruitlessly spread the information signal over a huge bandwidth. In the case of musical waveforms this spreading of the signal's frequency spectrum can be fruitfully applied as timbral enrichment.

An FM instrument is a little more complicated to understand than an additive-synthesis instrument. At least two oscillators—a carrier oscillator and a modulation oscillator—are necessary in FM synthesis. Both oscillators usually generate simple sinusoidal waveforms, whose attack and decay shapes are controlled by envelope generators. Essentially the frequency of the carrier-wave oscillator is continuously shifted by an amount proportional to the amplitude of the waveform generated by the modulation oscillator. Hence the frequency of the carrier is no longer constant; it is the sum of the average carrier frequency and the continuously varying output of the modulation oscillator.

If the average carrier frequency and the modulation frequency are the same, it turns out that the fundamental period of the frequency-modulated wave will be the same as that of the unmodified carrier wave. What has changed, however, is the carrier's waveform. It can be shown that as the amplitude of the modulation signal is increased, the number and strength of higher-frequency harmonic partials in the carrier waveform increase.

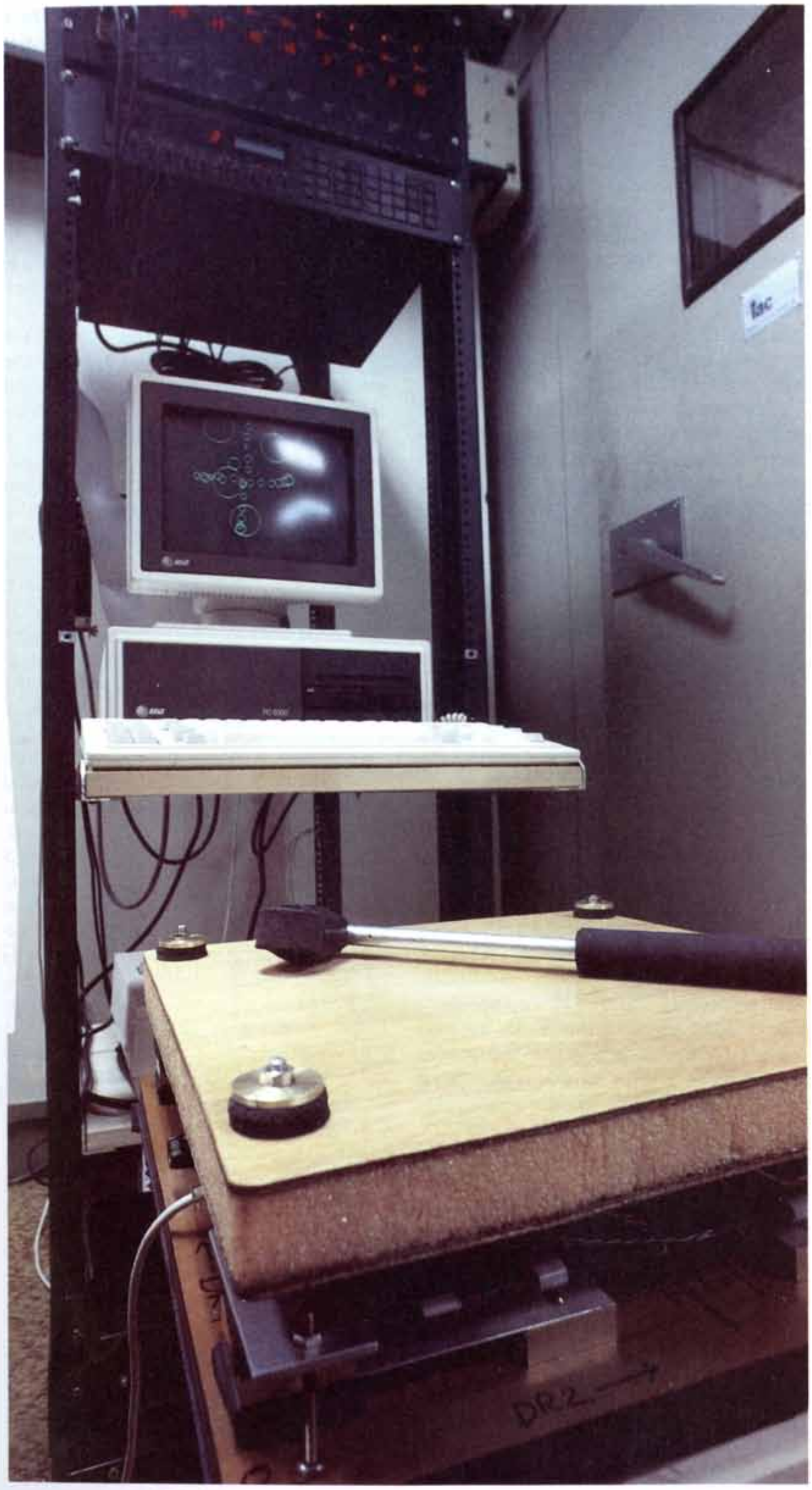
Suppose the modulator's envelope has a flatter attack than that of the carrier. In this case the high-frequency partials will build up slowly to their steady-state amplitude values. This is just what is needed for a brass-instrument timbre. Actually the computer program we have just described produces a passable brass timbre, and it requires only two oscillators and two envelope generators. In contrast, synthesizing a comparable brass tone by means of the additive-synthesis technique typically requires 10 separate oscillators and 10 separate envelope generators.

If the frequency of the modulator is not equal to that of the carrier, the partials in the frequency-modulated signal are inharmonic. The frequency spectrum of the sound consists of a cluster of partials centered on the carrier frequency and spaced apart from one another by an amount equal to the modulation frequency. If the modulation frequency is very low, a dense set of partials forms and in general a harsh, dissonant timbre is the result. If the modulation frequency is greater than the carrier, the inharmonic partials are spread out to yield a percussive timbre.

Although FM cannot produce any arbitrarily specified frequency spectrum, Chowning showed that a great array of musically interesting sounds could be efficiently produced merely by choosing the right frequencies and the right envelopes for the carrier and modulator. Moreover, the basic FM sound can be further enriched by simply adding together several FM waveforms. The Yamaha DX-7 synthesizer, which is based on FM synthesis, has six oscillators for each of 16 simultaneous voices. These oscillators are often grouped into three carrier-modulation pairs whose resultant FM signals are then summed.

Additive and FM synthesis are early methods of producing musical sounds by means of a computer that are still applied today. More recently developed approaches employ digital hardware designed specifically for musical purposes. Several synthesizer manufacturers, for example, have exploited the dramatic drop in the cost of memory chips to store copies of actual instrumental waveforms in sampled form. For each instrumental sound the waveform's attack and a part of its steady state or decay must be stored. Tones of various pitches are produced by speeding up or slowing down the rate at which the stored waveforms are strung together. The tonal quality and duration can be adjusted by averaging and smoothly fitting together different parts of the stored waveforms.

Although digitally stored, "natural" waveforms can produce "natural"



"INTELLIGENT" MUSICAL INSTRUMENT in the laboratory of one of the authors (Mathews) consists of a personal computer hooked up to a custom-made sensor and a commercially available digital sound synthesizer. The instrument enables one to "conduct" a programmed piece of music: tapping with a cushioned hammer on the rectangular "daton," or drum baton (*foreground*), sets the tempo; the placement of the hammer taps controls the instrumental balance and loudness. On the computer screen above the daton a series of strikes are displayed: a circle whose radius is proportional to the force of the strike indicates where the hammer hit the daton surface. The pair of black boxes on the rack above the computer are the sound synthesizer and a digital sound-processing device.

sounds, the waveforms cannot adequately capture all the loudness, pitch and timbre nuances in the phrasing of a musical passage. Hence much work is still spent on improving the "naturalness" of synthesized sounds, including those constructed from basic waveforms, such as sawtooth waves.

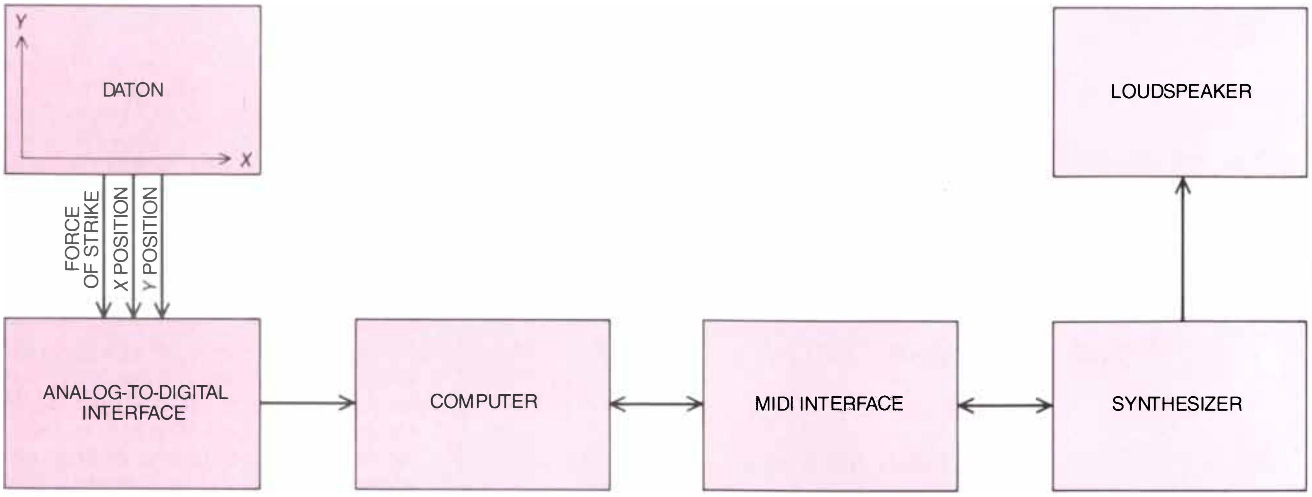
A sawtooth wave is similar to the waveform of the sound produced by bowing a violin string, yet when the sawtooth wave is played through loudspeakers, it sounds unpleasantly buzzy. The difference between a violin and a loudspeaker subjected to a sawtoothlike waveform is that the body of the violin naturally favors certain frequencies (the resonant frequencies) and damps others. Similarly, the human singing voice originates in vibrations of the vocal folds, which in themselves produce a nondescript sound. The timbre of the voice is formed almost entirely by the resonances of the vocal tract.

These principles of musical sound were understood at the beginning of computer music, but they were difficult to apply. New developments, primarily related to the cost and speed of integrated circuits, now make it possible to simulate by computer the characteristic resonances of natural sound-producing systems.

Sound synthesis based on this notion has two aspects: a general excitation, such as that provided by a sawtooth waveform, and a subsequent oscillatory decay of several major resonances. The key to getting a computer to produce a violinlike sound from a sawtooth waveform therefore lies in the simulation of the oscillatory decay of various frequency components. The computer can do this by solving linear difference equations, the discrete analogues of linear differential equations. The solution of a linear difference equation is a sum of damped sinusoids, and these can simulate the decaying oscillations of a violin soundboard or of the vocal tract.

The general problem with this type of synthesis is that it takes a lot of computing to solve a difference equation. Carver A. Mead and John C. Wawrzynek of the California Institute of Technology have recently manufactured integrated-circuit chips specifically designed for musical synthesis by the solution of linear difference equations. If such chips become available commercially, synthesis through this approach may leap forward.

Today there is no lack of strategies for the digital generation of complex and satisfying musical sounds. The chief problem with most strategies is the amount of computation nec-



MUSICAL-INSTRUMENT DIGITAL INTERFACE (MIDI) is a protocol for digitally coding musical data that was adopted by manufacturers of electronic instruments a few years ago. It allows a synthesizer to communicate with a computer. In the daton instrument (*see illustration on preceding page*) analog electrical signals that specify where and how hard the daton surface was hit are converted into binary numbers and passed on to a computer. The computer combines this information with the notes that have been stored in its memory and sends the combined information in MIDI form to the synthesizer, which then plays the appropriate tones.

essary to produce sounds with a lush timbre. A rough estimate is a million operations (multiplications and additions) per second of sound per instrument for a total of from 10 to 20 million operations per second of sound in a large work. The prodigious computations involved in synthesizing complex sounds have prevented the effective generation of sounds by general-purpose computers in real time—that is, without the need to record the sound samples at a low speed before they can be heard at normal speed.

Early computers did not have nearly enough capacity to synthesize music in real time. Instead composers synthesized their pieces slowly and deliberately, recording the sound samples on digital magnetic tapes and listening to the result on playback. Much fine music was created in this way, and it still may be the best way to make records and fill sound tracks, because a composer can apply a great degree of control in creating, evaluating and revising his musical scores.

Nevertheless, the recording process eliminates one kind of musician—the "live" performer. Interpretation of performance nuance must be written into the score by the composer; otherwise it will be missing. Also missing is the active pleasure of playing music, which is very important to both professional and amateur performers; listeners to such music may also regret the lack of a chance to play along vicariously with musicians at concerts.

Although general-purpose computers having enough power for real-time musical capability now exist, they are generally too expensive and cumbersome to serve as practical musical instruments in the concert hall or the home. The solution certainly lies in the production of efficient, special-purpose chips, such as those in the Yamaha synthesizers and those made in experimental form by Mead and Wawrzynek. In fact, digital musical instruments based on such chips are in many cases less expensive than some traditional acoustic instruments.

In this connection, the establishment of a standard musical-instrument digital interface (MIDI) protocol among manufacturers of commercial electronic instruments is good news for computer musicians: it allows a computer to be hooked up to such instruments, endowing them with some "intelligence." MIDI was originally intended to standardize the transmission of control information between various brands of synthesizers. Pressing a key on the keyboard of a MIDI-capable synthesizer not only causes a tone to be played but also transmits some data bits on an output cable that identify which key was pressed and how hard it was struck. A synthesizer can also have a MIDI input cable. If it receives key-play information through this cable, it will play a tone exactly as though one of its own keys had been pressed. In principle, anything that can be done on a synthesizer can be locally controlled by sensors on the machine (such as keys, buttons or knobs) or remotely controlled through MIDI.

Although the creators of MIDI may not have intended it to serve as the mode of communication between a computer and a synthesizer, it works well in this role. It certainly makes life beautiful for people who love to play with computers and synthesizers: an ordinary personal computer has plenty of power for most music-control purposes, since the waveform generation is taken care of by specialized circuits in the synthesizer itself. The essential parts of such an intelligent instrument are the synthesizer, a sensor on which the performer plays, the computer and, last but far from least, the software that ties all the components together.

Although the methods we have described for synthesizing sounds have been used to imitate traditional instruments, digital electronic equipment can just as easily create entirely new classes of sounds. More important, a sampled waveform—whether it is digitally recorded from a "real" instrument or manufactured in a computer—lends itself to easy manipulation. By means of digital sound processing a particular sound can be readily transformed into a completely different sound. For example, the frequency spectra and envelopes characterizing waveforms of human speech can be molded so that they sound like, say, a lion's roar.

Further study of musical sound by means of computers will certainly lead to more accurate imitations of traditional instruments and to the devising of means for the subtle and rapid control of their sound qualities, which is so important in actual musical performance. Also, computers have a key role in elucidating the subjective response that sounds elicit. This is particularly important for the modern composer because he or she is no longer limited to arranging sounds that can be produced from conventional instruments; it is now possible to call forth any sound that is imaginable—and even some that are not.

THE AMATEUR SCIENTIST

The secret of a microwave oven's rapid cooking action is disclosed

by Jearl Walker

The secret of a microwave oven's fast action is that water in the food rapidly absorbs the energy of the waves. At first thought this is surprising. In most cases where a material absorbs light or other types of electromagnetic radiation there must be a resonant match between the energy of the radiation and the energy changes possible in the atoms and molecules of the material. Water absorbs infrared radiation resonantly, but the frequency of microwaves is far too low for a resonant match. How, then, is energy transferred from the microwaves to the random thermal motion of the molecules of water? Several hypotheses I shall discuss represent a groping toward the answer. First one needs a bit of background on the physics of microwaves and the water molecule (H_2O), consisting of two atoms of hydrogen and one atom of oxygen.

Microwave frequencies range from 1×10^9 hertz (cycles per second) to 5×10^{12} hertz, well below the frequency of visible light (approximately 6×10^{14} hertz). My oven emits waves at the frequency of 2.45×10^9 hertz, or 2.45 gigahertz. Except for the frequency, microwaves and visible light are similar. Each can be pictured as a wave of oscillating electric fields.

A water molecule can be rotated by the electric field of a microwave because of the arrangement of the electric charge within the molecule. The electrons (negatively charged) associated with the hydrogen atoms are shifted toward the oxygen atom because of their strong attraction to the eight positively charged protons in the oxygen. The shift leaves the oxygen end of the molecule negative and the hydrogen ends positive. Such a charge distribution is called an electric dipole. Although the molecule as a whole is electrically neutral, it contributes an electric field to its surroundings and can be rotated by an electric field imposed on it. The dipole moment is the product of the net charge at each end and the separation between the charges. The moment is represented by a vector that points from the negative oxygen end along the line of symmetry between the hydrogen ends.

Ordinarily the dipole moments in water are randomly oriented. If an electric field is imposed, however, it creates a torque on each molecule. The torque makes the molecule rotate to align its dipole moment with the field. Picture an electric field produced by two charged, parallel plates. The direction of the field is from the positive plate to the negative one. The water molecule rotates so that it presents its negative end to the positive plate and its positive ends to the negative plate, thereby aligning itself with the field.

Every molecule is constantly buffeted by the random thermal motion of the surrounding molecules. The random motion, which sometimes goes by the name of Brownian motion, is related to the temperature of the water. Heat gives the molecules more kinetic energy, so that in their random motion they strike one another more vigorously. The temperature rises.

The polarization of water is expressed as the net dipole moment per unit of volume. It is zero when the dipole moments are randomly oriented, because for every moment pointing in one direction another moment points in the opposite direction. When an electric field begins to align the dipole moments, polarization increases. It would be at a maximum if all the dipoles were in alignment. Random molecular motion, however, continuously knocks dipoles out of alignment, keeping the polarization below the maximum level.

Now for the first hypothesis. Early in this century the eminent Dutch physicist Peter J. W. Debye demonstrated mathematically why microwave energy is strongly absorbed by water. The key fact is that water molecules cannot rotate instantaneously into alignment with the electric field. Since they have a mass spread over a certain volume, it takes time for torque to make them rotate. The retarding forces from surrounding molecules also affect rotation.

The response time of water can be considered in terms of the decay of its polarization if an electric field is suddenly turned off. Random molecular motion then begins to destroy the alignment of the dipole moments. The effect is to reduce the polarization exponentially with time.

Water's response time determines whether the dipole moments can keep up with the oscillating electric field in an electromagnetic wave. At low frequencies the time taken by the electric field to change direction is longer than the response time of the dipoles, and polarization keeps in phase with the electric field. The field provides energy to make the molecules rotate into alignment. Some of the energy is transferred to the random motion each time a dipole is knocked out of alignment and then realigned. The transfer of energy is so small, however, that the temperature hardly rises. If the electric field oscillates rapidly, it changes direction faster than the response time of the dipoles. Since the dipoles do not rotate, no energy is absorbed and the water does not heat up.

In the microwave range of frequencies the time in which the field changes is about the same as the response time of the dipoles. They rotate because of the torques they experience, but the resulting polarization lags behind the changes in the direction of the electric field. When the field is at maximum strength, say in an upward direction, polarization may still be low. It keeps rising as the field weakens. The lag indicates that the water absorbs energy from the field.

Microwave ovens operate at a frequency that is lower than the frequency at which absorption is greatest. The practical reason is that the user wants to heat food throughout its interior. If the frequency is optimal for a maximum heating rate, the microwaves are absorbed in the outer regions of the food, penetrating only a short distance. If the frequency is lower, say 2.45 gigahertz, the penetration improves. Some ovens operate at a frequency of .915 gigahertz, and so the penetration is even greater.

As neat as Debye's mathematical so-

Powder initiates boiling in water superheated by microwaves lution is, it leaves unanswered the question of exactly how energy is transferred from the microwaves to the random motion of the molecules. Debye put forward a simple model, which serves as the second hypothesis. Consider a water molecule as being spherical. When the sphere is rotated by the electric field of a microwave, it experiences viscous drag from the water around it. The drag is important only at microwave frequencies. At lower frequencies the sphere rotates too slowly to encounter drag. At higher frequencies it does not rotate at all.

At microwave frequencies it rotates fast enough so that the drag fights the rotation, requiring the field to supply additional energy. This is the energy that goes into the random motion of the molecules surrounding the sphere, raising the temperature. Exactly why the sphere experiences a viscous drag and how the energy is transferred to the surrounding molecules was not understood in Debye's time and is still not understood in detail. A simple explanation can be made. Suppose the sphere is initially in some brief state of equilibrium with the electrical forces from the surrounding molecules. If the sphere is to turn, it must upset the equilibrium and move the molecules, thereby increasing the energy in their random motion.

As crude as Debye's simple model is, it is surprisingly accurate in predicting the frequency at which the absorption rate is greatest. Debye assumed that the molecule has a radius of 2×10^{-10} meter and that the viscosity it encounters is the viscosity of bulk water. With his model he calculated that the response time should be about 2.5×10^{-11} second. Since the frequency at which the absorption rate is greatest is approximately the inverse of the response time, it should be about 40 gigahertz, which is about right in terms of the model.

A more refined model explaining how microwaves heat water has since been proposed. It is the third hypothesis. In addition to individual molecules water is made up of many short-lived groups of molecules. The groups are held together by hydrogen bonds: the hydrogen ends of one molecule are attracted to the oxygen end of another molecule. Every time two or more molecules form a group they lower their electric-potential energy. The difference in energy is added to the kinetic energy of the group's random motion. The temperature does not rise because on the average the number of molecular groups that form equals the number of groups broken apart by collisions from the random motion, leaving no net gain in the energy of the random motion.

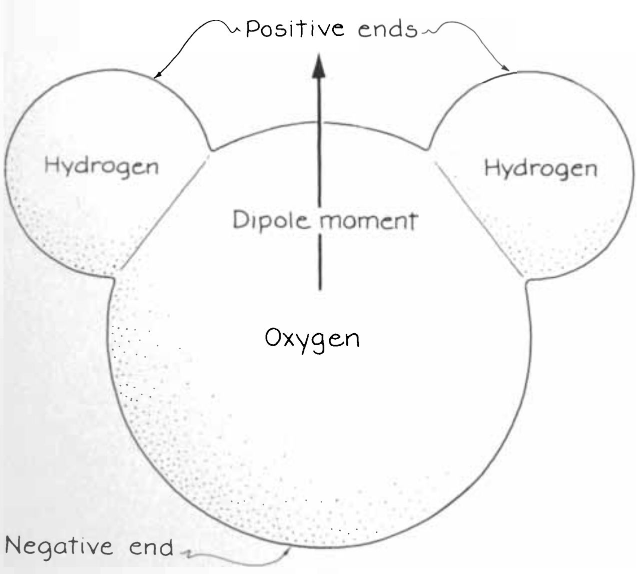
A microwave adds energy to the random motion if the torques it places on the molecules break some of the hydrogen bonds in the groups. (The electric field supplies the energy to break the bonds.) When a liberated molecule again forms a hydrogen bond with a group, the decrease in potential energy goes into the random motion of the molecular group.

Which of the various groups might be susceptible to such bond breaking? Individual molecules do not take part because the torque need not break a bond to rotate the molecule. Groups of two molecules are also not likely to participate because each molecule can be rotated about the hydrogen bond holding them together, with no bond breaking. Groups of four or more molecules probably do not participate either, because several bonds would have to be broken.

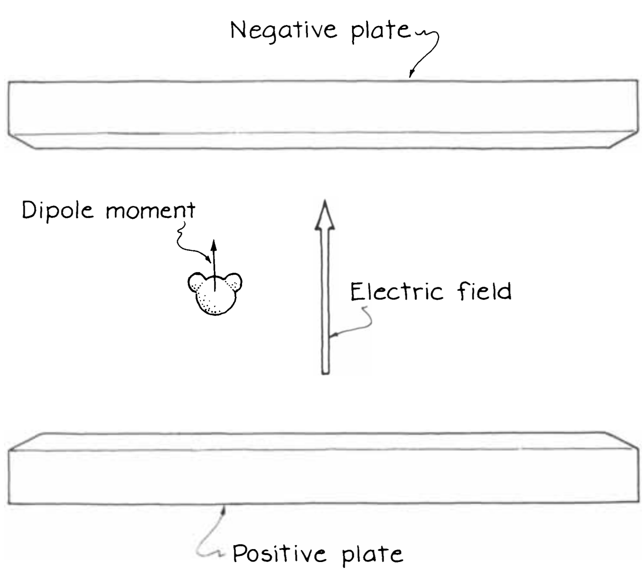
The best candidate for the process is a group of three water molecules in a particular arrangement [*see lower illustration at left on page 138*]. The middle molecule is attached to the second molecule by one of its own hydrogen ends. The other hydrogen end is free. Attached to the oxygen end of the middle molecule is the third water molecule. The attachment is by way of a hydrogen end on the third molecule.

The position of the attachment is important. Calculations indicate that there are two sites at the oxygen end of the middle molecule where the attachment would lower the potential energy the most. Attachments at other places are therefore less likely. For example, an attachment is not likely to be made near the free hydrogen end of the middle molecule because that hydrogen end will repel the hydrogen end of the third molecule.

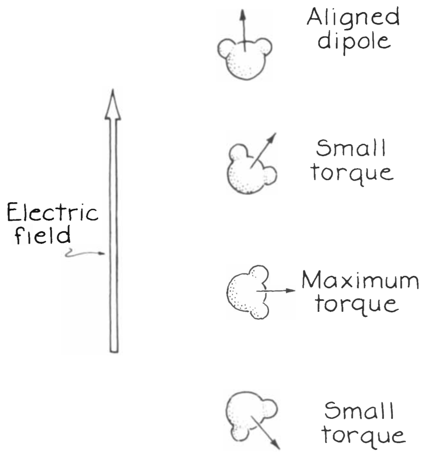
The torque on the middle molecule from the electric field of a microwave may be large enough to break the hydrogen bond with the third molecule. After the middle molecule rotates, the third one could reestablish its hydrogen bond at the other low-energy spot on the oxygen end of the middle molecule. Such a process puts energy into the random motion of the group. The field supplies energy to break the initial hydrogen bond. When the bond is



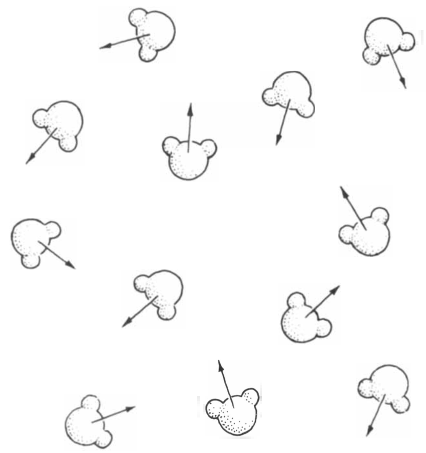
A water molecule

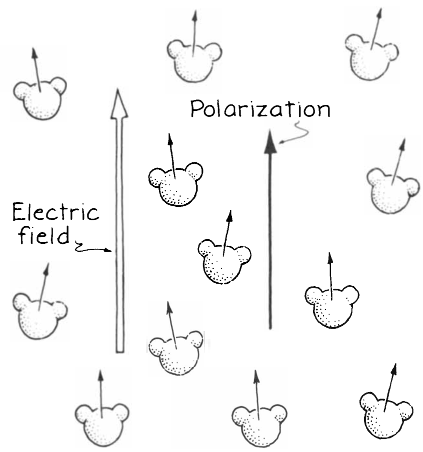


A dipole aligned in an electric field



Torques on dipoles


Randomly oriented water dipoles



Dipoles aligned

reestablished, the energy goes into the kinetic energy of the group. In this way the microwave energy ends up as heat.

So much for hypotheses on how the water in food rapidly absorbs microwave energy. I shall now turn to some of the other things that go on in a microwave oven.

The presence of sodium chloride in the water increases the heating rate. The salt separates into positive sodium ions and negative chlorine ions. The positive ions are surrounded by up to four molecules of water, the negative ions by up to seven. In each case the positive or negative end of a water molecule is electrically attracted to the charged ion. The electric field of a microwave drives the hydrated ions through the water, pushing the sodium ions in the direction of the field and the chlorine ions in the opposite direction. Whenever the hydrated ions bump into water molecules, energy is transferred to the random motion of those molecules, heating the water.

The water molecules locked into the crystalline structure of ice cannot absorb energy from microwaves because they are immobile. How then does a microwave oven melt ice or cook frozen foods? The answer is that the objects are not completely frozen. Within seconds after you remove an ice cube from the freezer its surface begins to melt. If you put it in a microwave oven, the liquid layer on the outside absorbs microwaves, heats up and melts the ice.

Frozen foods exposed to air may also be coated with a layer of liquid. They melt there and at many internal points where the water is liquid. If they are exposed to microwaves, the pockets heat up rapidly. As their surroundings thaw, the new water begins to absorb even more energy from the microwaves. This situation is called runaway heating. In order to avoid it food should be thawed at low power or exposed to high power only periodically so that the heat can get into the frozen sections by conduction, thereby thawing the food more uniformly.

In conventional cooking the food is kept in a hot environment so that heat can move from the surface into the interior by conduction or convection. When a beef roast is prepared in this way, the environment may reach 170 degrees Celsius (338 degrees Fahrenheit), considerably hotter than the boiling temperature of water. The interior probably never reaches a temperature higher than 70 or 80 degrees C. At that temperature the myoglobin pigment in the meat changes to oxymyoglobin, which is bright red. Meanwhile the surface of the meat becomes so hot that its oxymyoglobin denatures, turning brown. The high temperature also changes the flavor and aroma of the suface of the meat.

When beef is cooked in a microwave oven, the water heats the solid material. The surface never reaches a temperature higher than 100 degrees C., the boiling point of water. Since that temperature is not high enough to denature the oxymyoglobin fully, the surface never becomes dark brown. The meat also never develops the flavor and aroma of meat cooked in a conventional oven.

If the meat is thin, the entire interior cooks by the direct absorption of microwaves. If the meat is thick, as a roast usually is, the microwaves are absorbed before they reach the center. Heat is conducted to the center from the region directly heated by the microwaves. Because conduction takes time, a large roast must be allowed to sit after it has been heated in a microwave oven.

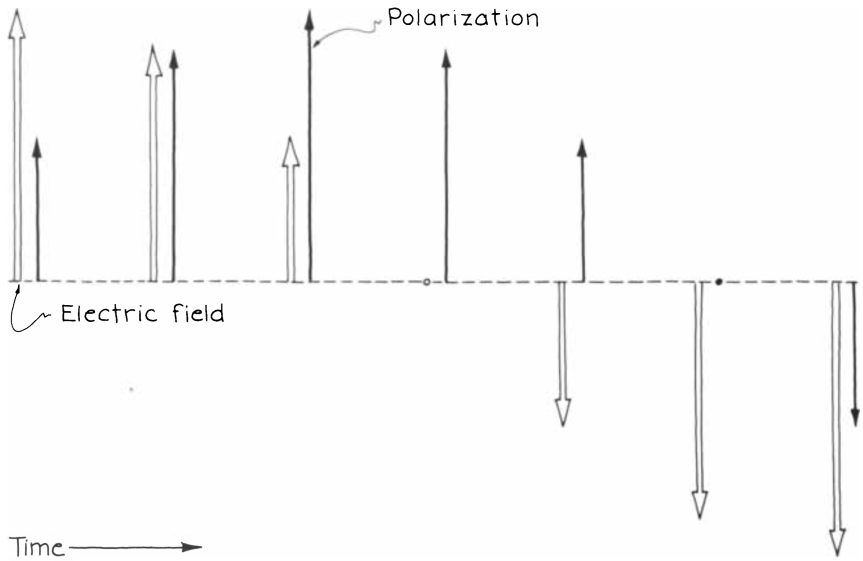
Liquid water is heated so rapidly by microwaves that steam can present a problem. If an egg or even an exposed unbroken yolk is heated by microwaves, the rapid interior production of steam can make the object explode. If food is cooked in a closed container, an escape route must be provided for the steam. That is why plastic cooking bags must be slit at the top.

Metal containers should be avoided in microwave cooking for several reasons. Metal reflects microwaves, thereby shielding the food and possibly returning enough energy to the microwave emitter to overload it. Because metal conducts electricity, sparks jump between the container and the bottom or walls of the oven. As I learned by accident, even a wire twist closing a plastic bag can produce frightening sparks if it is near the bottom of the oven. Aluminum foil is sometimes used in microwave cooking to shield parts of a fowl that might overcook. The foil is not dangerous if there is not much of it and it is kept at least several centimeters from the bottom and walls of the oven. The containers usually recommended for microwave cooking are often plastic or glass. They contain materials that absorb microwaves poorly or not at all. The containers get hot only because energy is conducted into them from the food. Sometimes the containers do absorb a small amount of microwave energy, with the result that they help to heat the food.

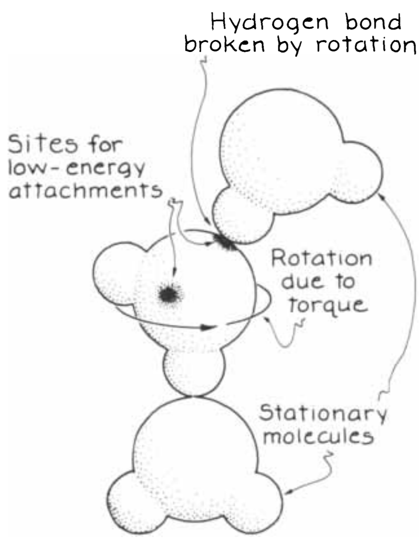
In 1986 Robert E. Apfel of Yale University and Richard L. Day of the Yale University School of Medicine pointed out that when water is heated in a glass container in a microwave oven, it can superheat up to 110 degrees C. with no sign of boiling. The water in the center of the container heats so fast that convection to the top surface and the subsequent vaporization are insufficient to prevent superheating. Bubbles of water vapor do not form in the cooler water along the sides of the container, partly because the container does not absorb microwaves and conducts heat from the water. If an ice cube is dropped into the superheated water, vapor bubbles rapidly form in the abundant microscopic crevices on the surface of the ice. The cube continues to initiate boiling until the temperature of the surrounding water drops to about 102 degrees.

I investigated a similar example of superheating by partially filling a glass beaker with water and then pouring a layer of corn oil on the water. When the water superheated, vapor bubbles formed on the glass sides and then forced their way through the oil. In one trial the water apparently superheated so much that water and oil were blown throughout the oven by a sudden and extensive vaporization in the water. Similar minor explosions can result when grease from cooking meat lies on top of water in the roasting pan. I avoid a mess in the oven by enclosing the meat in a plastic bag that I have slit in a few places.

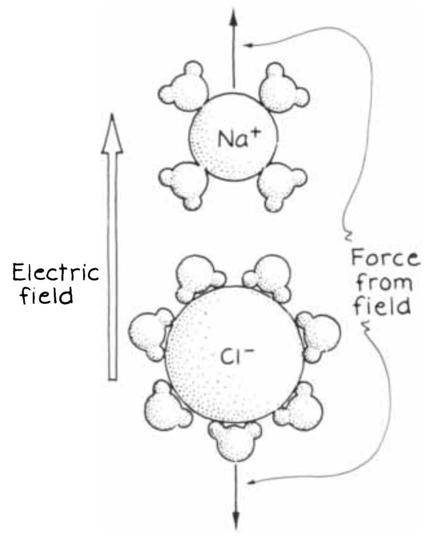
Apfel and Day also observed that a spoon can initiate boiling in superheated water. Anthony E. Siegman of Stanford University had previously written to me about the phenomenon. Anthony Parsons of the University of York recently noted that powder added to superheated water results in such rapid boiling that the contents may overflow the container.



How polarization lags behind an electric field



A molecule rotated within a group



Forces on hydrated ions

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

THE ORIGIN OF SPACEWAR. J. M. Graetz in *Creative Computing,* Vol. 7, No. 8, pages 56–67; August, 1981.

WINNING WAYS FOR YOUR MATHEMATICAL PLAYS, VOL. 2: GAMES IN PARTICULAR. Elwyn R. Berlekamp, John H. Conway and Richard K. Guy. Academic Press, 1982.

MICROCOMPUTER DISPLAYS, GRAPHICS, AND ANIMATION. Bruce A. Artwick. Prentice-Hall, Inc., 1984.

HUNGER IN THE U.S.

HUNGER IN AMERICA: THE GROWING EPIDEMIC. Physician Task Force on Hunger in America. Wesleyan University Press, 1985.

INCREASING HUNGER AND DECLINING HELP: BARRIERS TO PARTICIPATION IN THE FOOD STAMP PROGRAM. Physician Task Force on Hunger in America. Harvard University, 1986.

PRIMARY PREVENTION OF FAILURE-TO-THRIVE: SOCIAL POLICY IMPLICATIONS. D. Frank, D. Allen and L. Brown in *New Directions in Failure-to-Thrive: Implications for Future Research and Practice,* edited by Dennis Drotar. Plenum Press, 1986.

THE MICROTUBULE AS AN INTRACELLULAR ENGINE

STUDIES ON THE MOTILITY OF THE FORAMINIFERA, II: THE DYNAMIC MICROTUBULAR CYTOSKELETON OF THE RETICULOPODIAL NETWORK OF *ALLOGROMIA LATICOLLARIS.* Jeffrey L. Travis, James F. X. Kenealy and Robert D. Allen in *The Journal of Cell Biology,* Vol. 97, No. 6, pages 1668–1676; December, 1983.

NEW OBSERVATIONS ON CELL ARCHITECTURE AND DYNAMICS BY VIDEO-ENHANCED CONTRAST OPTICAL MICROSCOPY. Robert Day Allen in *Annual Review of Biophysics and Biophysical Chemistry,* Vol. 14, pages 265–290; 1985.

GLIDING MOVEMENT OF AND BIDIRECTIONAL TRANSPORT ALONG SINGLE NATIVE MICROTUBULES FROM SQUID AXOPLASM: EVIDENCE FOR AN ACTIVE ROLE OF MICROTUBULES IN CYTOPLASMIC TRANSPORT. Robert Day Allen, Dieter G. Weiss, John H. Hayden, Douglas T. Brown, Hideshi Fujiwake and Marcia Simpson in *The Journal of Cell Biology,* Vol. 100, No. 5, pages 1736–1752; May, 1985.

THE OLDEST PULSARS IN THE UNIVERSE

A MILLISECOND PULSAR. D. C. Backer, Shrinivas R. Kulkarni, Carl Heiles, M. M. Davis and W. M. Goss in *Nature,* Vol. 300, No. 5893, pages 615–618; December 16, 1982.

A NEW CLASS OF RADIO PULSARS. M. A. Alpar, A. F. Cheng, M. A. Ruderman and J. Shaham in *Nature,* Vol. 300, No. 5894, pages 728–730; December 16, 1982.

BLACK HOLES, WHITE DWARFS, AND NEUTRON STARS: THE PHYSICS OF COMPACT OBJECTS. Stuart L. Shapiro and Saul A. Teukolsky. John Wiley & Sons, Inc., 1983.

INTENSITY-DEPENDENT QUASI-PERIODIC OSCILLATIONS IN THE X-RAY FLUX OF GX5–1. M. van der Klis, F. Jansen, J. van Paradijs, W. H. G. Lewin, E. P. J. van den Heuvel, J. E. Trümper and M. Sztajno in *Nature,* Vol. 316, No. 6025, pages 225–230; July 18, 1985.

GENETIC RECOMBINATION

GENETIC RECOMBINATION: THINKING ABOUT IT IN PHAGE AND FUNGI. Franklin W. Stahl. W. H. Freeman and Company, 1979.

ROLES OF RECBC ENZYME AND CHI SITES IN HOMOLOGOUS RECOMBINATION. G. R. Smith, S. K. Amundsen, A. M. Chaudhury, K. C. Cheng, A. S. Ponticelli, C. M. Roberts, D. W. Schultz and A. F. Taylor in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. 49, pages 485–495; 1984.

ROLES OF DOUBLE-STRAND BREAKS IN GENERALIZED GENETIC RECOMBINATION. F. W. Stahl in *Progress in Nucleic Acid Research and Molecular Biology,* Vol. 33, pages 169–193; 1986.

BALLISTIC ELECTRONS IN SEMICONDUCTORS

BALLISTIC TRANSPORT IN SEMICONDUCTOR AT LOW TEMPERATURES FOR LOW-POWER HIGH-SPEED LOGIC. Michael S. Shur and Lester F. Eastman in *IEEE Transactions on Electron Devices,* Vol. ED-26, No. 11, pages 1677–1683; November, 1979.

DIRECT OBSERVATION OF BALLISTIC TRANSPORT IN GAAS. M. Heiblum, M. I. Nathan, D. C. Thomas and C. M. Knoedler in *Physical Review Letters,* Vol. 55, No. 20, pages 2200–2203; November 11, 1985.

THE QUEST FOR BALLISTIC ACTION. Trudy E. Bell in *IEEE Spectrum,* Vol. 23, No. 2, pages 36–38; February, 1986.

WHALES AND WALRUSES AS TILLERS OF THE SEA FLOOR

ECOLOGY AND BIOLOGY OF THE PACIFIC WALRUS, *ODOBENUS ROSMARUS DIVERGENS* ILLIGER. Francis H. Fay. U.S. Department of the Interior, Fish and Wildlife Service, North American Fauna No. 74. U.S. Government Printing Office, 1982.

THE GRAY WHALE, *ESCHRICHTIUS ROBUSTUS.* Edited by Mary Lou Jones, Steven L. Swartz and Stephen Leatherwood. Academic Press, 1984.

SIDE-SCAN SONAR ASSESSMENT OF GRAY WHALE FEEDING IN THE BERING SEA. Kirk R. Johnson and C. Hans Nelson in *Science,* Vol. 225, No. 4667, pages 1150–1152; September 14, 1984.

MARRIAGE, MOTHERHOOD AND RESEARCH PERFORMANCE IN SCIENCE

FAIR SCIENCE: WOMEN IN THE SCIENTIFIC COMMUNITY. Jonathan R. Cole. The Free Press, 1979.

CAREER OUTCOMES IN A MATCHED SAMPLE OF MEN AND WOMEN PH.D.S: AN ANALYTICAL REPORT. Nancy C. Ahern and Elizabeth L. Scott. National Academy Press, 1981.

THE PRODUCTIVITY PUZZLE: PERSPECTIVE AND CHANGE IN PATTERNS OF PUBLICATION OF MEN AND WOMEN SCIENTISTS. Jonathan R. Cole and Harriet Zuckerman in *Advances in Motivation and Achievement,* Vol. 2, pages 217–258; 1984.

THE COMPUTER AS A MUSICAL INSTRUMENT

THE TECHNOLOGY OF COMPUTER MUSIC. Max V. Mathews. The MIT Press, 1969.

THE SCIENCE OF MUSICAL SOUND. John R. Pierce. W. H. Freeman and Company, 1983.

COMPUTER MUSIC. Charles Dodge and Thomas A. Jerse. Schirmer Books, 1985.

THE AMATEUR SCIENTIST

AQUEOUS DIELECTRICS. J. B. Hasted. Chapman and Hall, 1973.

INITIATING BOILING WITH ICE. Robert E. Apfel and Richard L. Day in *Nature,* Vol. 321, No. 6071, page 657; June 12, 1986.

Dependable.
A reputation earned on six continents, in 77 countries, and 150 cities around the world.

Lufthansa
German Airlines

See your Travel Agent.