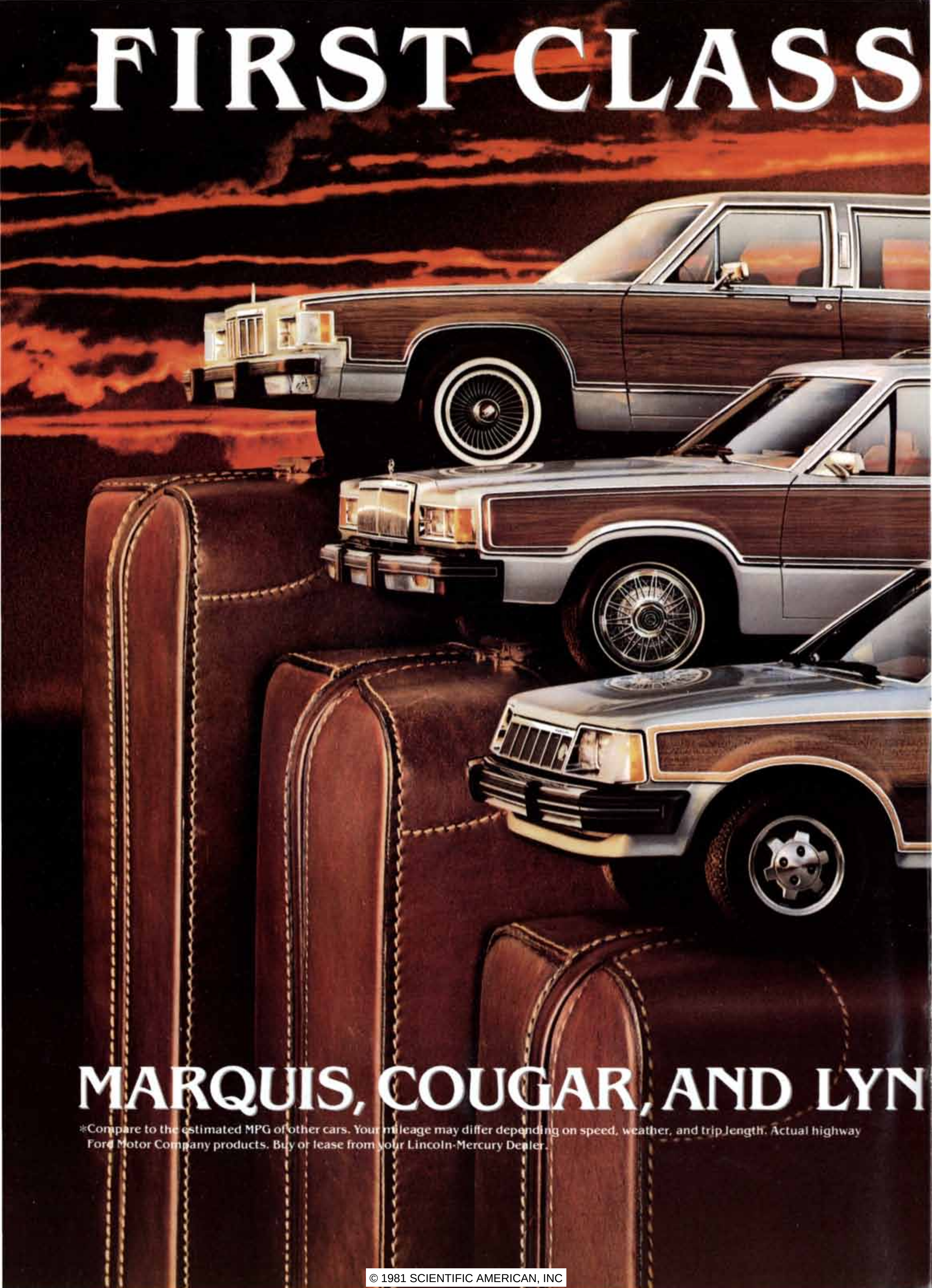
SCIENTIFIC AMERICAN

BUTTERFLY COLOR PATTERNS

$2.00

November 1981


FIRST CLASS
MARQUIS, COUGAR, AND LYN
*Compare to the estimated MPG of other cars. Your mileage may differ depending on speed, weather, and trip length. Actual highway
Ford Motor Company products. Buy or lease from your Lincoln-Mercury Dealer.


TRAVELERS
No matter where your journeys take you, this year you can travel first class. Because this year, Lincoln-Mercury has a wagon with the comfort features and amenities sized just right for you.
MARQUIS, 53 CUBIC** FEET 17 *EPA EST. MPG 26 EST. HWY.
COUGAR, 42 CUBIC** FEET 19 *EPA EST. MPG 26 EST. HWY.
LYNX, 28 CUBIC** FEET 28 *EPA EST. MPG 44 EST. HWY.
Lynx, of course, is the mileage leader of American wagons and has a special inner world for your passengers and your cargo. There's a totally new Cougar wagon, marked by its distinctive grille, generous capacity, and individualized interior. Plus the 1982 Marquis wagon, which offers traditional space with contemporary mileage. Visit your Lincoln-Mercury Dealer and find your wagon. You'll find it's the difference between coach and first class.
LINCOLN-MERCURY DIVISION
X FROM LINCOLN-MERCURY
mileage probably lower. *With optional 3.8-liter engine. **Based on EPA Interior Volume Index figures. All comparisons exclude other

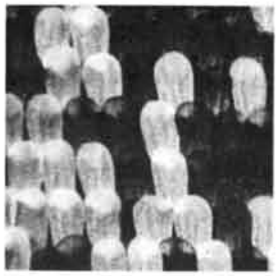

BEAM
THE WORLD'S FINEST BOURBON
JIM BEAM

Established 1845 **SCIENTIFIC AMERICAN** November 1981 Volume 245 Number 5

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $21 PER YEAR, U.S., ITS POSSESSIONS AND CANADA, $27 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.



THE COVER

The photograph on the cover shows a few of the scales of a butterfly's wing at a magnification of about 280 diameters. Each scale is the flattened outgrowth of a single cell of the wing; together the scales, overlapping like roof tiles, make up a mosaic that forms the color pattern of the wing (see "The Color Patterns of Butterflies and Moths," by H. Frederik Nijhout, page 140). Here the scales form part of the pattern of the underside of the wing of the butterfly *Limenitis astyanax*. The blue scales derive their color from the interference of reflected light. This type of color is called structural because it arises from structural features of the scales. The color of the single reddish scale is pigmentary, arising from melanin in the scale. Each pigmentary scale is thought to contain only one pigment, but the amount varies from scale to scale, so that the intensity of a given color can range from, say, red to buff. A typical butterfly or moth pattern consists of no more than five colors and often of only three or four.

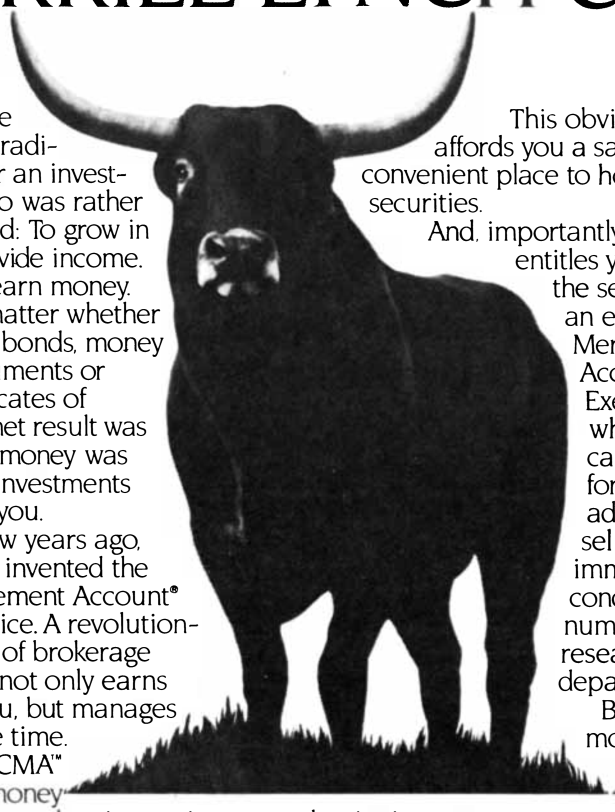
THE ILLUSTRATIONS

Cover photograph by Fritz Goro

Page	Source
24–39	Mitchell J. Feigenbaum
54	Social and Behavioral Sciences Laboratory, University of South Carolina
56–57	Albert E. Miller
58–59	Social and Behavioral Sciences Laboratory, University of South Carolina
60–61	Albert E. Miller
63	Ralph Morse
64–66	George V. Kelvin
68	Ralph Morse
69	Adolf R. Asam, ITT Electro-Optical Products Division, Roanoke, Va. (*top*); Michael Lampton (*bottom*)
70–71	George V. Kelvin
73	Paul Howard-Flanders
74–80	Patricia J. Wynne
104	Jet Propulsion Laboratory
106–107	Walken Graphics
108	Jet Propulsion Laboratory
112	Walken Graphics
117–118	Jet Propulsion Laboratory
120	Walken Graphics
125	Jet Propulsion Laboratory
126–128	Walken Graphics
141	Fritz Goro
142–143	Tom Prentiss
145	H. Frederik Nijhout
146	H. Frederik Nijhout (*left*), Tom Prentiss (*right*)
147	Tom Prentiss
148	H. Frederik Nijhout
150	Tom Prentiss
151	H. Frederik Nijhout and Tom Prentiss
155	Florence Moog
156–160	Carol Donner
162	Florence Moog (*top*), Susumu Ito (*bottom*)
166	Andrew Christie
171	Dominique Maestracci
174–176	Andrew Christie
181	Rob McLachlin (*top*), Peter Waddell (*bottom*)
182	Memorial University of Newfoundland
184	Andrew Tomko
186	Lester Ross, Parks Canada (*top*); J. Martin, Memorial University of Newfoundland (*bottom*)
187	J. Martin, Memorial University of Newfoundland (*top*); Canadian Conservation Institute (*bottom*)
188	Memorial University of Newfoundland
195	Kenneth A. Connors
196	John M. Franke
197–198	Michael Goodman
200	Adam Lloyd Cohen

IF YOUR INVESTMENTS ARE ONLY EARNING YOU MONEY, THEY AREN'T DOING NEARLY ENOUGH.

PRESE
THE MERRILL LYNCH CASH



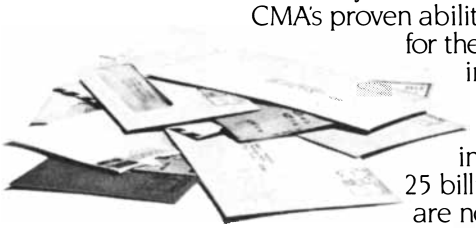
Until quite recently, the traditional role for an investment portfolio was rather strictly defined: To grow in worth. To provide income. Basically, to earn money.

But, no matter whether it was stocks, bonds, money market instruments or simply certificates of deposit, the net result was that earning money was all that your investments could do for you.

Then a few years ago, Merrill Lynch invented the Cash Management Account® financial service. A revolutionary new kind of brokerage account that not only earns money for you, but manages it at the same time.

In effect, CMA™ allows your money to continually work at its maximum potential, without relinquishing the least amount of liquidity, flexibility and convenience that you demand.

Today, there is no question about CMA's proven abilities and potential for the sophisticated investor.



In fact, more than 450,000 investors with over 25 billion dollars invested are now relying on its many features. The only question remaining should be how the CMA program can begin working for you, immediately.

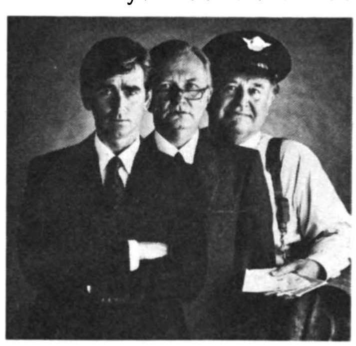
QUALIFICATION. To open a CMA account requires that you place a minimum of $20,000 in cash or securities, or a combination of both, in a Merrill Lynch brokerage account.

This obviously affords you a safe and convenient place to hold your securities.

And, importantly, it entitles you to the services of an experienced Merrill Lynch Account Executive whom you can rely upon for sound advice, counsel and an immediate conduit to our number-one research department.

But, much more, this is just the beginning your assets need to give them the special qualities that make the CMA service so unique.

MAXIMIZATION. There is virtually no such thing as idle cash in a CMA account. On opening your account, any cash you originally place there is automatically invested in your choice of three professionally managed money market funds. The choice of which fund is best for your particular investment strategy is, of course, your decision. But the CMA Money Fund, Tax Exempt Fund or the new Government Securities Fund can provide definite advantages for any portfolio.



Daily, the dividend on your

CMA™ is not available in all states. The Cash Management Account program is proprietary to Merrill Lynch. U.S. Patents Pending

NTING
MANAGEMENT ACCOUNT.®

money market fund is reinvested for you. And any cash that is generated on your securities (i.e. dividends, interest, proceeds from sales) is also automatically invested in your money fund. Without you, your broker or the postman ever getting involved. So, your money continually works earning more money, without you having to work at doing anything.

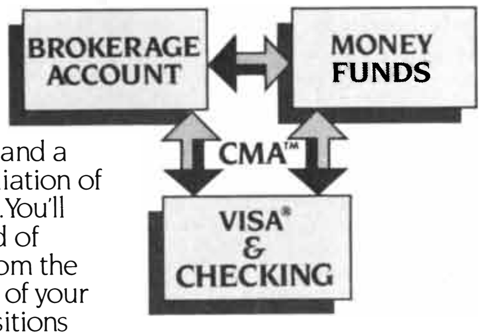
LIQUIDITY. Now, let's assume that while your money is busy working for you, it's necessary to take an important business trip. Or, on

vacation in Mexico City, you spot a rare antique.

With the CMA service, you have instant access to your assets, anywhere in the world, at any time. You can make purchases. And although Merrill Lynch is not a bank, you can get cash advances when you need to.

CMA's special checking and VISA® card (provided by Bank One of Columbus, Ohio, N.A.) are all the credentials you'll need to use your money when you want to. If there is not enough cash to cover your purchases from the balance in your money market fund, you are entitled to borrow up to the full margin value of your securities, instantly. This constitutes an automatic loan on *our* part. With no time limit of repayment on *your* part.

CONVENIENCE. At the end of each month, you'll receive CMA's comprehensive computerized statement. It lists all your transactions, and a daily reconciliation of your account. You'll have a record of everything from the market value of your security positions including annual percent yield and estimated income, as well as all your VISA and checking activities, on one convenient form.



And, if on any business day during the month, you want to know exactly how much money is available to you, you only have to call a special toll-free 800 number.

So, there you have it. CMA. A unique service that keeps your money working productively, while giving you immediate access to your assets.

For more complete information, including a prospectus containing all sales charges and expenses, call 800-526-3030 (in New Jersey, call 800-742-2900). Read the prospectus carefully before you invest or forward funds. We're convinced that once you realize all that your investments can do for you, you won't want to be without it.

THE MORE YOU DEMAND OF YOUR MONEY, THE MORE YOU NEED CMA.™

Merrill Lynch
Merrill Lynch Pierce Fenner & Smith Inc.
A breed apart.

LETTERS

Sirs:

Paul F. Walker hypothesizes ["Precision-guided Weapons," by Paul F. Walker; SCIENTIFIC AMERICAN, August] that tens of thousands of cheap missiles proliferated in a theater of warfare will put expensive machines such as tanks out of business. In developing this thesis he reckons without a responsive enemy, a "dirty" battlefield and the practical military and engineering problems of finding opposing tanks and delivering the precision-guided weapons against them.

The weapons cannot be fired against targets without launchers to start from and men to operate them. Launchers can vary from hand-held tubes to cannon to simple vehicle-mounted rails. Exposure and vulnerability of the men, the missiles and the launchers invite artillery fire, which, particularly if it is delivered according to published Russian concepts for massed use of artillery, would destroy large numbers of all three unless they were protected by armor. Armored fighting vehicles able to launch precision-guided antitank missiles approach tanks in cost, so that the values of the two kinds of weapon system must be weighed against each other to acquire a balanced force within a budget. Although the somewhat cheaper antitank-fighting vehicle or self-propelled armored cannon affords a very effective defensive capability, tanks are needed for the counterattacks that will inevitably be required against local enemy successes even in a defensive war. Similar considerations apply in the case of air-delivered precision-guided weapons, where heavy expense must be incurred to overcome increasingly effective air defenses (even using the precision weapons) to allow attack aircraft to operate against the armor itself....

The "dirty" battlefield is full of dust, smoke, "killed" targets and false targets such as sun-baked rocks and water-filled shell holes. All these considerations conspire to make weapon seekers less effective than they would be under ideal conditions, or they may require that the seekers be made so sophisticated and expensive that the weapons can no longer be purchased in large quantities. A responsive enemy will capitalize on these target-acquisition difficulties by using countermeasures that block or confuse the "vision" of the sensors, by presenting decoys and by other means to inhibit the seekers from finding and homing on targets; he will thereby cause the weapons to be used inefficiently and wastefully, significantly reducing their value.

Precision-guided weapons represent but one more step, although an important one, in antitank warfare. Such warfare has seen the use of "Molotov cocktails," bazookas, guns such as the German "88" of World War II and tanks themselves. Each major increase in antitank capability caused a change in the way tanks were used but did not make them obsolete. The chief value of armored forces (including tanks) is in providing mobile, protected infantry and artillery to attack and defeat opposing forces where they are weak, usually in rear areas. In World War II typical doctrine for an armored attack called for infantry and artillery to destroy the forward defenses, allowing the armor to pass through and exploit the advantage. Subsequently the growing power of armored forces led to a preference for direct assault by tanks, supported by tactical aviation and artillery, to make a breakthrough. The advent of precision-guided weapons in defense could cause a return to earlier tactics or a shift to still another tactical variation designed to neutralize the strength of the defense.

All of this is not to say that the impact of these weapons will be inconsequential. On the contrary, both sides will have them, and the net effect is likely to be very high loss rates of major equipment for both sides in a conflict. One of two results then becomes possible. It may be that ultimately neither side will be able to afford to build enough military forces that have both tanks and precision-guided weapon systems to sustain a war for long. The result could be a stalemate. This favors the defense, but it exacts a very high price....

The "bottom line" is that, as in so many other areas, these new inventions do not simplify life, as Walker implies, but rather they make it infinitely more complicated for the military planner.

S. J. DEITCHMAN

Institute for Defense Analyses
Arlington, Va.

Sirs:

Mr. Deitchman's criticism appears to be in partial agreement with my argument that precision-guided weapons will change the face of battle. He states that their impact will not be inconsequential. In fact, in his 1979 book on military technology, *New Technology and Military Power: General Purpose Military Forces for the 1980s and Beyond,* he writes that the "impact of ATGM [antitank guided missiles] has been to change the quantitative relationships—the potential of relative loss—very much in favor of the defense.... They inhibit the free run of a battlefield [and] can be used decisively against armor."

Yet he fails to draw the next logical conclusion: that major pieces of capital equipment become highly vulnerable and therefore questionable investments. He instead argues that combined arms—tanks, artillery and infantry—will still carry the day on a major battlefield after a heavy attrition of forces. Here he is falling into the trap of applying old tactics to new weaponry. I sought to point out that all concentrated forces—tank columns, supply depots, artillery batteries, military bases—are vulnerable to missile attack on today's battlefield (and even more so on tomorrow's). This being the case, it is doubtful that they would be left to win the battle after the first engagements. Deitchman also argues that precision-guided weapons will become almost as expensive and vulnerable by being mounted on armored vehicles. Here he misses the point; one does not want to gold-plate precision-guided weapons and thereby make them just as immobile, visible, expensive and vulnerable as their opposition.

I did not seek to "simplify" the battlefield, as Deitchman alleges, by ignoring "a responsive enemy, a 'dirty' battlefield and the practical military and engineering problems." Much of the article discusses just these remaining serious limitations to effective use of precision-guided weapons: reliability, maintenance, supply, cost, weather. Yet compared with the same limits of major capital items such as tanks, battleships, aircraft carriers and aircraft, precision-guided weapons appear much more cost-effective.

PAUL F. WALKER

Cambridge, Mass.

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 60 cents.

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

Name ______

New Address ______

______

______

Old Address ______

______

______

HOW THE NEW *Cimarron* BEATS THE IMPORTS AT THEIR OWN GAME.

	CIMARRON	AUDI 5000	BMW 320i	VOLVO GLE	SAAB 900S SEDAN
EPA MILEAGE RATINGS WITH STD. TRANS. HWY. EST./ EPA EST. MPG*	42/[26]	33/[19]	36/[25]	25/[16]	33/[21]
FRONT-WHEEL DRIVE	STANDARD	STANDARD	NOT AVAILABLE	NOT AVAILABLE	STANDARD
POWER-ASSISTED RACK AND PINION STEERING	STANDARD	STANDARD	RACK AND PINION ONLY	STANDARD	STANDARD
FOUR-SPEED MANUAL INCLUDING OVERDRIVE	STANDARD	STANDARD 5-SPEED	STANDARD 5-SPEED	STANDARD	STANDARD 5-SPEED
TACHOMETER	STANDARD	EXTRA COST	STANDARD	STANDARD	STANDARD
EPA PASSENGER COMPARTMENT VOLUME	89 CU. FT.	90 CU. FT.	82 CU. FT.	89 CU. FT.	89 CU. FT.
ALUMINUM ALLOY WHEELS	STANDARD	EXTRA COST	EXTRA COST	STANDARD	STANDARD
AIR CONDITIONING	STANDARD	EXTRA COST	EXTRA COST	STANDARD	DEALER INSTALLED OPTION
LEATHER-WRAPPED STEERING WHEEL	STANDARD	NOT AVAILABLE	EXTRA COST	DEALER INSTALLED OPTION	NOT AVAILABLE
LEATHER SEATING AREAS FOR FIVE	STANDARD	EXTRA COST	NOT AVAILABLE	STANDARD	NOT AVAILABLE
MSRP**	$12,131 (F.O.B.)	$11,540 (P.O.E.)	$13,105 (P.O.E.)	$15,590 (P.O.E.)	$12,700 (P.O.E.)

*Use estimated mpg for comparison. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Some Cadillacs are equipped with engines produced by other GM divisions, subsidiaries, or affiliated companies worldwide. See your dealer for details.

**Manufacturer's Suggested Retail Price including dealer prep. as of 7/7/81. Tax, license, destination charges and optional equipment additional. Destination charges vary by location and may affect this comparison. *Level of standard equipment varies.*

For years foreign car manufacturers have boasted about their gas mileage, standard features and interiors. But now, there's a car that beats these imports at their own game. Cimarron by Cadillac.

As the chart shows, Cimarron has features the imports have, plus Cadillac comfort and convenience, with reclining body-contoured front bucket seats and genuine leather seating areas for five. Cimarron has front-wheel drive, just as the Cadillac Eldorado and Seville do. It comes with Cadillac's exclusively tuned, road-hugging touring suspension and a four-speed manual transmission including overdrive. What's more, Cimarron behaves like a civilized car should. Nimble . . . easy to maneuver . . . with a smooth, refined ride.

If you've been thinking about buying an import, it's time to re-think your decision. It's time for Cimarron. Now available at all Cadillac dealers.

Cimarron BY CADILLAC

A NEW KIND OF CADILLAC
FOR A NEW KIND OF CADILLAC OWNER.

GM

THEN THE BEST PICTURES.

picture that's every bit as large as the sound. As for the picture quality, well, just look at the picture of Liza below. Hard to believe, it's an actual picture taken right off the screen.

But with Pioneer LaserDisc you don't just sit back and watch. For example, with the "How to Watch Pro Football" disc, you can go backwards, forwards, in fast motion, slow motion, stop motion, study it one frame at a time.

There are discs that teach you golf, tennis, cooking, step by step. Then there's The First National Kidisc. For the first time children learn at their own rate.

The only way to believe all this new technology is to see it. So we've arranged for you to actually take a LaserDisc player and several discs home with you to marvel at. Absolutely free. For the nearest dealer who can loan it to you, call us at **800-621-5199**. (In Illinois, 800-972-5855.)

PIONEER®
We bring it back alive.

Liza in Concert*

The Pioneer 50" Projection TV.

The Pioneer LaserDisc Player.

Optional Remote Control.

LaserVision

*This LaserVision mark certifies compatability with other laser-optical products bearing the mark

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1931: "Some eighteen months have passed since representatives of Great Britain, Japan and the United States affixed their signatures to the London Naval Treaty by which the 5–5–3 principle of the Washington treaty of 1922 on capital ships was extended to cover other categories. Recent figures given out by the Navy Department show that while the building programs of all the signatories lag behind the construction permissible for 1930 and 1931, that of the United States lags worst of all. The United States obtained parity with Great Britain—on paper—and now it is up to the nation to achieve it on the high seas. Popular demand should see to that. It is true, of course, that such a demand would place the Administration in an embarrassing position. With a steadily growing Treasury deficit Mr. Hoover has to consider national economy ahead of all other things."

"Scientists of the National Institute of Health are seeking to ascertain the crops having the highest pellagra-preventive values that may be grown most easily by farmers in the areas in which pellagra is prevalent. The nutrition specialists of the United States Public Health Service, working at the Institute, hope that their studies will result in a practical solution of the problem of preventing pellagra, a nutritional disease, which has become widespread in parts of the area. In conjunction with its study of the nutritive value of crops the Institute is attempting to concentrate—and if possible to isolate—the pellagra-preventive vitamin. The Institute, or Hygienic Laboratory, as it was then known, discovered the cause of pellagra. This discovery is probably one of the most significant steps forward in public health during the past decade. The late Dr. Joseph Goldberger of the Public Health Service found that pellagra was caused by lack of a certain nutritive substance in the diet. This finding threw a new light on pellagra and has made possible the beneficial work now being done by the Public Health Service in attempting to eradicate the disease."

"Hailed as a prodigy of American science, Linus Pauling, who at 30 has published nearly 50 papers on original research and who has won a full professorship at the California Institute of Technology, has been awarded the A. C. Langmuir Prize of the American Chemical Society. Dr. Langmuir, sponsor of the prize of $1,000, called Pauling 'a rising star, who may yet win the Nobel prize.' In singling out Pauling, he explained, American chemists, breaking with tradition, are honoring a scientist at the daybreak of his career instead of at the sunset. The work of Pauling has had to do with crystal structure, the quantum theory of the dielectric constant of gases, atomic and molecular structure, and the determination of the nature of chemical bonds."

SCIENTIFIC AMERICAN

NOVEMBER, 1881: "The Edison Electric Light Company has laid about three miles of conductor in an area three-quarters of a mile square in New York. When this district is complete, there will be 14 miles of conductor under the streets. These conductors will supply 16,000 lamps and 400 horse power for driving machinery. The conductors are of drawn copper and are supported throughout their entire length by insulating material in an iron pipe. Throughout the system the conductors are of the same size, and they are connected together at the corners of the blocks so as to increase the capacity of the system. The central lighting station is to be provided with 12 large Edison generators requiring 2,200 horse power. These machines are in the process of construction."

"During the past decade the production of breadstuffs in the United States, as shown in the latest census reports, has been nearly doubled. During the same period the exportation of breadstuffs has increased fourfold. It is now more than 10 times as great as it was 20 years ago and more than 20 times what it was 30 years ago. The enormous and wonderfully rapid increase in our grain crops is attributable to several causes. Primarily we have the invention and improvement of agricultural machinery, by which the cultivation of the great West has been made possible. Next we have the vast extension and improvement of our railway and water lines, making possible the profitable transportation of the large surplus to Eastern and foreign markets. With this extension of means has come an important lowering of freight charges, which has made it possible to place American grain in the markets of Europe at prices at which it can compete successfully with European grain, especially that from Russia, Hungary, Austria and Germany."

"An English statistician has lately brought out the following fact. It appears that 107,000 men, women and children lost their lives or were injured

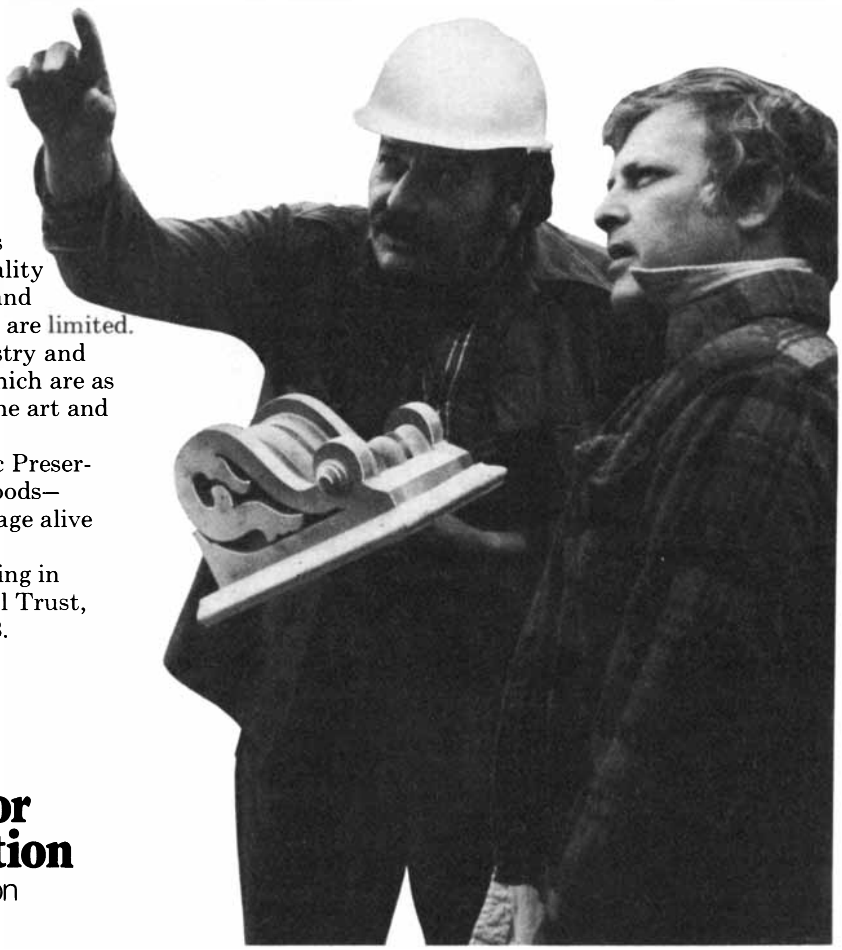
in English mines and factories, on railways and by boiler accidents during the four years preceding 1877. On this basis it is estimated that half a million workmen will lose their lives in 10 years—300,000 in mines, 70,000 on railways and 130,000 in factories. This aggregate is appalling and ought to be inquired into in this country as well as in England, but it is difficult to prescribe efficient legislative measures to meet the case."

"Ordinary human eyesight has certain imperfections. The eye requires a certain definite, though short, time in which to take in the visual image. Then there are cases of an opposite kind, in which the eye fails to recognize objects or their details because of their exceeding faintness. For instance, a certain photograph of the sun's surface has been so rapidly taken that it shows details no astronomer has ever actually seen. On the other hand, a picture of the Great Nebula in Orion was two hours and 20 minutes forming itself on the retina of the photographic eye. We see now but the beginning of the use of the photographic eye, which can see in the five-thousandth part of a second if need be, or if need be can rest its gaze for many hours on the same object, seeing more and more as minute after minute passes on."

"The invention of wood pulp has revolutionized paper making and paper prices. It has brought good books, good newspapers and writing paper within the means of thousands of the common people who could never have afforded such luxuries had rags remained the only available material for papers of good quality. In the busy manufacturing town of Manayunk, Pa., the operations of wood pulp making can be seen on a large scale in the extensive works of the American Wood Paper Company, where 12,000 cords of poplar from the forests of Virginia are annually converted into paper fiber. The capacity of these works is 18 to 20 tons of fine pulp every 24 hours. The product is highly esteemed wherever it has found its way for its superior quality. Large quantities are lately being shipped to France."

"For some time past there has been on exhibition at the Jardin d'Acclimatation in Paris a party of Fuegians, who have attracted great attention from visitors. These people are inhabitants of Tierra del Fuego in South America. When we read in the works of travelers a description of their country, we are no longer surprised at the deep degradation of these people. The Fuegians in Paris, 11 in number (four men, four women and three children), were brought thither by Mr. Waalen, after he had deposited some $3,000 in the hands of the Chilian Governor of Punta-Arenas as a security for their safe return to their native land after being exhibited in Europe."


STEINHARDT · HANSON
OFFICE PRODUCTS
BUSINESS FURNITURE
OFFICE MACHINES
GIFTS · GREETING CARDS
DOUGLAS



THE AUTHORS

PHILIP M. HAUSER ("The Census of 1980") is Lucy Flower Professor Emeritus of Urban Sociology and director emeritus of the Population Research Center at the University of Chicago. From 1938 to 1947 he was deputy director of the Bureau of the Census and in 1949–50 he was acting director. Hauser was the U.S. representative to the Population Commission of the United Nations from 1947 to 1951. He received his undergraduate education and his Ph.D. at the University of Chicago. This article is Hauser's fifth on the census for SCIENTIFIC AMERICAN; the first appeared after the census of 1950.

MICHAEL LAMPTON ("The Microchannel Image Intensifier") is a research physicist at the Space Sciences Laboratory of the University of California at Berkeley and a payload specialist in training for flight duty on the *Spacelab 1* mission. His degrees are in physics: a B.Sc. from the California Institute of Technology in 1962 and a Ph.D. from Berkeley in 1967. He has been associated with several space missions, including the Apollo-Soyuz flight in 1975. In his spare time he skis and designs loudspeakers.

PAUL HOWARD-FLANDERS ("Inducible Repair of DNA") is professor in the department of molecular biophysics and biochemistry and the department of therapeutic radiology at the Yale University School of Medicine. Born in England, he studied physics and mathematics at the Imperial College of Science and Technology. He got his Ph.D. from the University of London, where he worked under Louis H. Gray and George R. Popjak of the Medical Research Council at Hammersmith Hospital. Howard-Flanders' early work included the design of the first microwave linear accelerator employed in radiotherapy. In 1958 he was a lecturer in biophysics at the University of California at Berkeley. The following year he moved to Yale. His research has focused on the mechanisms by which cells repair damaged DNA; he is a codiscoverer of excision repair and postreplication repair.

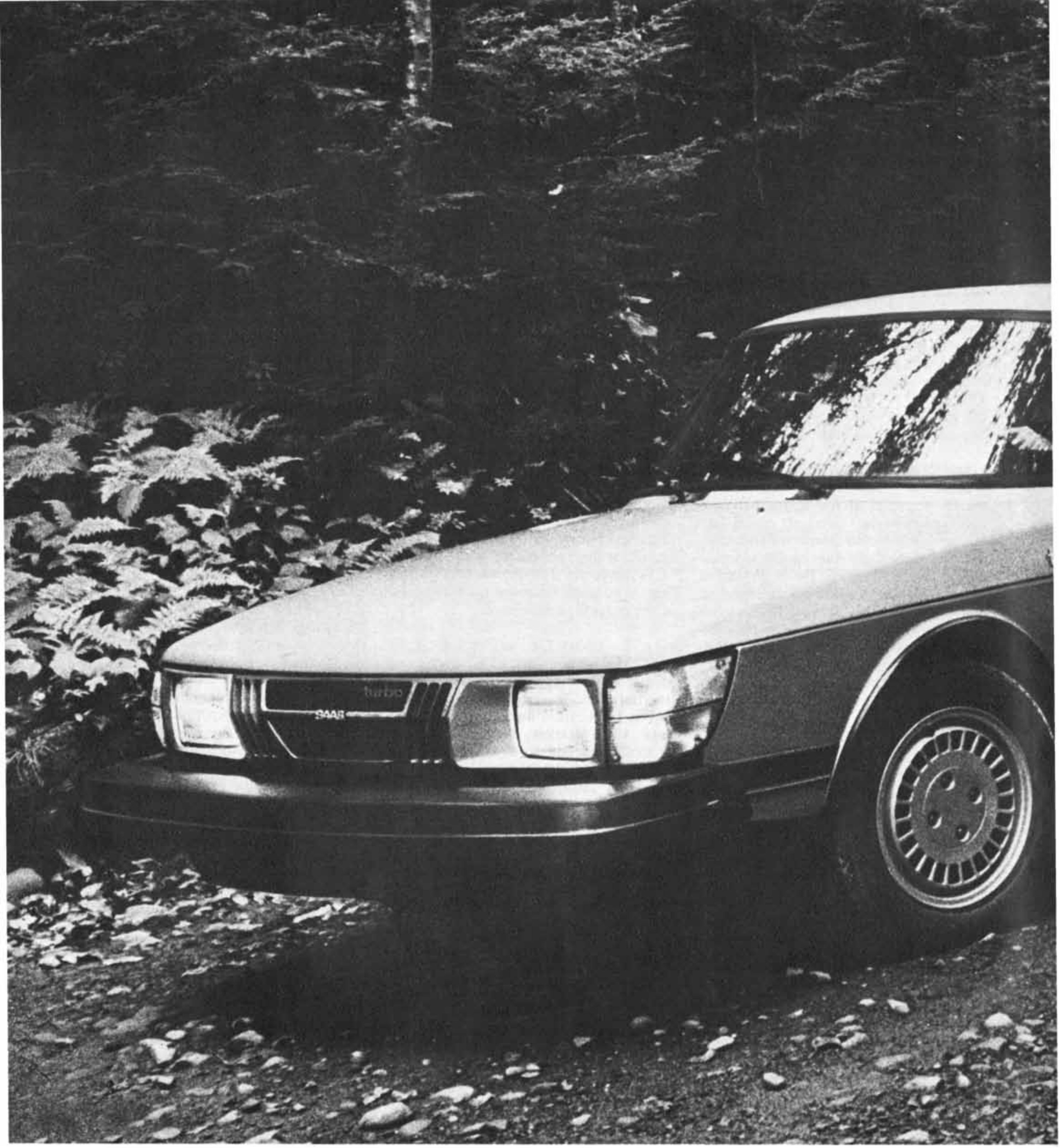
JAMES B. POLLACK and JEFFREY N. CUZZI ("Rings in the Solar System") work at the Ames Research Center of the National Aeronautics and Space Administration. Pollack is also chief scientist of the NASA Aerosol Climatic Effects Program. He earned his bachelor's and master's degrees in physics respectively at Princeton University in 1960 and the University of California at Berkeley in 1962; he received his Ph.D. in astronomy from Harvard University in 1965. Five years later he joined NASA, where he participated in the Pioneer Venus mission in 1978, the *Mariner 9* mission in 1971, the Viking missions to Mars in 1976 and the Voyager mission to Saturn in 1980. Pollack's work has earned him two medals from NASA and one award from the American Institute of Aeronautics and Astronautics. Cuzzi did his undergraduate work in engineering physics at Cornell University and went on to obtain his Ph.D. in planetary science from the California Institute of Technology in 1972. Before going to Ames in 1974 he held research positions at the University of Massachusetts and at the Grand Ronde Center for Basic Studies. He writes: "I have studied the rings of Saturn and Jupiter for several years, both observationally and theoretically, most recently as a member of the Voyager Imaging Science team. I am particularly interested in what the dynamical behavior of rings may teach us about the formation of the planets themselves."

J. A. BARKER and DOUGLAS HENDERSON ("The Fluid Phases of Matter") have been staff members of the International Business Machines Research Laboratory in San Jose, Calif., since 1969. Barker, born in Western Australia, earned three degrees at the University of Melbourne: a B.Sc. in physics in 1944, a B.A. in mathematics in 1945 and a D.Sc. in chemistry in 1958. From 1950 to 1967 he worked in the Chemical Research Laboratories of the Commonwealth Scientific and Industrial Research Organization (C.S.I.R.O.) in Melbourne. The year before he went to IBM he was professor of applied mathematics and physics at the University of Waterloo in Ontario. He is a fellow of the Australian Academy of Science and a Fellow of the Royal Society of London. "My research interests," he writes, "have been in statistical mechanics, the study of intermolecular forces and the theory of liquids, solids, gases and surfaces." Within these fields, Barker notes, he has concentrated on "the development of computational techniques and computer-simulation methods." Henderson, who was born in Canada, obtained his bachelor's degree in mathematics at the University of British Columbia in 1956 and his doctoral degree in physics from the University of Utah in 1961. Like Barker, he was professor of applied mathematics and physics at the University of Waterloo (from 1964 to 1969) and a worker in the Chemical Research Laboratories of the C.S.I.R.O., where, he reports, "many of the ideas in this article were formulated." A fellow of the Institute of Physics of the United Kingdom, the American Physical Society and the American Institute of Chemists, Henderson is interested in "statistical mechanics with particular emphasis on the theory of liquids and liquid interfaces. A further interest is in the theory of disordered solids."

H. FREDERIK NIJHOUT ("The Color Patterns of Butterflies and Moths") is associate professor of zoology at Duke University. Born in the Netherlands, he received his bachelor's degree in biology at the University of Notre Dame in 1970 and his Ph.D. from Harvard University in 1974. He worked as a postdoctoral fellow at the University of Washington in Seattle and as a staff fellow at the National Institute of Allergy and Infectious Diseases before going to Duke in 1976. "My work on color patterns," he reports, "is one aspect of a general interest in the developmental biology of insects. I am also doing research on the physiology of molting and metamorphosis and am particularly interested in the physiological processes that regulate the secretion of insect developmental hormones."

FLORENCE MOOG ("The Lining of the Small Intestine") is Charles Rebstock Professor of Biology at Washington University. She was graduated from New York University in 1936 and went on to Columbia University to earn her A.M. in 1938 and her Ph.D. in 1944. At Columbia, she reports, "I was attracted to experimental embryology because it seemed on the verge of solving some of the most profound problems of biology (still unsolved in 1981). Virtually all my research has been in the field of development, particularly of vertebrates. In the late 1940's I turned my attention to the embryonic intestine, and quite unintentionally I became a pioneer in what has subsequently become the very lively area of research on the fetal and suckling intestine. As for my nonscientific interests, I am a dedicated opera buff and an avid reader of history, biography and novels. I consider myself to be at least a minor authority on Anthony Trollope's exceptionally sympathetic understanding of the position of women in Victorian society."

JAMES A. TUCK and ROBERT GRENIER ("A 16th-Century Basque Whaling Station in Labrador") are respectively professor of archaeology at the Memorial University of Newfoundland and head of the Marine Excavation Unit of Parks Canada. Tuck received his Ph.D. from Syracuse University in 1968 and has been at Memorial University since then. His archaeological field work has been done mainly in Newfoundland and Labrador. Grenier got his master's degree at Laval University in Quebec and did his first archaeological work at the Cape Breton Island fortress site of Louisbourg.


turbo
SAAB

How simple it would be to turn to the Audi, the BMW or the Mercedes and declare that herein lies all the performance you're likely to encounter in a sports sedan.

You wouldn't be alone. Car buyers following other car buyers claiming to be driving enthusiasts have trodden that well-worn path for years.

The editors of *Road & Track*, on the other hand, have long earned the esteem of the automotive world, not to mention their very livelihood, by looking into these things for themselves.

And they recently voted the Saab 900 Turbo the best sports sedan for the eighties. Unanimously.

Now, you may still end up buying whatever car you assumed you were going to buy in the first place.

But before you actually plunk down the money, why not indulge that little twinge of curiosity you already feel and ask yourself what the gentlemen from *Road & Track* magazine *saw* in that car.

All you need is an open mind, and an open road, to find out.

SAAB

The most intelligent car ever built.

METAMAGICAL THEMAS

Strange attractors: mathematical patterns delicately poised between order and chaos

by Douglas R. Hofstadter

"You can't know how happy
I am that we met,
I'm strangely attracted to you."
—COLE PORTER,
"It's All Right with Me"

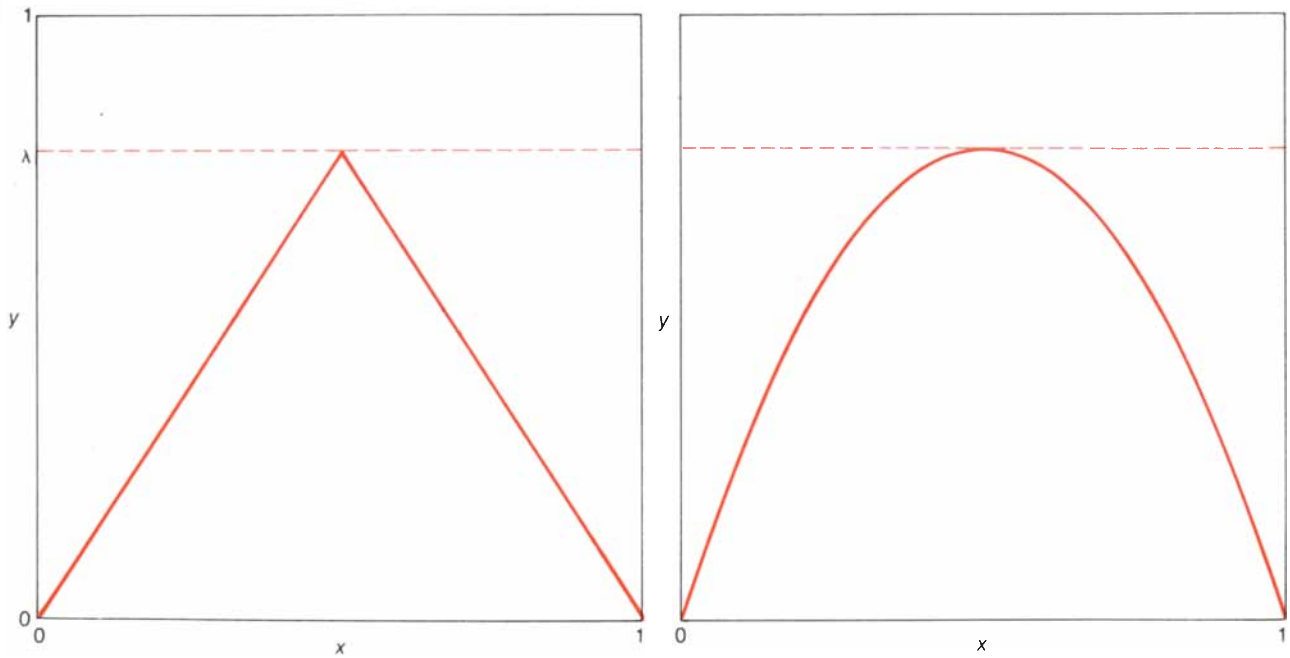
A few months ago, while walking through the corridors of the physics department of the University of Chicago with a friend, I spotted a poster announcing an international symposium titled "Strange Attractors." My eye could not help but be strangely attracted by this odd term, and I asked my friend what it was all about. He said it was a hot topic in theoretical physics these days. As he described it, it sounded quite wonderful and mysterious.

I gathered that the basic idea hinges on looking at what might be called mathematical feedback loops: expressions whose output can be fed back into them as new input, the way a loudspeaker's sounds can cycle back into the microphone and come out again. From the simplest of such loops, it seemed, both stable patterns and chaotic patterns (if this is not a contradiction in terms!) could emerge. The difference was merely in the value of a single parameter. Very small changes in the parameter could make all the difference in the world as to the orderliness of the behavior of the loopy system. This image of order melting smoothly into chaos, of pattern dissolving gradually into randomness, was exciting to me.

Moreover, it seemed that some unexpected "universal" features of the transition into chaos had recently been unearthed, features that depended solely on the presence of feedback and that were virtually insensitive to other details of the system. This generality was important because any mathematical model featuring a gradual approach to chaotic behavior might provide a key insight into the onset of turbulence in all kinds of physical systems. Turbulence, in contrast to most phenomena successfully understood in physics, is a nonlinear phenomenon: two solutions to the equations of turbulence do not add up to a new solution. Nonlinear mathematics is much less well understood than linear mathematics, which is why a good mathematical description of turbulence has eluded physicists for a long time, and would be a fundamental breakthrough.

When I later began to read about these ideas, I found out that they had actually grown out of many disciplines simultaneously. Pure mathematicians had begun studying the iteration of nonlinear systems by using computers. Theoretical meteorologists and population geneticists, as well as theoretical physicists studying such diverse things as fluids, lasers and planetary orbits, had independently come up with similar nonlinear mathematical models featuring chaos-pregnant feedback loops and had studied their properties, each group finding some quirks that the others had not found. Moreover, not only theorists but also experimentalists from these widely separated disciplines had simultaneously observed chaotic phenomenona that share certain basic patterns. I soon saw that the simplicity of the underlying ideas gives them an elegance that in my opinion rivals that of some of the best of classical mathematics. Indeed, there is an 18th- or 19th-century flavor to some of this work that is refreshingly concrete in this era of staggering abstraction.

Probably the main reason these ideas are only now being discovered is that the style of exploration is entirely modern: it is a kind of experimental mathematics, in which the digital computer plays the role of Magellan's ship, the astron-



Examples of "folded" functions can generate a sharp peak (left) *and a parabola* (right)


THE TASTE BEYOND
12-YEAR-OLD SCOTCH
Discover more.
Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.
Most premium Scotch is blended and depends on several whiskies for taste and smoothness.
Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.
Try the taste beyond premium Scotch.
The Glenlivet
Unchanged since 1824.
The
GLENLIVET
AGED 12 YEARS
The
GLENLIVET
12 YEARS OLD
Unblended all malt
Scotch Whisky
Distilled and Bottled in Scotland
GEORGE & J. G. SMITH LIMITED

omer's telescope and the physicist's accelerator. Just as ships, telescopes and accelerators must be ever larger, more powerful and more expensive in order to probe ever more hidden regions of nature, so one would need computers of ever greater size, speed and accuracy in order to explore the remoter regions of mathematical space. By the same token, just as there was a golden era of exploration by ship and of discoveries made with telescopes and accelerators, characterized by a peak in the ratio of new secrets uncovered to money spent, so one would expect there to be a golden era in the experimental mathematics of these models of chaos. Perhaps this era has already occurred, or perhaps it is occurring right now. And perhaps after it we will witness a flurry of theoretical work to back up these experimental discoveries.

In any case it is a curious and delightful brand of mathematics that is being done. This way of doing mathematics builds powerful visual imagery and intuitions directly into one's understanding. Bypassing the traditional theorem-proof-theorem-proof brand of mathematics by exploiting the power of computers allows one to arrive quickly at empirical observations and discoveries that may reinforce each other and form such a rich and coherent network of ideas that in the long run it may turn out to be easier to find the requisite proofs because the conceptual territory will have been so thoroughly mapped out in advance. One of the strongest proponents of this style of mathematizing has been Stanislaw M. Ulam, who, when computers were still young, turned them loose on problems of nonlinear iteration as well as on problems from many other branches of mathematics. It is from Ulam's early studies with Paul Stein that many of the ideas to be sketched here follow.

So much for romance. Let us now work our way up to the notion of a strange attractor by beginning with the more basic concept of an attractor. The entire field is founded on one concept: the iteration of a real-valued mathematical function, that is, the behavior of the sequence of values x, $f(x)$, $f(f(x))$, $f(f(f(x)))$, . . . , where f is some interesting function. The initial value of x can be called the "seed." The idea is to feed f's output back into f as new input over and over again, to see if some kind of pattern emerges.

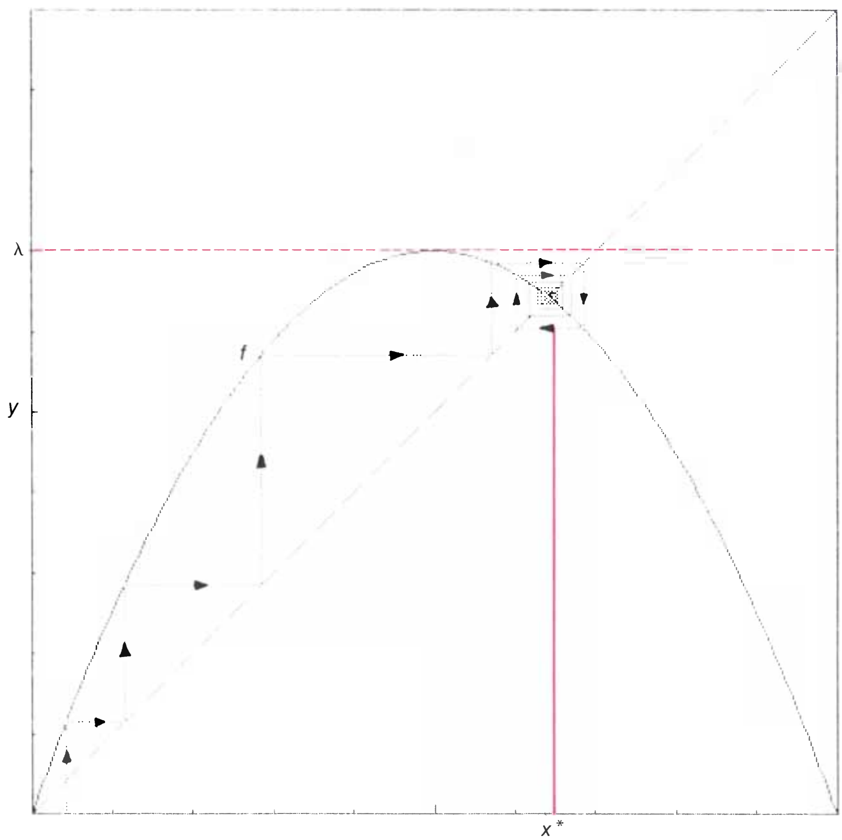
An interesting and not too difficult problem concerning the iteration of a function is this: Can you invent a function p with the property that for any real value of x, $p(x)$ is also real, and where $p(p(x))$ equals $-x$? The condition that $p(x)$ be real is what gives the problem a twist; otherwise the function $p(x) = ix$ (where i is the square root of -1) would work. In fact, you can even think of the challenge as that of finding a real-valued "square root of the minus sign." A related problem is to find a real-valued function q, whose property is that $q(q(x)) = 1/x$ for all x other than zero. Note that no matter how you construct p and q, each will have the property that, given any seed, repeated iteration creates a cycle of length four.

Now, more generally, what kinds of functions, when repeatedly iterated, are likely to exhibit interesting cyclic or near-cyclic behavior? A simple function such as $3x$ or x^3, when iterated, does not do anything like that. The nth iteration of $3x$, for example, is $3 \times 3 \times 3 \times \ldots \times 3 \times x$ with n 3's—that is, $3^n x$—and the nth iteration of x^3 is just $(((x^3)^3)^3 \ldots)^3$ with n 3's again, which amounts to x^{3^n}. Nothing cyclelike here; the values just keep going up and up and up. To reverse this trend one needs a function with some sort of switchback—a little zigzag or twist. A more technical way of putting it is that one needs a *nonmonotonic* function: a function whose graph is folded, that is, it starts moving one way—say upward—and then bends back the other way—say downward.

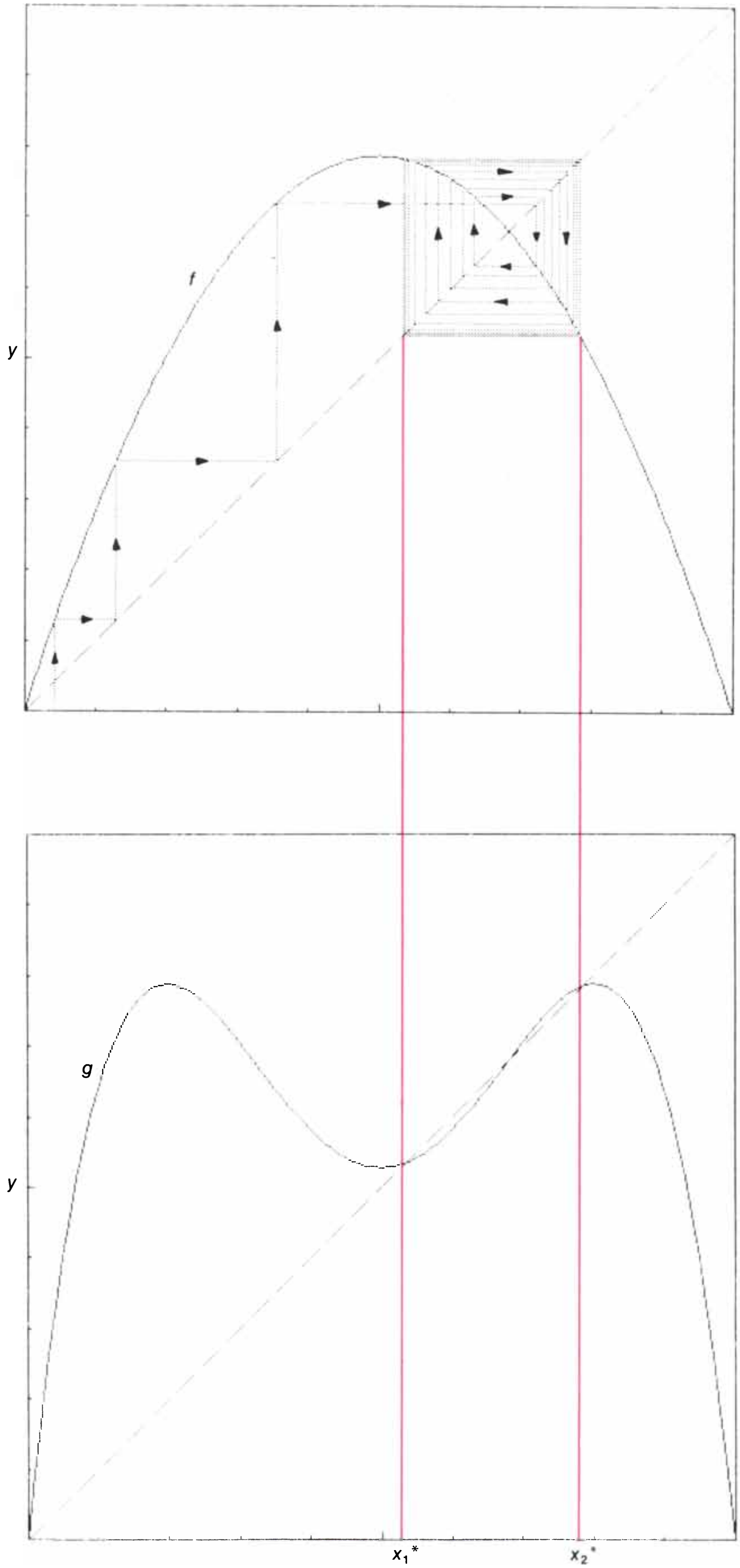
In the figure at the left in the illustration on page 22 we have a sawtooth with a sharp point at its top, and in the figure at the right a smoothly bending parabolic arc. Each of them rises from the origin, eventually reaches a peak height called λ and then comes back down for a landing on the far side of the interval. Of course there are uncountably many shapes that rise to height λ and then come back down, but these two are among the simplest. And of the two the parabola is perhaps the simpler, or at least the more mathematically appealing. Its equation is $y = 4\lambda x(1 - x)$, with λ not exceeding 1.

We shall allow input—values of x—only between 0 and 1. As the graph shows, for any x in that interval the output—y—always is between 0 and λ. Therefore the output value can always be fed back into the function as input, which ensures that repeated iteration will always be possible. When you repeatedly iterate a "folded" function like this, the successive y values you produce will sometimes go up and sometimes down—always hovering, of course, between 0 and λ. The fold in the graph guarantees interesting effects when the function is iterated—as we shall see.

It turns out that the spectacular differences in the degree of regularity of patterns I mentioned above are due to variations in the setting of what we might call the "λ knob." Depending on the value the knob is set at, the function yields an incredible variety of "orbits," that is, of sequences x, $f(x)$, $f(f(x))$, and so on. In particular, for λ below a certain critical value ($\lambda_c = 0.892486417967\ldots$) the orbits are all regular and patterned (although there are various *degrees* of patternedness; generally the lower λ is, the



Setting the "λ knob" at 0.7 for the function **f(x)** *gives rise to this graph*



A spiral to a stable 2-cycle is at the top; the elements of the cycle, x_1* *and* x_2*, *are at the bottom*

more simply the orbit is patterned), but for λ at or beyond this critical value, hold onto your hat! An essentially chaotic sequence of values will be traced out by the values x, $f(x)$, $f(f(x))$, . . ., no matter what positive seed value of x you choose. In the case of the parabola the critical role played by varying the λ knob seems to have been first realized by P. J. Myrberg in the early 1960's, but his work was published in an obscure journal and did not attract much attention. Some 10 years later Nicholas C. Metropolis, Paul Stein and Myron Stein rediscovered the importance of the λ knob not only for the parabola but also for many functions. Indeed, they discovered that so far as certain topological properties were concerned the function did not matter—only the value of λ did. This has come to be called "structural universality."

In order to see how such a nonintuitive dependence on the setting of the λ knob comes about one must develop a visual sense for the process of iterating $f(x)$. This is readily done. Suppose we set λ to 0.7. The graph of $f(x)$ appears in the illustration on page 24. In addition the line $y = x$ appears as a 45-degree broken line. (This graph and most of the others in this article were produced on a small computer by Mitchell J. Feigenbaum of the Los Alamos National Laboratory.)

Consider the two x values where the 45-degree line and the curve intersect. They are at $x = 0$ and $x = 9/14 \cong 0.643$. Let us designate the nonzero value as x^*. By construction, then, $f(x^*)$ equals x^*, and repeated iteration of f at this x value will get you into an infinite loop. The same happens if you start iterating at $x = 0$: you get stuck in an endless loop. There is, however, a significant difference between these two "fixed points" of f. It is best indicated by taking some other initial value of x, say one close to 0.04, as is shown in the same illustration. Call this starting x value x_0. There is an elegant graphical way to generate the orbit of any seed x_0. A vertical line at x value x_0 will hit the curve at height $y_0 = f(x_0)$. To iterate f we must draw a new vertical line located at the new x value equal to this y value. This is where the 45-degree line $y = x$ comes in handy. Staying at height y_0, we simply move over horizontally until we hit that 45-degree line. Then, since along this line y equals x, both x and y equal y_0. Let us call this new x value x_1. We now draw a second vertical line. This one will hit the curve at height $y_1 = f(x_1) = f(y_0) = f(f(x_0))$. Now we just repeat.

In brief, iteration is realized graphically by a simple recipe: (1) Move vertically until you hit the curve and (2) then move horizontally until you hit the diagonal line. Repeat steps (1) and (2) over and over again.

The results of this procedure with seed $x_0 = 0.04$ are also shown in the il-

lustration. We are led in a merry chase 'round and 'round the point whose x coordinate and y coordinate are x^*. Gradually we close down on that point. Thus x^* is a special kind of fixed point because it *attracts* iterated values of $f(x)$. It is the simplest example of an attractor: every possible seed (except 0) is drawn, through iteration of f, to this stable x value. This x^* is therefore called a stable fixed point. In contrast, 0 is a repellent or unstable fixed point, since the orbit of any initial x value, even one infinitesimally removed from 0, will proceed to move *away* from 0 and toward x^*. Note that sometimes the iterates of f will overshoot x^*, sometimes they will fall short—but they inexorably draw closer to x^*, zeroing in on it like swallows returning to Capistrano.

What accounts for this radical qualitative difference between the two fixed points (0 and x^*) of f? A moment's look at the illustration will show that it is the fact that at 0 the curve is sloped too steeply. In particular, the slope there is greater than 45 degrees. It is the local slope of the curve that controls how far you move horizontally each time you iterate f. Whenever the curve is steeper than 45 degrees (whether it is going uphill or downhill) it tends to pull you farther and farther away from your starting point as you repeatedly iterate by rules (1) and (2). Hence the criterion for the stability of a fixed point is: the slope at the fixed point should be less than 45 degrees. Now, this is the case for x^* when λ equals 0.7. In fact, the slope there is about 41 degrees, whereas at 0 it is much greater than 45 degrees.

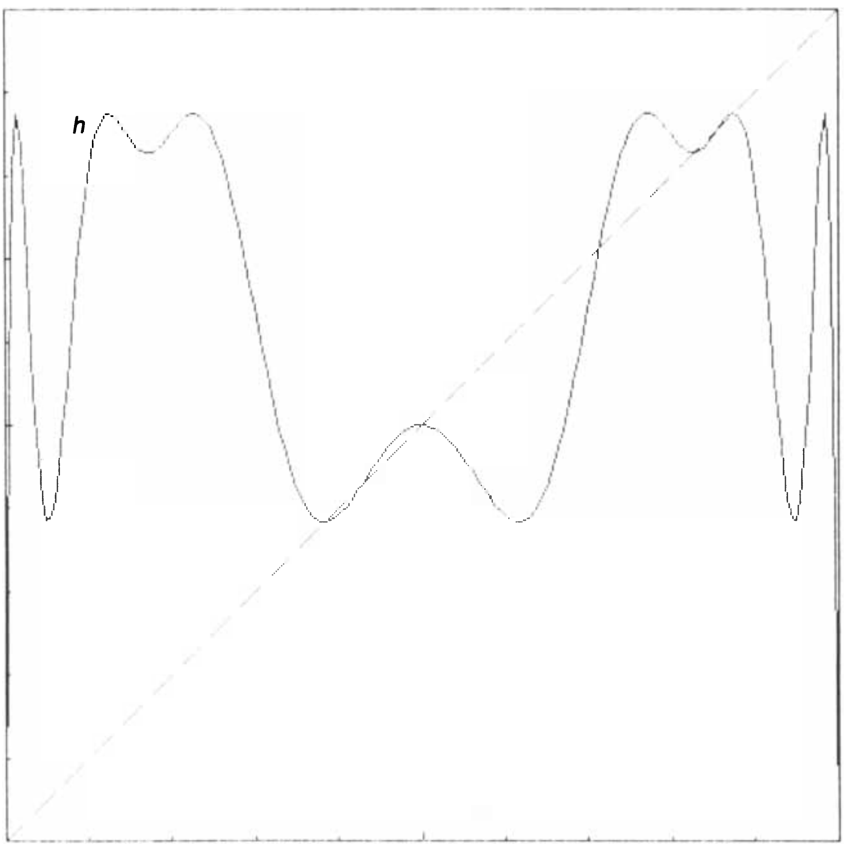
What happens if we increase λ? The position of x^* (x^* is by definition the x coordinate of the point where the curve f and the line $y = x$ intersect) will change, and the *slope* of f at x^* will increase as well. What happens when the slope hits 45 degrees or exceeds it? This occurs when λ is 3/4. We will call this special value of the λ knob Λ_1. Let us look at a picture for a slightly greater λ-knob setting, namely $\lambda = 0.785$ [*see illustration on page 26*].

What if we begin with some random seed instead, again say $x = 0.04$? The resulting orbit is shown in the top half of the illustration. As you can see, a very pretty thing happens. At first the values move up toward the vicinity of x^* (now an unstable fixed point of f), but then they spiral gradually outward and settle down smoothly to a kind of square dance converging on two special values x_1^* and x_2^*. This elegant oscillation is called a 2-cycle, and the pair of x values that constitute it (x_1^* and x_2^*) is again called an attractor, in particular an attractor of period two. This term means that our 2-cycle is stable: it attracts x values from far and wide as f is iterated. The orbit for any positive seed value (except x^* itself) will eventually fall into the same dance. That is, it will asymptotically approach the perfect 2-cycle composed of the points x_1^* and x_2^*, although it will never quite reach it exactly. From a physicist's point of view, however, the accuracy of the approach soon becomes so great that one can just as well say that the orbits have been "trapped" by the attractor.

An enlightening way to understand this is to look at a graph of a new function made from the old one. Consider the graph of $g(x) = f(f(x))$, shown in the bottom half of the illustration. This two-humped camel is called the *iterate* of f. First of all observe that any fixed point of f is also a fixed point of g, so that 0 and x^* will be fixed points of g. But secondly observe that since $f(x_1^*)$ equals x_2^*, and conversely $f(x_2^*)$ equals x_1^*, g will have two *new* fixed points: $g(x_1^*) = x_1^*$ and $g(x_2^*) = x_2^*$. Graphically, x_1^* and x_2^* are easily found: they are intersection points of the 45-degree line with the two-humped graph of $g(x)$. There are four such points (0 and x^* being the other two). As we have seen, the criterion for the stability of any fixed point under iteration is that the slope at that point should be less than 45 degrees. Here we are concerned with fixed points of g, and hence with g's slope (as distinguished from f's slope). Indeed, in the same illustration you can clearly see that at 0 and at x^*, g is sloped more steeply than 45 degrees, whereas at both x_1^* and x_2^*, g's slope is less than 45 degrees. In fact, quite remarkably, not only are both slope values less than 45 degrees, but also, as it turns out through a simple bit of calculus, they are equal (or "slaved" to each other, as it is sometimes put).

We have now seen an attractor of period one get converted into an attractor of period two at a special value of λ ($\lambda = 3/4$). Precisely at that value the single fixed point x^* splits into two oscillating values, x_1^* and x_2^*. Of course they coincide at "birth," but as λ increases they separate and draw farther and farther apart. This increase of λ will also cause g's slope at these two stable fixed points (of g) to get steeper and steeper until finally, at some value of λ, g, like its progenitor f, will reach its own breaking point (that is, the identical slopes at both x_1^* and x_2^* will exceed 45 degrees), and each of these two attracting points will break up, spawning its own local 2-cycle. (Actually the cycles are 2-cycles only as far as g is concerned; for f the new points are elements of an attractor of period four. You must be careful to keep f and g straight in your mind!) These two splittings will happen at exactly the same moment (that is to say, at the same λ-knob setting), since the value of the slope of g at x_1^* is slaved to the value of the slope at x_2^*. This λ-knob setting will be called Λ_2, and it has the value of 0.86237....

Here, as with a joke, you may antici-



Graph of **h(x)** *at a λ-knob setting of 0.87 is even bumpier than that of* **g(x)**

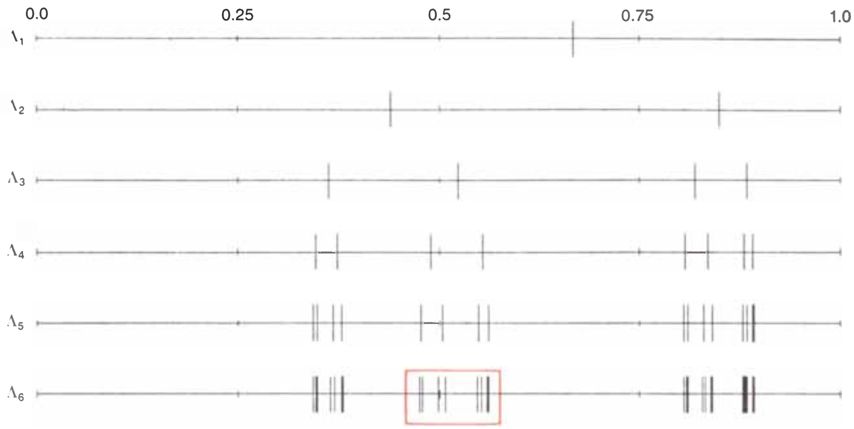

On a scale of one to ten,
your friends deserve a twelve.
Canadian Club
"The Best In The House".
To send gifts of Canadian Club by phone, dial 800-528-6148
6 Years Old. Imported in Bottle from Canada by Hiram Walker Importers Inc., Detroit, Mich. 86.8 Proof. Blended Canadian Whisky. ©1981

pate the punch line by the time you have heard the theme and one variation. Hence by now you have probably surmised that at some new value Λ_3 all four points in f's attractor will simultaneously fission, yielding a periodic attractor consisting of eight points; and thereafter this pattern will go on and on, doubling and redoubling as various special λ-knob settings are reached and passed. If this is your guess, you are entirely correct, and the underlying reason is the same each time: the (identical) slopes at all the stable fixed points of some graph reach the critical angle of 45 degrees. In the case of the first fission (at Λ_1) it was the slope of f itself at the single point x^*. The next fission was due to the slopes at g's two stable fixed points $x_1{}^*$ and $x_2{}^*$ simultaneously reaching 45 degrees. Analogously, Λ_3 is that value of λ at which the slope of $h(x) = g(g(x)) = f(f(f(f(x))))$ hits 45 degrees simultaneously at the four stable fixed points of h. And so it goes. The illustration that appears on page 28 shows the bumpy appearance of $h(x)$ at a λ value of approximately 0.87.

In the illustration on this page the locations on the x axis of the stable fixed points of f are shown for Λ_1 through Λ_6 (by which time there are 32 of them, some clustered so closely that they cannot be distinguished). The points are pictured just at the moment of their becoming unstable, each one like a cell on the verge of division. Notice the neat pattern in the distribution of the attracting points. Looking at these graphs of the spacings of the elements of the successive period-doubled attractors of f, you can see that each line can be made from the one above it through a recursive geometric scheme whereby each point is replaced by two "twin" points below it. Each local clustering pattern of points echoes the global clustering pattern, simply reduced in scale (and also, in alternating local clusters, left and right are reversed). For example, in the bottom line a local group of eight points has been outlined in color. Notice how the group of points is like a miniature version of the global pattern two lines above it.

The discovery of this recursive regularity, made on a little calculator by Feigenbaum, is one of the major recent advances in the field. It states in particular that to make line $n + 1$ from line n you simply let each point on line n give birth to "twins." The new generation of points should be packed in about 2.5 times more densely than the old generation was. More exactly stated, the distance between new twins should be α times smaller than the distance between their parent and its twin, where α is a constant, approximately equal to 2.5029078750958928485.... This rule holds with greater and greater accuracy the larger n becomes.

What about the values of the Λ's? Are they headed asymptotically toward 1? Surprisingly enough, no. These Λ values are quickly converging on a particular critical value λ_c, of size roughly 0.892486418.... And their convergence is remarkably smooth, in the sense that the distance between successive Λ's is shrinking geometrically. More precisely, the ratio $(\Lambda_n - \Lambda_{n-1})/(\Lambda_{n+1} - \Lambda_n)$ approaches a constant value called δ by Feigenbaum, its discoverer, but more often referred to as Feigenbaum's number by others. Its value is approximately 4.669201660910299097....



Stable attractors become unstable and "fission" at a series of Λ-knob settings **(top to bottom)**

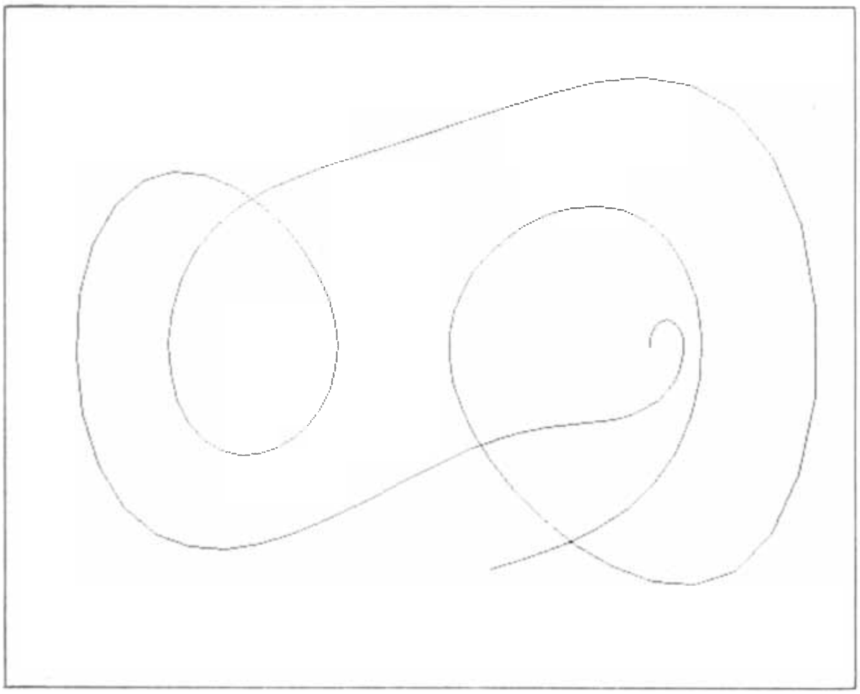
In short, as λ approaches λ_c, at special λ values predicted by Feigenbaum's constant δ, f's attractor doubles in population, and its increasingly many elements are geometrically arranged on the x axis according to a simple recursive plan, the main determining parameter of which is Feigenbaum's other constant, α.

Then for λ beyond λ_c—called the chaotic regime—the results of iterating f can for some seed values yield orbits that converge to no finite attractor. These are aperiodic orbits. For most seed values the orbit will remain periodic; the periodicity will, however, be very hard to detect. First of all, the period will be extremely high. Secondly, the orbit will be much more chaotic than before. A typical periodic orbit, instead of quickly converging to a geometrically simple attractor, will meander all over the interval [0,1], and its behavior will appear indistinguishable from total chaos. Such behavior is termed ergodic. Furthermore, neighboring seeds may, within a very small number of iterations, give rise to utterly different orbits. In short, a statistical view of the phenomena becomes considerably more reasonable beyond λ_c.

Now, what do such concepts as the iteration of folded functions, period doubling, chaotic regime and so on have to do with the study of turbulence in hydrodynamic flow, the erratic population fluctuations in predator-prey relations and the instability of laser modes? The basic idea is embedded in the contrast between laminar flow and turbulent flow. In a peacefully flowing fluid the flow is laminar—a soft and gentle word that means that all the molecules in the fluid are moving like cars on a multilane freeway. The key features are (1) that each car follows the same path as its predecessor and (2) that two nearby cars, whether they are in the same lane or in different ones, will, as time passes, slowly separate from each other—in essence in proportion to the difference in their velocities, that is, linearly. These features also apply to molecules of fluid in laminar flow; there the lanes are called streamlines or laminas.

In contrast, when a fluid is churned up by some external force, this smooth behavior turns into turbulent behavior, as is seen in breakers at the beach and cream being stirred into coffee. Even the word "turbulent" sounds much harsher and more angular than the soft word "laminar." Here the image of a multilane freeway no longer holds; the streamlines separate from each other and tangle in the most convoluted of ways. In such systems there are eddies and vortexes and all kinds of unnamable whorls on many size scales at once, and consequently two points that were initially very close may soon wind up in totally different regions of the fluid. Such quickly diverging paths are the hallmark of turbulence. The distance between points can increase exponentially with time, instead of just linearly, and the coefficient of time in the exponent is called the Lyapunov number. When one speaks of chaos in turbulent flow, it is this rapid, nearly unpredictable separation of neighbors that is meant. Such behavior is strikingly reminiscent of the rapid separation, in the chaotic regime of λ, of two orbits whose seeds might originally have been very close together.

This suggests that the scenario (as it is called) by which pretty, periodic orbits gradually give way to the messy, chaotic orbits of our parabolic function might conceivably be mathematically identical to the scenario underlying the transi-



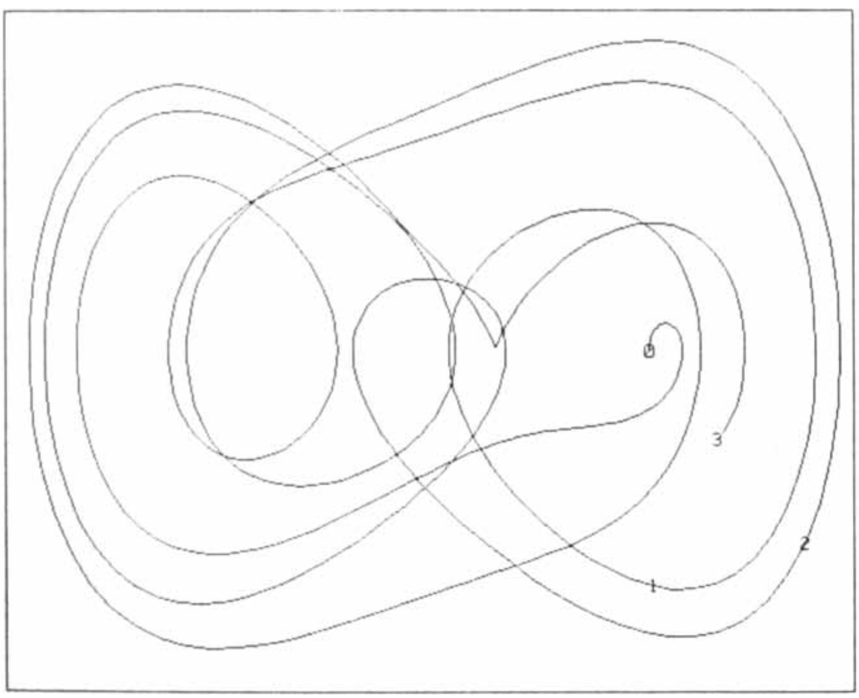
A curve generated by Duffing's equation resembles the shadow of a fly's path on a wall

tion to turbulence in a fluid or other system. Exactly how this connection is established, however, requires some more detailed setting of context. In particular, we must briefly consider how the spatiotemporal flow of a fluid or some other entity, such as population density or money, is mathematically modeled.

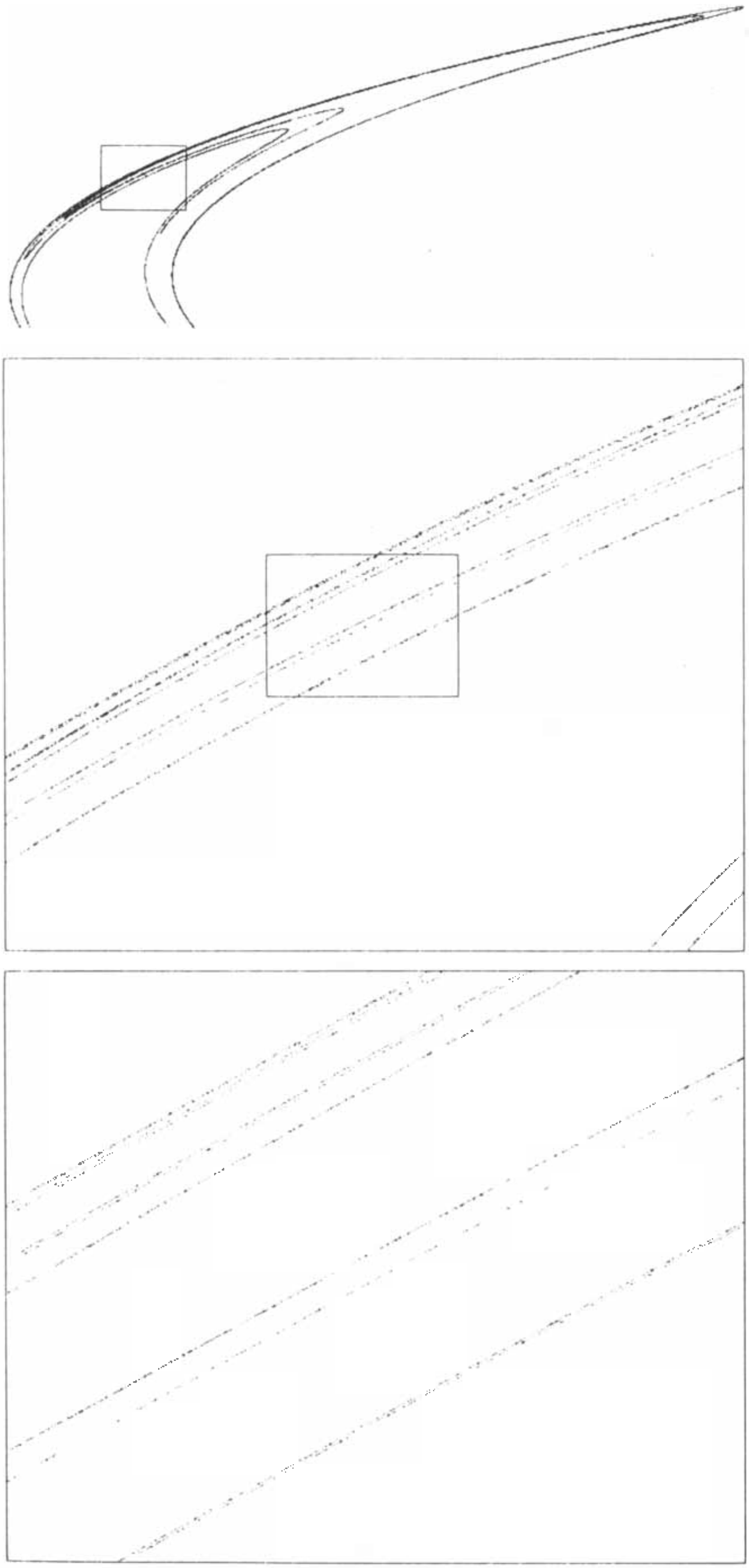
In such real-world problems the most successful equations yet found to model the phenomena are differential equations. A differential equation connects the continuous rate of variation in some quantity to that quantity's current size and the current sizes of other quantities. Moreover, the time variable is itself continuous, not jerking from one discrete instant to the next as some strange clocks and watches occasionally do, but indivisibly flowing like a liquid. One way to visualize the patterns defined by differential equations is to imagine a multidimensional space—it could have thousands of dimensions or merely a few—in which a point is continuously tracing out a curve. At any one moment the single point contains all the information about the state of the physical system. Its projections along the various axes give the values of all the relevant quantities that pin down a unique state. Clearly the space—called phase space—would need to have an enormous number of dimensions for a mere point to summarize the shape of a wave breaking on a beach. On the other hand, in a simple predator-prey relation only two dimensions suffice: one variable, say x, giving the predator population and the other, say y, giving the prey population.

As time progresses x and y determine each other in an intertwined manner. For example, a large population of predators will tend to reduce the population of prey, whereas a small population of prey will tend to reduce the population of predators. In such a system x and y constitute a single point (x,y) that swirls around smoothly in a continuous orbit on the plane. (Here the sense of "orbit" is different from the preceding one—that of the discrete, or jumping, orbits we saw when our parabolic function was iterated.) One such possible orbit appears in the top illustration on this page; it is generated by a differential equation called Duffing's equation. It looks like the path of a buzzing fly in your bedroom—or rather, it looks like the *shadow* of the fly's path on a wall. As a matter of fact this self-intersecting two-dimensional curve *is* the shadow of a non-self-intersecting three-dimensional curve. The motion of a point in phase space must *always* be non-self-intersecting. This arises from the fact that a point in phase space representing the state of a system encodes all the information about the system, including its future history, so that there cannot be two different pathways leading out of one and the same point.

In particular, in Duffing's equation there is a third variable, z, that I have not mentioned so far. If you think of x and y as representing predator and prey populations, then you can think of z as representing a periodically varying external influence, such as the sun's azimuth or the amount of snow on the ground. Now, if you will allow me to mix my buzzing-fly image into the predator-prey example, imagine a bedroom with a fly buzzing periodically back and forth between two walls. Let us say it takes the fly a year to cross the room and come back. (Perhaps it is a rather large bedroom, or maybe just a slow fly.) In any case, as the fly flies, its shadow on one of the two walls traces out the curve shown in the illustration. If the fly ever chances to come back to a point in the room it has passed through before, it is doomed to loop forever, following the path it took the preceding time over and over



Numbers on the fly's-path curve are the positions of the fly's shadow at regular intervals



The attractor of Hénon is blown up in two stages, revealing its infinite detail

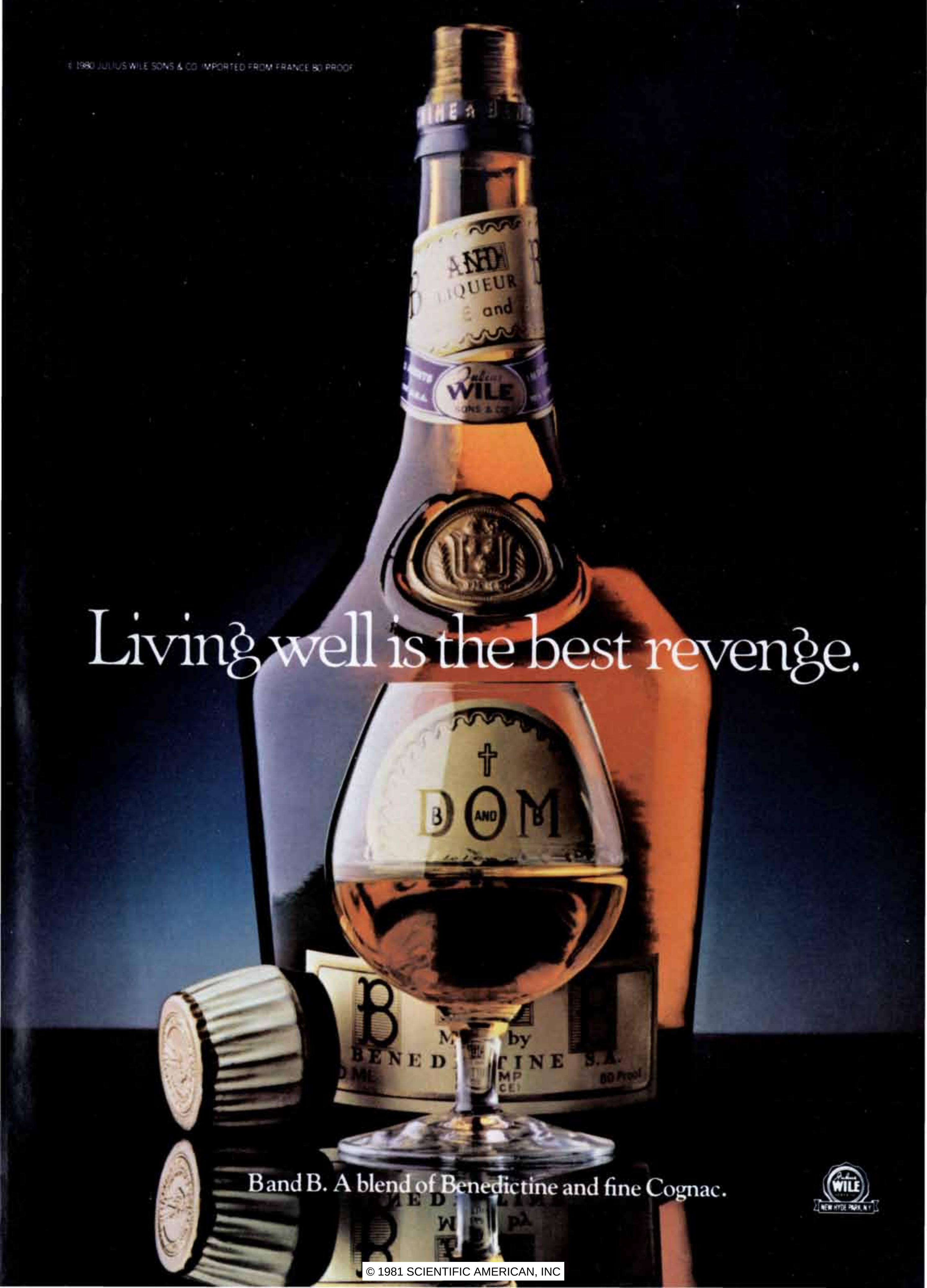
again. This gives you a picture of the continuous orbit of a point in phase space representing the state of a dynamic system controlled by differential equations.

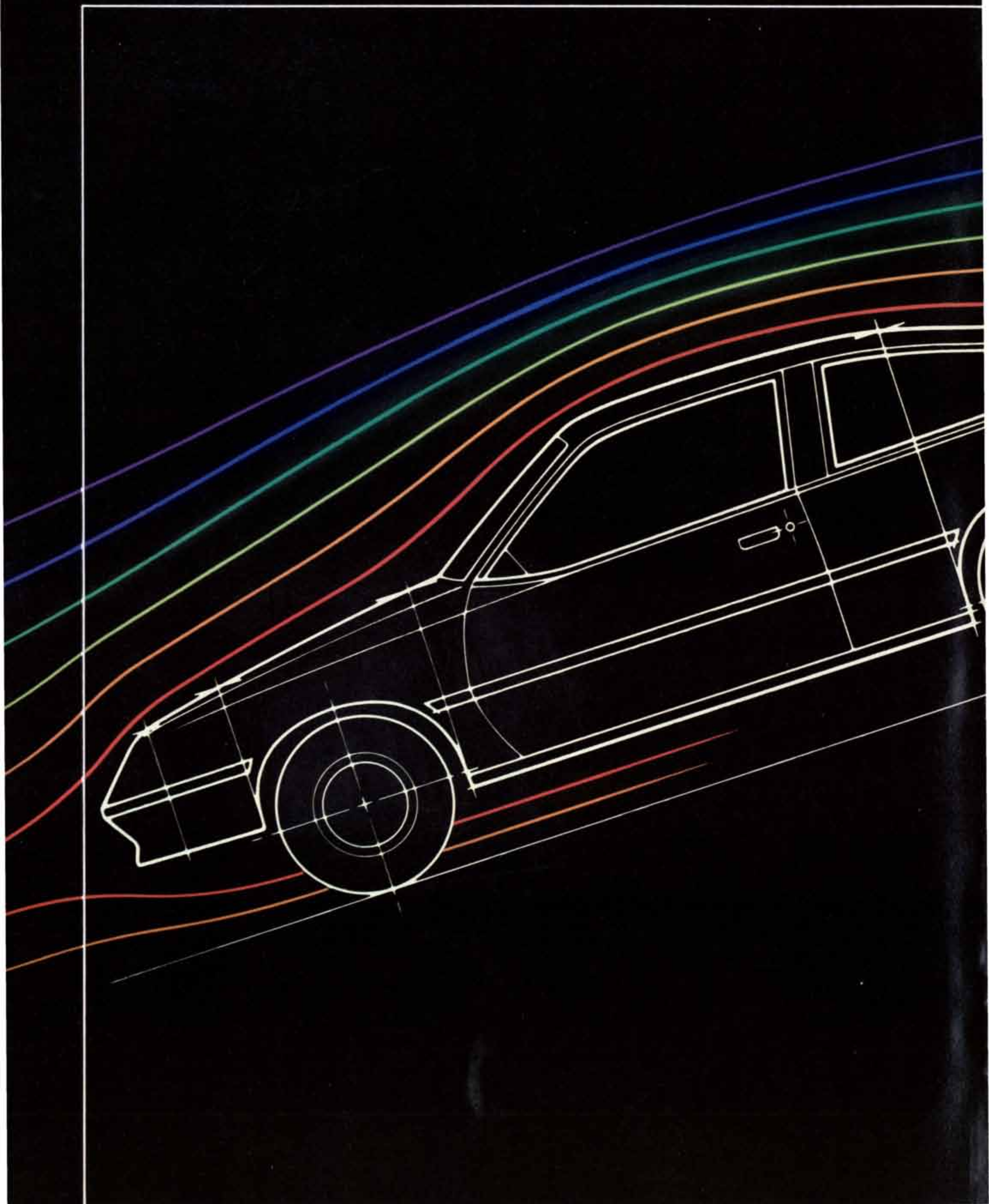
Now suppose we wanted to establish some connection of these systems to *discrete* orbits. How might we do so? Well, the values x, y and z need not be watched at all moments; they can be sampled periodically, at some natural frequency. In the case of animal populations a year is the obvious natural period. The sun's azimuth is exactly periodic, and the weather at least *tries* to repeat itself a year later. Thus a natural sequence of discrete points $(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots$ can be singled out—one per year. It is as if a strobe light blinked regularly and froze the fly on special annual occasions—perhaps at midnight every Halloween. At all other times its peregrinations around the room are unseen. The bottom illustration appearing on page 32 shows a sequence of discrete points along the fly-path's shadow, marked by numbers telling when they occurred. Gradually, as many "years" elapse, enough of these discrete points will accumulate that they will start to form a recognizable shape of their own. This pattern of points is a *discrete* "orbit," and so it is closely related to the discrete orbits defined by the iteration of our parabola $f(x)$. In that parabolic case we had a simple one-dimensional recurrence relation (or iteration): $x_{n+1} = f(x_n)$. Here we have a two-dimensional recurrence: $x_{n+1} = f_1(x_n, y_n)$ and $y_{n+1} = f_2(x_n, y_n)$.

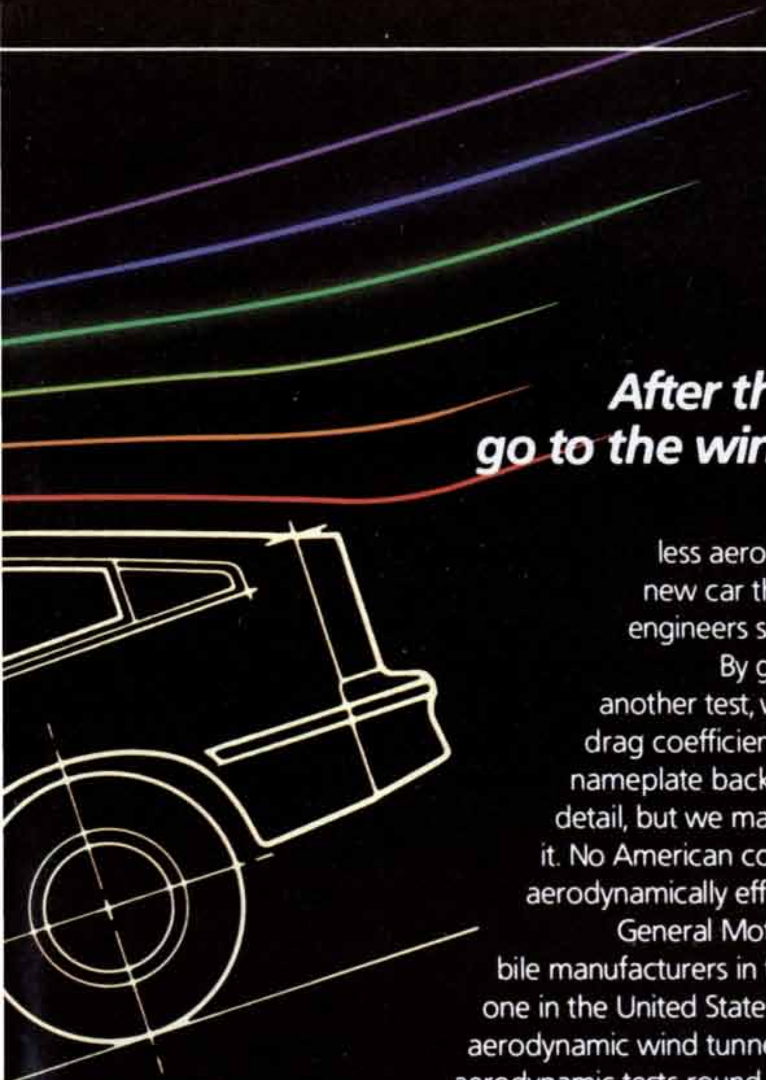
This is a system of *coupled* recurrence relations, in which output values of the nth generation (x_n, y_n) are fed right back into f_1 and f_2 as new inputs, to produce the $n + 1$st generation. On and on it goes, generation after generation. In higher-dimensional cases, of course, there are more such equations. Nevertheless, the skeleton of all these systems remains the same: a multidimensional point $(x_n, y_n, z_n, \ldots)$ jumps from one discrete location in phase space to another, as a discrete variable (n, representing time jumping ahead in discrete units) is incremented.

Notice that we have finessed our way around the continuous time variable that is involved in differential equations. We have done it by focusing on the way the point is connected to its predecessor one "year" earlier (or whatever natural period is involved). But is there always a "natural period" at which to look at a system of mutually intertwined differential equations? Not always. In some situations, however, there is, and this happens to be the case in all situations where turbulent behavior occurs.

Why is this so? All systems that exhibit turbulent behavior are "dissipative," which means that they dissipate, or degrade, energy from more usable forms such as electricity into the less usable


© 1980 JULIUS WILE SONS & CO. IMPORTED FROM FRANCE 80 PROOF
Living well is the best revenge.
B and B. A blend of Benedictine and fine Cognac.
Julius WILE
NEW HYDE PARK, N.Y.





After the aerodynamics are right, go to the wind tunnel again.

The original goal had been achieved. It took less aerodynamic energy per pound to move our new car than any other in its size category. But GM engineers sought even better results.

By going into the wind tunnel for yet another test, we found that we could reduce the drag coefficient even more by moving the car's nameplate back less than an inch. It was a small detail, but we made the change. And it was worth it. No American company had ever built a more aerodynamically efficient family car.

General Motors is one of the few automobile manufacturers in the world, and the only one in the United States, with its own full-scale aerodynamic wind tunnel. That means we can run aerodynamic tests round-the-clock, whenever necessary. After the test, we can make improvements right at the site, and test the cars again.

That's our idea of how to use technology to build cars. Attention to details where you don't see them, as well as where you do. Appearance and comfort may sometimes sell a car, but today's customers demand real value.

Our goal, as the company that sells more cars than any other manufacturer in the world, is to maintain our lead by using new technology to build cars that run better and last longer, with lower maintenance costs, than those built by any competitor–American or foreign.

General Motors

The future of transportation is here.

GM

30 years of computer experience: $1,565† and up.

Some big ideas start small.

Yet in the world of computers, it seems like just the opposite. They've become smaller. Faster. Less complicated and less expensive.

In the past 30 years, IBM has contributed quite a bit to that evolution. And today, one of our latest efforts is the IBM Personal Computer.

Our "big computer" experts got the chance to put their experience and knowledge into *your* IBM computer. Their enthusiasm was boundless. So were their demands.

They insisted that the system be expandable enough to add memory up to 256KB. Or to add disk drives. Or a printer for word processing.

They demanded that the system be flexible enough so the components could be moved around with ease.

Our IBM Personal Computer Software Publishing Department insisted that the programs be versatile enough for your professional, business, entertainment and personal needs. And the people who were writing our literature were adamant that an encyclopedia of "computerese" *not* be required reading.

We gave them all *they* wanted. So you could have the personal computer *you* want. For flexibility, performance and ease of use, no other personal computer offers as many advanced capabilities (see the box).

And there's more good news. The IBM Personal Computer starts at less than $1,600 for a system that, with the addition of one simple device, hooks up to your home TV and uses your audio cassette recorder.

You'll find the IBM Personal Computer at the nationwide chain of 150 ComputerLand® stores, and Sears Business Systems Centers. Of course, our own IBM Product Centers will sell and service the system. And the IBM Data Processing Division will serve those customers who want to purchase in quantity.

So come in and see the computer IBM made for *you*.

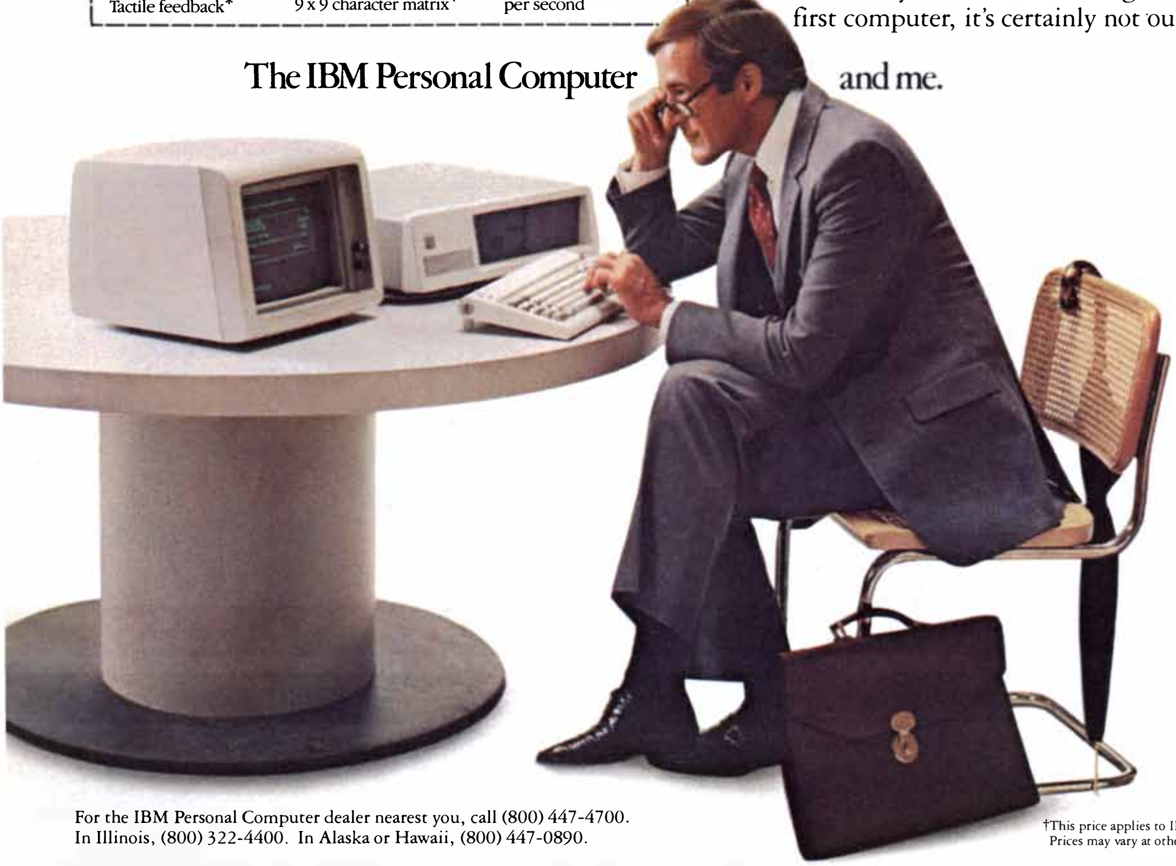
And rest easy. Because even though this may be your first computer, it's certainly not ours. IBM®

IBM PERSONAL COMPUTER SPECIFICATIONS
*ADVANCED FEATURES FOR PERSONAL COMPUTERS

User Memory
16K - 256K bytes*
Permanent Memory
(ROM) 40K bytes*
Microprocessor
High speed, 8088*
Auxiliary Memory
2 optional internal diskette drives, 5¼", 160K bytes per diskette
Keyboard
83 keys, 6 ft. cord attaches to system unit*
10 function keys*
10-key numeric pad
Tactile feedback*

Display Screen
High-resolution (720h x 350v)*
80 characters x 25 lines
Upper and lower case
Green phosphor screen*
Diagnostics
Power-on self testing*
Parity checking*
Languages
BASIC, Pascal
Printer
Bidirectional*
80 characters/second
12 character styles, up to 132 characters/line*
9 x 9 character matrix*

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



For the IBM Personal Computer dealer nearest you, call (800) 447-4700. In Illinois, (800) 322-4400. In Alaska or Hawaii, (800) 447-0890.

†This price applies to IBM Product Centers. Prices may vary at other stores.

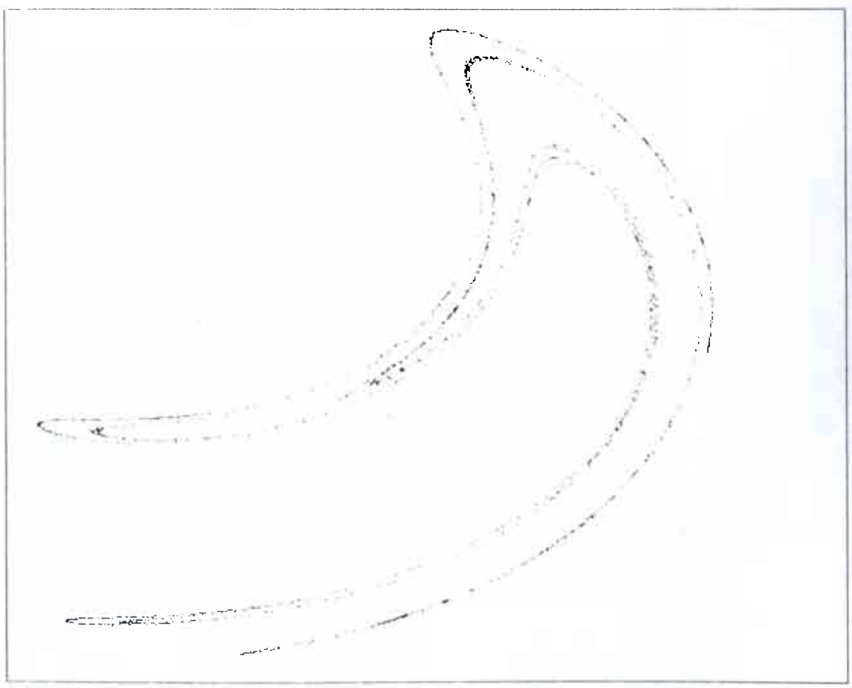
form of heat. This is caused by friction in the case of hydrodynamic flow, and by abstract analogues of friction in the other systems we have been considering. A familiar consequence of friction is that moving objects will grind to a halt unless energy is pumped in. Now, if we "drive" a dissipative system with a *periodic* driving force (you can imagine, for example, stirring a cup of coffee with a spoon in a periodic, circular way), then, of course, the system will not grind to a halt; it will head for some kind of steady state. Such a steady state is a stable orbit—or in our terms, an attractor in phase space. Since we have driven the system with a periodic spoon, however, we have defined a natural frequency at which to flash our strobe light and freeze the system's state, namely each time the spoon comes swinging around and passes some fixed mark on the cup, such as its handle. This will constitute our "year." In this way continuous time can be replaced by a series of discrete instants, as long as we are dealing with a dissipative system driven by a periodic force. And so continuous orbits can be replaced by discrete orbits, which brings iteration back into the picture.

If the driving force itself has no natural period (it may be simply a constant force), there is still a way to define a natural period, as long as some variable in the system swings back and forth between extremes. Just flash your strobe whenever that variable hits its extreme value, and the fly will still be caught at discrete instants. This type of discrete representation of the fly's motion in a multidimensional space is called a Poincaré map.

This stirring argument is only hand waving, of course, and needs much more rigor to be convincing to a mathematician. It nonetheless gives the flavor of how the study of a set of coupled differential equations can be replaced by the study of a set of coupled discrete recurrence relations. This is the vital step that brings us back to the recent discoveries about the parabola.

In 1975 Feigenbaum discovered that his numbers α and δ do not actually depend on the details of the shape of the curve defined by $f(x)$. Almost any smooth convex shape that peaks in the same spot will do as well. Inspired by the structural universality discovered by Metropolis, Stein and Stein, Feigenbaum tried working with a sine curve instead of a parabola. He was flabbergasted by the reappearance of the same numerical values, to many decimal places, of the numbers α and δ, which had characterized the period-doubling and the onset of chaos for the parabola. Just as for the parabola, there is for the sine curve a height-parameter λ and a set of special λ values that converge to a critical point λ_c. Moreover, the onset of chaos at λ_c is governed by the same numbers α and δ. Feigenbaum began to suspect that there was something universal going on here. In other words, he suspected that what is more important than f itself is the mere fact that f is being iterated over and over. In fact, he suspected that f itself might play no role in the onset of chaos.

It is not quite that simple, of course. Feigenbaum soon discovered that what matters about f is the nature of the peak at its very center. The long-term



A strange attractor is generated by Duffing's equation

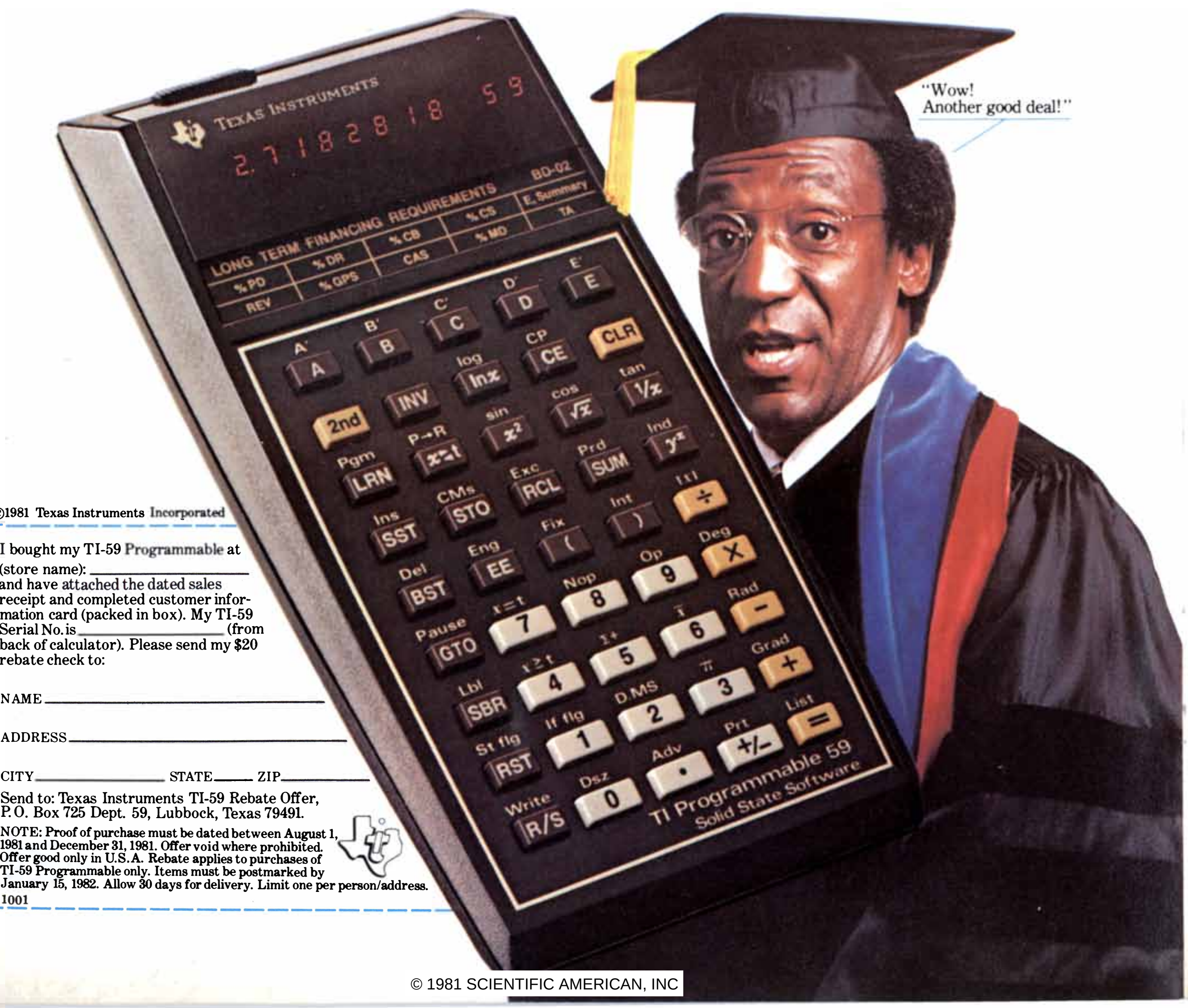
Get a $20 rebate on the TI-59 Programmable.

Even without the $20 rebate, the TI-59 is special—it's our most powerful programmable, and we've never offered it at a lower price.

The TI-59 gives you up to 960 program steps, or up to 100 memories, plus magnetic card read/write capability. You can also slip in one of TI's Solid State Software™ modules and successfully attack complex engineering, business, statistical and scientific problems. And by adding the optional PC-100C printer, you can record your calculations. But if that's not enough, any TI-59 owner can join our Professional Program Exchange for access to over 2500 additional programs.

So if you like the idea of having real programmable power, take us up on the rebate offer. Buy a TI-59 now, and fill out the coupon below. The offer ends December 31, 1981.

TEXAS INSTRUMENTS
INCORPORATED



I bought my TI-59 Programmable at (store name): ________ and have attached the dated sales receipt and completed customer information card (packed in box). My TI-59 Serial No. is ________ (from back of calculator). Please send my $20 rebate check to:

NAME ________

ADDRESS ________

CITY ________ STATE ____ ZIP ________

Send to: Texas Instruments TI-59 Rebate Offer, P.O. Box 725 Dept. 59, Lubbock, Texas 79491.

NOTE: Proof of purchase must be dated between August 1, 1981 and December 31, 1981. Offer void where prohibited. Offer good only in U.S.A. Rebate applies to purchases of TI-59 Programmable only. Items must be postmarked by January 15, 1982. Allow 30 days for delivery. Limit one per person/address.

1001

behavior of orbits depends only on an infinitesimal segment at the crest of the graph, and ultimately it depends only on the behavior at the very point where the maximum occurs. The rest of the shape, even the region close to the peak, is irrelevant. A parabola has what is called a quadratic maximum, as a sine wave, a circle and an ellipse do. In fact, the behavior of a smooth, randomly produced function at a typical maximum would be expected to be of the quadratic type in the absence of any special coincidences. Hence the parabolic case, rather than being a quirky exception, begins to seem like the rule. This empirical discovery by Feigenbaum, involving two fundamental scaling factors α and δ that characterize the onset of chaos through period-doubling attractors, represents a new kind of universality, known as *metrical* universality to distinguish it from the earlier-known structural universality. This metrical universality was proved to be correct (in the more traditional sense of proof) in the one-dimensional case by Oscar E. Lanford III.

A truly exciting development occurred when Feigenbaum's constants unexpectedly turned up in some messy models of actual physical systems that exhibit turbulence, not just in pretty and idealized mathematical systems. Valter Franceschini of the University of Modena in Italy adapted the Navier-Stokes equation, which governs all hydrodynamic flow, for computer simulation. To do so he turned it into a set of five coupled differential equations whose Poincaré maps he could then study numerically on his computer. He first found that the system exhibited attractors with repeated period doubling as its governing parameters approached the values where turbulence was expected to set in. Unaware of Feigenbaum's work, he showed his results to Jean-Pierre Eckmann of the University of Geneva, who immediately urged him to go back and determine the rate of convergence of the λ values at which period doubling occurred. To their amazement Feigenbaum's α and δ values—accurate to about four decimal places—appeared seemingly out of nowhere! For the first time an accurate mathematical model of true physical turbulence revealed that its structure was intimately related to the humble chaos lurking in the humble parabola $y = 4\lambda x(1 - x)$. Subsequently Eckmann, Pierre Collet and H. Koch showed that in the behavior of a multidimensional driven dissipative system all dimensions but one tend to drop out after a sufficiently long period of time, and so one should *expect* the characteristics of one-dimensional behavior—namely Feigenbaum's metrical universality—to reappear.

Since then experimentalists have been keeping their eyes peeled for period-doubling behavior in actual physical systems (not just computer models).

Such behavior has been observed in certain types of convective flow, but so far the measurements are too imprecise to lend very strong support to the idea that the parabola contains the clues revealing the nature of genuine physical turbulence. Still, it is tantalizing to think that somehow all that really matters is that a dissipative set of coupled recurrence relations is being iterated—but that the detailed properties of those recurrences can be entirely ignored if one is concentrating on understanding the route to turbulence.

Feigenbaum puts it this way. One often sees a pattern of clouds in the sky—a celestial trellis composed of a myriad of small white puffs stretching from horizon to horizon—that clearly did not happen "by accident." Some systematic hydrodynamic law has got to be operating. Yet, says Feigenbaum, it must be a law operating at a higher level, or a larger scale, than the Navier-Stokes equation, which is based on infinitesimal volumes of fluid and not on large "chunks." It seems that in order to understand such beautiful sky patterns one must somehow bypass the *details* of the Navier-Stokes equation and come up with some coarser-grained but more relevant way of analyzing hydrodynamic flow. The discovery that iteration gives rise to universality—that is, independence of the details of the function (or functions) being iterated—offers hope that such a view of hydrodynamics may be well on its way to emerging.

We have covered attractors and turbulence; what of strange attractors? We have now built up the necessary concepts to understand this idea. When a periodically driven two-dimensional (or higher-dimensional) dissipative system is modeled by a set of coupled iterations, the set of points caught by the successive flashes of the periodic strobe light traces out a shape that plays the role for this system that a simple orbit did for our parabola. Now, the possibilities are richer when one is operating in a space of more than one dimension. Certainly it is possible to have a stable fixed point, or an attractor of period one. This would just mean that at every flash of the strobe the point representing the system's state is exactly where it was last time. It is also possible to have a *periodic* attractor: one where after some finite number of flashes the point has returned to a preceding position. This would be analogous to the 2-cycles, 4-cycles and so on that we saw occurring for the parabola.

There is, however, another option. It is that the point never returns to its original position in phase space; successive flashes reveal it to be jumping around quite erratically inside a restricted region of phase space. Over a period of time this region may take shape before an observer's eyes as the strobe flashes periodically. In the majority of such cases so far studied a most unexpected phenomenon has been observed to take place: the erratically jumping point gradually creates a delicate filigree that recalls the "faint fantastic traceries made by frost on glass." (I owe this poetic image to the American critic James Huneker, who used it to describe the magical effect of one of Chopin's piano études, Op. 25, No. 2.) The delicacy is of a rather specific kind, closely related to the "fractal" curves described by Benoit Mandelbrot in his book *Fractals: Form, Chance, and Dimension.* In particular, any section of such an attractor, when blown up, reveals itself to be just as exquisitely detailed as was the larger picture from which it was taken. In other words, there is an infinite regress of detail, a never ending nesting of pattern within pattern. One of the earliest of such structures to be found, called the attractor of Hénon, is shown in the illustration on page 34. It is generated by the sequence of points (x_n, y_n) defined by the recurrence relations: $x_{n+1} = y_n - ax_n^2 - 1$ and $y_{n+1} = bx_n$. Here a is equal to 7/5 and b to 3/10; the seed values are $x_0 = 0$ and $y_0 = 0$. The small square in the top illustration is blown up in the middle illustration to reveal more detail, and then another square in the middle one is blown up in the bottom illustration to reveal yet finer detail. Note that what we appear to have is a sort of three-lane highway each of whose lanes breaks up, when magnified, into more parallel lanes, the outermost of which is a new three-lane highway—and on and on it goes. Any perpendicular cross section of this highway would be what is called a Cantor set, formed by a simple and famous recursive process.

Begin with a closed interval, say [0,1]. ("Closed" means that the interval includes its end points.) Now eliminate some open central subinterval. (Since an open subinterval does not include its end points, those two points will remain in the Cantor set being constructed before your eyes.) Usually the deleted subinterval is chosen to be the middle third (1/3, 2/3), but this is not necessary. Two closed subintervals remain. Subject them to the same kind of process, namely eliminate an open central subinterval inside each of them. Repeat the process ad infinitum. What you will be left with at the end of your infinite toil will be a delicate structure consisting of isolated points stretched out along the original segment [0,1] like beads of dew on a wire. Their number, however, will be uncountably infinite, and their density will depend on the details of your recursive elimination process. Such is the nature of a Cantor set, and if an attractor's cross sections have this weird kind of distribution, the attractor is said to be strange, and for good reason.

Another beautiful strange attractor is generated by the "stroboscopic" points 0, 1, 2,... in the bottom illustration on page 32. Since this pattern comes out of Duffing's equation, it is called Duffing's attractor, and it is shown in a slightly expanded scale in the illustration on page 39. Notice its remarkable similarity to the attractor of Hénon. Perhaps this is universality showing its face again.

It is interesting that for the parabola, at the critical value λ_c, f's attractor becomes aperiodic and consists of infinitely many points. (After all, it is the culmination of an infinite sequence of period doublings.) Moreover, the arrangement of those uncountably many points on the interval [0,1] has been constructed by Feigenbaum's recursive rule involving his constant α, as you saw in the illustration on page 20. It is not hard to believe that this implies that this particular attractor is itself a Cantor set. Hence the fertile parabola has provided us with an example of a *one-dimensional* strange attractor!

In the chaotic regime of the more general k-dimensional case, long-term prediction of the path that a point will take is quite impossible. Two nearly touching points on a strange attractor will, after a few blinks of the strobe light, have wound up at totally different places. This is called "sensitive dependence on initial conditions" and is another defining criterion of a strange attractor.

No one knows at present just why, how or when strange attractors crop up in the chaotic regimes of iterative schemes representing dissipative physical systems, but they do seem to play a central role in the mystery of turbulence. David Ruelle, one of the prime movers of this entire approach to turbulence, wrote: "These systems of curves, these clouds of points, sometimes evoke galaxies or fireworks, other times quite weird and disturbing blossomings. There is a whole world of forms still to be explored, and harmonies still to be discovered."

Robert M. May, a theoretical biologist, in a now famous review article covering the field in 1976, concluded with a plea that I find most apt and would like to repeat:

"I would... urge that people be introduced to [the equation $y = 4\lambda x(1 - x)$] early in their mathematical education. This equation can be studied phenomenologically by iterating it on a calculator, or even by hand. Its study does not involve as much conceptual sophistication as does elementary calculus. Such study would greatly enrich the student's intuition about nonlinear systems.

"Not only in research but also in the everyday world of politics and economics we would all be better off if more people realized that simple nonlinear systems do not necessarily possess simple dynamical properties."

BOOKS

The preservation of corpses, the misnamed jungle and the essence of modern astronomy

by Philip Morrison

MUMMIES, DISEASE, AND ANCIENT CULTURES, edited by Aidan and Eve Cockburn. Cambridge University Press ($49.95). Hamlet had it straight from the First Clown: "... your water is a sore decayer of your whoreson dead body." That craftsman of mattock and epigram had his own observations to support his view, but every commonplace salt, dried or frozen haddock presents the same clear evidence. Without liquid water the little carnivorous larvae of many species do not thrive on dead bodies. What is more, those enzymes that catalyze the degradation of tissue are inactivated, whether they are introduced by bacteria or fungi or are an autolytic constituent of the tissues themselves. The process is patently a balance of rates: dehydration v. decomposition. In warm climates speedy drying is needed to prevent decay; in icy climates the process is slowed and dehydration can be leisurely.

For bodies long buried in the wet bogs of Denmark most enzymatic attack is slowed by the absence of dissolved oxygen in the cold water of the peat layers, the activity of the anaerobic bacteria present in the body is hindered by the cold and the acid soils act as antibodies. Decomposition becomes a millennial reaction of the fixed tissues, as slow as medieval forecasts of resurrection. The left-handed amino acids do slowly degrade into their right-handed counterparts, with a rate quite sensitive to temperature. A few percent of some high-molecular-weight proteins have remained otherwise intact for as long as 4,000 years; the protein antigens of tissue cells, in particular the familiar ABO blood-group antigens, have been studied in mummies with success for decades.

At the level of the optical microscope the tissues often repay normal pathological study; under the scanning electron beam or the higher-power transmitted electron beams there is cellular detail in plenty. The radiographer can work more or less normally; skeletal structure is remarkably well preserved. The sex and age of a mummy at the time of death can best be told by reading bone: pelvis and skull details separate male and female fairly well, although with noticeable overlap. The age of the skeleton is so far best tallied from the progressive closing of the gaps in various bones (but beware of the skull sutures: they are too hard to see on the X-ray plate).

A compendium of paleopathology worldwide, this big volume presents some 20 well-organized papers that sum up the contemporary state of this intimate investigation of the human past. The evidence is powerful, if sometimes enigmatic and always macabre. It is the Egyptian mummies, long collected and examined, that hold first place, but the mummies of North and South America, Denmark, Australia, Japan, western Siberia and the Canary Islands are not overlooked. The final fourth of the book treats the rapid growth in modern methods of study, so far "music of the future."

Millions of mummies lie as they were interred in the dry tombs and sands of Egypt before the entry of Islam; they are found whenever a road or an airstrip is built or irrigation floods new lands. Some hundreds of thousands have certainly been lost, often consumed most prosaically, for example imported into Canada in the 19th century for the linen of their wrappings. (Rags good for papermaking were too scarce in that woolclad nation; what happened to the bodies is not known.) Herodotus himself is the authority for the techniques of Egyptian New Kingdom embalming, a trade that by his time had reached such maturity that he describes three recognized grades, like the recommendations of a mail-order catalogue.

Detailed chapters discuss recent examinations of Egyptian specimens. One, designated PUM II for Pennsylvania University Museum, had attracted all the antique skills: the organs were preserved, the wrappings were complete, the sarcophagus was elegantly decorated. Both the wrappings and the tissues were so hard that their dissection called for surgical saws. The embalmer's hot fluid resin had polymerized into a solid organic glass. That material was identified by X-ray diffraction; it is mainly an oil of the juniper tree, with minor additions of camphor oil and the gum myrrh. Tissue analyses support the inference (counter to the text of Herodotus) that the principal drying step was covering the body for more than a month with drying natron: an impure natural sodium carbonate familiar as a desert mineral deposit. Before that step the organs were removed, treated with molten resin and packed in special jars. It is inferred that before the slow chemical desiccation the body cavity was washed with a palm wine containing alcohol; then came the spreading of the hot liquid resin and the careful multiple wrappings, with more resin. The modern outline runs to 13 distinct steps for which there is either direct evidence or a logical inference.

In Toronto they examined ROM I. The body of a 12th-century-B.C. teenaged temple weaver, it had simply been wrapped and stored in a painted and inscribed coffin, to be desiccated naturally "in the dry, hot air of Thebes." The money saved on embalming was put into the coffin by a loving family of the "upper ranks of the working class." ROM I showed unmistakable infection with schistosomes, tapeworm and trichinae, but the cause of death is not known.

The account of Peruvian mummification centers on the fact that the arid land may hold a million bodies well preserved one way or another. The world's first museum of paleopathology has recently been founded in Lima. One of its early findings is that in Peru bone tuberculosis is pre-Columbian; it is also present in the Old World as far back as the Egyptian mummies go. The ethnographic material is even more remarkable. We see one pottery vessel that plainly shows a mummy being borne on a special litter by two attendants, background to the elaborate near-contemporary accounts of the funeral ceremonies that followed the death of the Inca, most puissant of sovereigns.

Dressed and wrapped in abundant fine cloth, specially woven for tribute, the body was stored at the family estate with many possessions, among sacrifices that may have included wives and retainers. Custodians cared for the royal mummies, gave them food and drink, changed and washed their clothing and even brought in visitors with whom the Inca wished to "speak." At coronations the royal remains were carried through the streets amid scenes of great reverence. There is no direct evidence for all of this; many details rest on somewhat later accounts, not quite independent of the contemporary chronicler better known for "impressionistic detail" than for veracity. In Australia and New Guinea we have the survival of the most archaic of treatments of the dead: in several distinct cases the entire dried body is smeared with red ocher. That blood-red pigment is found strewn on skeletal remains in the Lower Paleolithic of all Europe.

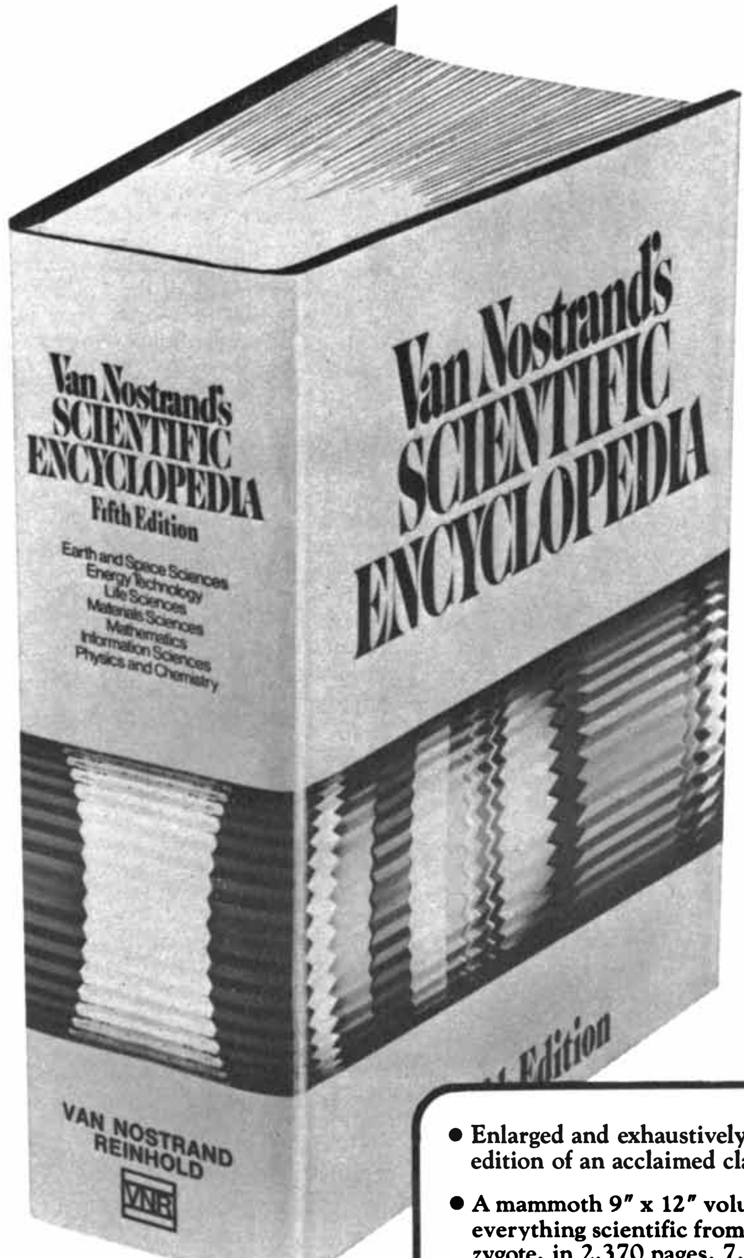
One true prize lies at the end of this somewhat fearful trail. It is a detailed

knowledge of the history of human mortality, its causes and changes. That goal is far off. Isolated results on the presence of this or that condition are not few, and they can be important. Vital statistics, however, call for large samples. The salvage archaeology of the first Aswan Dam before 1910 required that a heroic group of five specialists examine within a few months some 8,000 mummies. The autopsies were carried out with devotion and skill, although they were perforce "a very crude and rushed affair." They showed little malignant disease, no syphilis or rickets and only rare instances of inflammatory disease, suggesting either a "remarkable resistance to infection in ancient Nubia" or a low virulence of the pathogens. Most of the causes of death remain unknown, although many diseases of the present day are surely present in the Egyptian material overall.

Embalming is of course a very widespread ritual in our own time. The finest example is said to be that of the body of Eva Perón, "who appears to be merely asleep." The method was an Argentine innovation of this century: infiltration of the tissues with paraffin wax. Only a page or so is given to the remarkably preserved body of the wife of the Marquis of Tai, a high officer of the early Han of China. No other such mummy is reported anywhere; an hour-long film of the autopsy convinced the editors of this book of the unique elastic state of the skin, tissues and joints. Multiple hermetic sealing may have been the secret; the scientific publications show nothing of special note in the tissue. It seems that the body is now preserved in formaldehyde, perhaps an unfortunate decision of the moment. The anaerobic chemistry of the Tao worked a great deal better than that marvelous suit of joined tablets of magical jade, whose wearer, as museumgoers will recall, became but dust and bone. This volume is a rich, definitive treatment, although its index disappoints rather too many searches, and a few captions from the Russian are misleadingly translated.

JUNGLES, edited by Edward S. Ayensu. Crown Publishers, Inc. ($35). "Conceived, edited and designed" by a London company of book professionals, this attractive volume is a clear success within its genre. It is a big book but not a thick one; its large pages number about 200. The knowing text is brief, amounting to about a third of each page. Each spread is a combination of photographs, careful explanatory maps or diagrams and large, showy paintings of wildlife, many in color. Inset boxes here and there discourse on the history of the topic. The entire short work is an encyclopedic treatment without footnotes or other justification but prepared by experts. The publications engine is at full throttle: experienced artists, serious pic-

The most comprehensive and authoritative scientific reference in the world.



Van Nostrand's
Scientific Encyclopedia

a $69.50 value

Yours for only **$3.95**

when you join **The Library of Science.**
You simply agree to buy 3 more books—at handsome discounts—within the next 12 months.

Find the answers to virtually *all* your scientific questions, in the thoroughly updated and expanded fifth edition of Van Nostrand's SCIENTIFIC ENCYCLOPEDIA. It has been hailed by *The New York Times* as "an amazing book for both the general and scientific reader."

Thousands of photos, diagrams, graphs and charts completely illuminate the ENCYCLOPEDIA's clear and thorough coverage of every scientific field—including archaeology, astronomy, general science, biology, medicine, physics, ethology and meteorology. Whether you're looking for a basic explanation of earthquakes or a comprehensive analysis of nuclear reactors—you'll find it all in Van Nostrand's SCIENTIFIC ENCYCLOPEDIA.

What is **The Library of Science?** The Library of Science is a very special book club. A club dedicated to keeping members of the scientific community (and knowledgeable amateurs) aware of the important advances and the speculative thinking in their own and related fields.

Begin enjoying the Club's benefits by accepting Van Nostrand's SCIENTIFIC ENCYCLOPEDIA. It's the perfect guide to the marvels of science ... and it's a great bargain, too!

- **Enlarged and exhaustively updated edition of an acclaimed classic.**
- **A mammoth 9" x 12" volume covering everything scientific from abacus to zygote, in 2,370 pages, 7,200 editorial entries, 8,000 cross-references and 2.2 million words ... plus 2,450 photos, diagrams, and charts as well as 550 tables.**
- **Authoritatively compiled by over 180 internationally respected scientific experts.**

If reply card has been removed, please write:
The Library of Science—Dept. 2-AH7, Riverside, N.J. 08075
to obtain membership information and application.

ture research, a well-known editor. The text is mainly the product of three English academic research people, one an ecologist, one a botanist and one a zoologist who doubles in ethnography, with a good deal of help from their friends. It worked.

For starters the popular title is disclaimed. The term jungle suggests an impassable green thicket; this volume is about tropical rain forests, now a fifth of the world's forest area, where the dimly lit ground is open, predators are few and the specialized fauna are full of color. The jungle picture is correct only at the edge of a clearing or along a riverbank, where the explorers once paddled day after day past the wall of green that grew where sunlight came unfiltered to ground level. In the forest proper a multilayered canopy of leaves controls growth in stratified layers. Indeed, this environment, the home of the most diverse swarm of species, about half of the total number extant, is currently studied from walkways hung high among the treetops, steel towers rising into the leafy canopy, even cages hung up there to protect the night observer. Sun and rain nourish the tropical rain forest; it takes steady rainfall without seasonal drought at 80 inches per year or more to support it.

Open to a few of the spreads throughout the volume. One spread shows a two-mile line of waterfalls, cascading in two high steps a couple of hundred feet down into the Rio Paraná. It is Iguazú Falls, near the three-way junction of Brazil, Paraguay and Argentina. Iguazú is never dry. A column recites the names of the Amazon's explorers, from Vicente Pinzón in 1500 to Teddy Roosevelt and on to Colonel Percy Fawcett, who was lost in the Mato Grosso in 1925. Here is a wonderful shot of an emergent tree in purple blossom, flowering high above the green canopy. The spread offers good thumbnail color drawings to scale of five high trees: the world's tallest (a eucalyptus), a coastal redwood and three high rain-forest species, all shorter.

Here is a backwater in Brazil crowded with the giant lily pads of wild *Victoria regia,* at least 50 two-meter boatlike leaves afloat, with some flowers in blossom. Seven toucan heads blaze from a full-page painting with the great colorful bills of the various species; overleaf we see a color photograph of the toco, a species of toucan that is gifted with not only a signpost bill but also a pair of feet glistening in electric blue. There is a good spread on fruit bats, and one on lianas, so familiar and yet with their complex stem section and its physiology so little described. ("If a small cut is made in a liana stem, it is possible to hear a hissing sound as air is sucked into the severed vessels.") A pair of pages treat tersely the history, husbandry and marketing of "the food of the gods," a rain-forest understory tree: the cocoa.

The chapters cover the geography, botany, mammals, reptiles, birds, insects, human inhabitants and economic uses of the rain forest. The last pages are somber. Since the end of World War II the rain forests have been under serious threat. New technologies, from the chain saw to heavy logging transport vehicles, have come on the scene. The market has grown for the lighter forest woods, not merely the elegant teak, ebony and mahogany of the past. Where once hardworking men dragged scattered selected trees away with buffalo or elephant at high cost, entire stands are now "quarried": nothing is left.

The West African forest is seriously depleted now; Nigeria imports timber. The lowland forests of the Philippines and Malaya will be mostly gone in a couple of decades. South America is luckier so far; there the groups of species that in Southeast Asia can be milled into a single grade of lumber are not found. Sapele, abura and gaboon are some of the plywood species Europe has taken from the West African forest; white seraya, keruing and meranti flow even more copiously now from Southeast Asian rain forests to Japan, again chiefly as the material of plywood and veneers. Most of us will not learn to recognize these species before they are rarities, even though they cover much of our doors and walls, our furniture, even our packing crates (less in the U.S. than in Europe and Japan).

More and larger inviolate tropical-forest reserves are indispensable and belated. The tropical countries that require cropland and crave export income need help if the world is to retain any sensible fraction of these splendid forests. It might even be that the global environment will suffer from the onslaught on the rain forests; although that cannot now be predicted, "it would be very unwise to ignore." There is a place too for plantations and for large seminatural plots managed for economic use. A guide to tropical rain-forest products and a good index complete this book—handsome, useful and timely if not very deep.

COSMIC DISCOVERY: THE SEARCH, SCOPE, AND HERITAGE OF ASTRONOMY, by Martin Harwit. Basic Books, Inc. ($25). "Turning Points" is not perhaps the aptest term for the five events invoked in the opening pages of this highly original book. The events are Galileo's first use of the telescope and four of the 20th century: the balloon flight of 1912 on which Viktor Hess showed that cosmic rays reach the earth from the sky and the first successful observations in radio (1931), X-ray (1948) and gamma-ray astronomy (1973). They share one quality: each opens a new window onto the cosmos. Professor Harwit, himself a thoughtful Cornell astronomer dealing in unconventional wave bands, offers a provocative account of how new windows are opened, by whom and what they can show. He projects a plausible estimate of how many cosmic windows we shall open eventually, and even something of how many fine scenes we can hope to see through them.

The foundation of Professor Harwit's narrative is a list of 43 (at a later count 44) cosmic phenomena, from the stars as a whole (known, of course, since antiquity, although our sun gets short shrift here) through multiple stars and X-ray galaxies to the mysteries of the extraordinary double star SS 433, recognized in 1979. A third of the book discusses these discoveries; a couple of pages are spent on each, giving a compact history of the discovery and its significance at the level of the general scientific reader. On only a few could any reader demur at Professor Harwit's terse, direct account. From quite another line of argument he proceeds to define a phenomenon—the kind we name in texts, hold meetings about and become experts on—by the token that it is distinguished from all other cosmic observations by a substantial factor in at least one of the properties of the signals by which we detect it.

Those properties cannot be many: in the electromagnetic spectrum we exploit only a handful of channels, and we expect only a couple of other new modes of remote viewing to materialize, say through neutrinos and gravitational waves. Moreover, we cannot study an infinite range of wavelengths in any modality. Astronomy is by definition a science of severe limitations. In any laboratory and in any field science we do our best to turn over the crystal specimen, walk around the tree or map the entire mountain range. Only in deep-space astronomy are we constrained to stand still and watch. For cosmic study we are too small, our life is too short. We must study the thickness of the great spiral galaxy in Andromeda largely by looking for another galaxy we can see more nearly edge on. It turns out that with a sensible acceptance of approximation the number of really distinct observational channels we have exploited up to now is just about the same as the number of the phenomena Professor Harwit names; the argument closes on itself well. The two catalogues (they are very good reading even though they include tight lists of objects and observations) do not match one for one, but it seems clear that they are about the same length and that they speak to the same point.

In such a framework the argument can go much further. Once we recognize some phenomenon in two distinct channels we can argue that the number of phenomena overall is roughly fixed, including those we do not yet know. The argument is the same coincidence argument that lay behind Ernest Rutherford's 1910 studies of alpha particles by the flashes they gave in a phosphorescent compound viewed under the microscope. Innocent of electronics and photocells, he and his colleagues had to count the scintillations one by one in a darkened room. Could that count be relied on? Even the patient Hans Geiger must have blinked once in a while. (We can understand why he invented an early electronic counter!) No trouble: Rutherford set two people to counting the same sample, each to record his own flashes by pressing a key. The moving chart recorded the counts made by each observer and displayed which of the flashes both observers saw. It is not hard to see how the efficiency of each experimenter could be estimated; once it is known what the chance of missing a flash is the true number can be calculated, allowing statistically for unseen flashes. By analogy the same trick succeeds in astronomy, although the doubtful randomness of the searches, the smallness of the samples and the effects of various changes over the years, among other things, do not favor a simple analysis.

It seems likely that we could have uncovered half a dozen cosmic phenomena quite independently in two ways, say the planets or the spiral galaxies by their emission of both light and radio waves. We can therefore expect only 100 to 200 phenomena overall as distinct as the ones we now list. Even if there are those that show up intrinsically on only one channel, for example the enigmatic gamma-ray bursts, they too, judged by our present sample, cannot be very many. The discussion is full of interest as Harwit seeks to justify the choices he has made to define his channels and the like. The result is far from ironclad, but it does not seem to be off by an order of magnitude.

Professor Harwit, like many an operations analyst before him, thinks he may have a grip on some sturdy pillar of the cosmos. A reader can remain somewhat skeptical; what we see here looks a good deal like our own footprint. Planetary astronomy is almost outside the argument, as solar physics seems explicitly to be. There we know too much, we see and feel too much, and we do not call what we find in such a domain cosmic phenomena in quite the same way. Stars appear on Harwit's master list some 20 times, counting everything we know and call stars, from neutron stars and SS 433 to the stable sun. Since the study of stars is relatively mature, we do not now think of all the rich details of atmosphere and interior, say, as a set of cosmic phenomena, although by using "filters" designed for the study of particular spectral lines, sizes and so on, one could most likely satisfy Harwit's definitions. Form is strongly subordinate to wave band today, but will it always be so? In these results there is not only the structure of the uncaring cosmos but also the human structure of science and its changes fast and slow. The results are in some sense unaffected by that distinction; the continual branching that divides sciences as they grow, which for example largely removed fluid mechanics from physics to a distinct specialty, is the kind of phenomenon being examined here.

The end of this study is a scheme for the design of the science of astronomy itself. Most of the conclusions, however, do not depend much on the statistical analysis. One cheers the historical evidence that new channels mean new discoveries, often quickly, often under the control of the very investigator who first exploited the new sensor. It is clear that external technology (in our talkative, wary and warlike times often communications technology and military search technology) aids new astronomical discovery, and that nonastronomers therefore contribute a great deal. All these points are well made here.

One interesting argument suggests that any plan for astronomy ought to allow an increment of about 5 percent of the effort per year for new directions. This follows from the fact that 20 or 25 years is about all a big instrument at a new observatory has before obsolescence. The "planning wedge" of future funds that are not tightly committed should thus be on the average about that fraction, provided it is possible to end the life of an institution as its unique functions dwindle. The Max Planck institutes in Germany have reached some solution to the problem of program termination without real damage to the personal research careers involved. There a section or an institute is terminated or its leadership is transferred once a year; there are about 50 institutes in the system.

Again like other operations analysts Professor Harwit looks at our institutions with a somewhat ingenuous gaze. He does not much notice the many functions of public planning documents. Their role as ratification for an expected organizational direction is at least as important as the new schemes they present. Advisers are numerous, and decision makers well understand how both to seek out and to limit the advice they find themselves externally constrained to follow—or to avoid.

The apparatus and graphs of Harwit's provocative study are fresh and admirable: the unusual joint index and glossary, the tables of comparative sizes, energies, masses and the like, the graphical studies of bandwidth and angular resolution and more are valuable for the general reader and the specialist alike. Students of astronomy and physics particularly should seek out this lively and varied work.

NEW YORK IN AERIAL VIEWS: 68 PHOTOGRAPHS BY WILLIAM FRIED, with identifications by Edward B. Watson. Dover Publications, Inc. ($6.50). The Piper PA22 with its extra camera hatches flies low and slow, its transponder, like its air crew, alert to the perils of three-dimensional traffic in New York City airspace. William Fried and his company Skyviews have been shooting the Big Apple in black-and-white by day ever since 1946. "The greatest number of flying hours [are] spent photographing market areas before site-location decisions are made" for enterprises all the way from banks to restaurants.

Here is a fine set of the take from those decades. The boroughs are all here, with of course an emphasis on the tall towers of Manhattan and the long spans. Each photograph, nearly a spread, is flanked by a reduced copy on which a dozen features are marked by number. (An overall index would have been a rather modest luxury, well worthwhile.) A few views are from high aloft; you can make out misty Montauk Point or chemical New Jersey. Most of them, however, are city scale at its best, a few square kilometers of urban architecture, with the sense that the lucky airplane passenger occasionally shares of the elegant model being turned in the hand. The skies are clear; the city is recognizably intricate, beautiful and changing. The oldest photographs date from the first postwar years; up-to-date ones go to 1978.

Some changes shout from the images. Whereas in 1946 two dozen ships lay at the piers around the tip of Manhattan Island from South Street up the Hudson, there were three or four in 1966 and none in 1976. The Lower East Side was a sea of roofs in 1946; by 1978 the area is occupied by multiple crystalline blocks of public housing with ringing names such as Lillian Wald, LaGuardia and Baruch. The high twin Trade Towers now star, supported by that Citicorp camera top, along with the older Wall Street canyons, Hunts Point tanks, Brooklyn Heights, the big Ferris wheel at Coney Island and the East Side's P.S. 142, quite circular too. Even Ebbets Field is recorded; someone is on third! The airports and the parks are shown clearly. Ground truth is of course not attended to; that nifty tilted solar intake on the Citicorp tower never went to work, the streets are choked with problems, the South Bronx is full of blinded houses; none of this appears to the lens aloft. Given a little reflection, these views are at once an intimate and a detached look at a supremely complex artifact of our era, at the right scale. The bridges, ornamented by the flourishes of their approach roads, remain crisp pleasures; so does Mlle Liberté.


86 PROOF BLENDED SCOTCH WHISKY DISTILLED AND BOTTLED IN SCOTLAND. IMPORTED BY THE BUCKINGHAM CORPORATION, NEW YORK, N.Y.
"IT'S NEVER BEEN DONE BEFORE." A TERRIFYING THOUGHT TO SOME. HERE'S TO THOSE IT INSPIRES.
Kitty O'Neil is 5 ft. 3 in. tall, weighs 98 pounds and has been deaf since childhood. But that hasn't slowed her down.
She's set 26 world speed records on land, 2 on water and 1 on waterskis. (Including the Women's World Land Speed Record—with a top speed of 618 mph.) And she was the first stuntwoman to perform a 180 ft. high fall and a 90 ft. high fall while on fire (another world record).
Kitty O'Neil has been through a lot. And after a day spent doing things that haven't been done before, Kitty O'Neil does something she has done before: she enjoys a Cutty Sark.
Kitty O'Neil
The Scotch with a following of leaders.
CUTTY SARK®

Established 1845 SCIENTIFIC AMERICAN November 1981 Volume 245 Number 5

The Census of 1980

For the first time the population of rural areas and small towns grew faster than that of the metropolitan areas. The migration to the West and the South, begun during the 1960's, continued

by Philip M. Hauser

The returns of a U.S. census seldom reveal unforeseen developments in the nature of the American population. The first clues to the emergence of new trends in fertility, mortality and migration usually come from surveys of representative samples of the nation's population carried out between censuses. The surveys are conducted by demographers in the Bureau of the Census, in other Government agencies and in universities and corporations. What the census provides is information that is more complete geographically and more detailed than information obtained from smaller-scale surveys.

Analysis of the results now available from the 1980 census has begun to confirm a significant change in patterns of population movement and growth in the U.S. For the first time since the census was begun in 1790 rural areas and small towns, where about a fourth of the nation's people live, had greater rates of growth than the metropolitan centers. This trend was observed in all regions of the nation. At the same time there was an accelerated movement of people out of the older industrial regions of the North and into the South and West. The differences in regional growth rates, in large part the product of migration, may signal the beginning of a new pattern of population distribution, one marked by dispersal of the population and by increased employment in industries outside the older metropolitan centers. Preliminary census data also confirm some other findings of earlier surveys, including the decentralization of population within metropolitan areas, the aging of the U.S. population and differences in fertility and household size among racial and ethnic groups.

Much of the attention that has been given to the 1980 census has focused on questions that are essentially procedural. The census count exceeded the predictions of demographers by almost five million people. This was the largest discrepancy in the history of the census. Although improved methods of enumeration may account for some of the error in the estimates, much of it can probably be attributed to the influx of illegal aliens. A related question is that of how many people were not counted in the census tabulations. In the past the undercount has been disproportionately large for members of racial and ethnic minority groups. The size of the illegal-alien population and the possibility of selective undercounting have become political issues of extreme sensitivity: at stake is the apportionment of representatives in Congress and the distribution of billions of dollars in Federal revenues to be shared with the states.

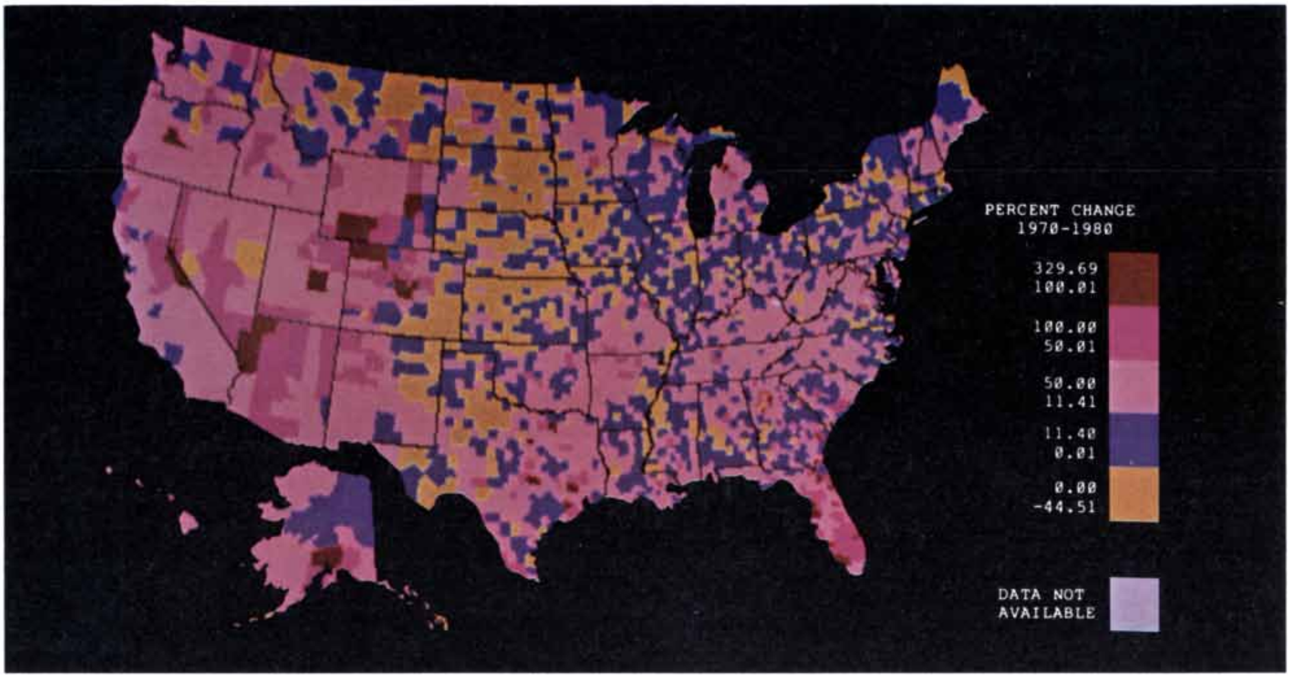
Perhaps the most unexpected result of the 1980 census was the size of the total count of the population. According to the census returns the resident population of the U.S. on April 1, 1980, was 226,504,825. This number represents an increase of 11.4 percent over the 1970 population, which was about 203 million. (Both numbers exclude citizens living abroad; their inclusion would raise the 1980 total to about 227 million.)

The increase in the population between the censuses was the result of births and immigration on the one hand and of deaths and emigration on the other. Most of the increase was brought about by the excess of births over deaths for U.S. citizens and residents; the contribution of immigration and emigration is not known precisely because of the entrance of illegal aliens and the poor quality of the data available on emigration.

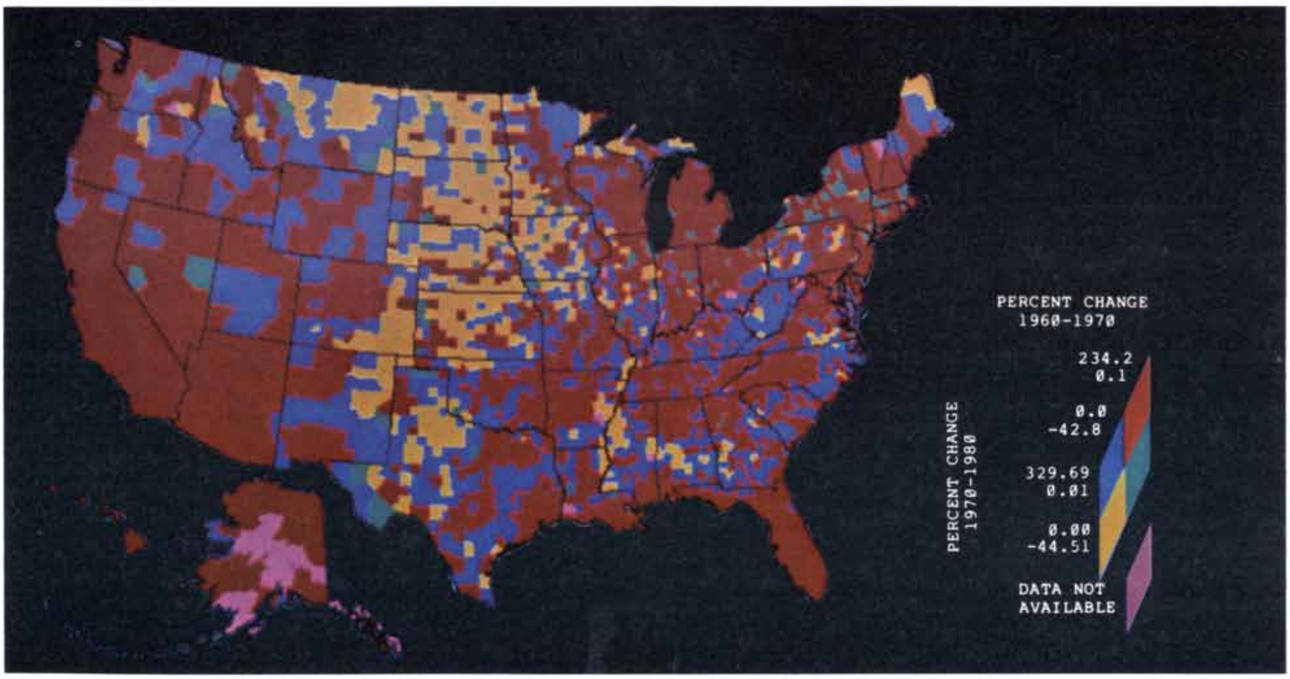
Before each decennial census the Bureau of the Census prepares an estimate of the expected total count. The estimate is derived from previous censuses and from information on births, deaths, immigration and emigration. The difference between the precensus estimate and the actual count is known as the error of closure. It is the magnitude of the error of closure in the 1980 census that has most surprised demographers: the actual count exceeded the estimate by some 4.8 million.

Part of the reason for the unexpectedly large error is that coverage of the population may have been more complete in this census than in any previous one. Census officials had a strong incentive to make the census exhaustive: anticipated population losses in the Northeast and North Central regions could be expected to give rise to political controversy, which would be aggravated by any undercount in those areas.

Several tactics were adopted to improve the coverage. It had been discovered in an analysis of the returns of the 1970 census that about 10 percent of the buildings then classified as vacant were actually occupied. In 1980 all housing units reported to be vacant were checked by a census canvasser. The verification of the correctness of mailing addresses, which was partial in 1970, was extended to all households in 1980. In addition there was more extensive cross-checking of the census rolls against other records, such as those that list holders of a driver's license. More intensive canvassing of places such as pool halls and transient hotels was done in an attempt to include a greater proportion of people who have no permanent address.



CHANGES IN THE POPULATIONS of U.S. counties between 1970 and 1980 reflect a sharp reversal in the patterns of demographic growth and movement that prevailed between 1900 and 1970. Until 1970 there was a consistent movement of population into the metropolitan centers of the Northeast and North Central regions. Returns of the 1980 census, however, show that rural areas and small towns gained proportionally more residents than urban areas for the first time since census taking was begun in 1790. The counties that lost residents were concentrated in urban areas of the North and in the Great Plains and the Middle Western corn belt, where farm employment declined. The counties where population increased fastest during the decade are in the West, Southwest and South, regions of rapid economic growth. The census returns suggest that a more dispersed pattern of population distribution is beginning to take hold in the U.S. This map and the other maps that accompany this article show data for individual counties; the maps were prepared with a computer at the University of South Carolina. The computer is part of the Decision Information Display System, which was originally developed by the National Aeronautics and Space Administration to aid in the analysis of satellite data on weather patterns.



SIMULTANEOUS DISPLAY of changes in the population of U.S. counties in 1960–70 and in 1970–80 shows several distinct patterns. Some rural areas lost population in both decades (*yellow*) as the trend toward reduced farm employment persisted over both decades. Other rural counties lost residents in the 1960's but gained population in the next decade (*blue*) as part of the growth of nonmetropolitan areas. The population of many suburban counties increased in both decades (*red*) in a continuation of the migration to commuter communities that began after World War II. Large areas of the West and Southwest gained residents in both decades, with the rate of increase accelerating in the later period. Many cities of the Northeast and Middle West gained residents before 1970, then lost population (*green*).

There was also a more aggressive advertising campaign than had been mounted in previous decades, much of it aimed at minority groups.

On the other hand, the increased alienation of minority racial and ethnic groups, whose hopes for a higher standard of living were not being fulfilled as rapidly as they wished, may have led to their failure to cooperate in the census. This factor may have partially offset the efforts of the Bureau of the Census to improve the completeness of the census count.

The effort to improve census coverage may have helped to reduce the number of U.S. citizens and other residents who went uncounted and contributed to the high final count. In addition the number of illegal aliens who were counted was probably substantial. Although it seems improbable, given the tendency of illegal residents to avoid making contact with Federal authorities, the pattern of low precensus estimates of the populations of individual states strongly suggests that many illegal aliens returned a census form. More than 2.3 million of the 4.8 million people unexpectedly counted were residents of either California or Texas, both of which are in the direct path of entry from Mexico. These two states and Arizona, Colorado, New Mexico, Utah and Washington accounted for two-thirds of the overall error of closure.

In metropolitan areas such as New York and Chicago large numbers of Hispanics not included in precensus estimates were counted. Many of them may be illegal aliens. In explaining the unexpectedly high count of people of Hispanic origin, it should be noted that much of the advertising that preceded the census was in Spanish. Census officials have suggested that, given the benevolent tone of the advertising, it is possible many illegal aliens were persuaded they had nothing to fear from filling in a census form.

Notwithstanding the fact that almost five million more people were counted than had been predicted, the aspect of the census that has gained the most political attention is the possibility that some people were not counted, particularly members of minority groups and residents of the older industrial areas of the Northeast and North Central regions. The controversy over the undercount reflects the political origin of the census and its political uses. The decennial census was mandated in Article I of the Constitution to provide a basis for apportioning representation in the House of Representatives and for apportioning taxation among the states. Census data have more recently been utilized for the purpose of allocating those Federal revenues that are shared with state or with local governments.

Concern that older cities and those with large minority populations would suffer disproportionately from a failure to count their entire population led officials of Detroit and New York to challenge the accuracy of the census findings in Federal district courts. The officials asked to have the census totals adjusted to compensate for the portion of those cities' minority populations that they maintained had been missed in the census. The district courts upheld this contention and ordered an adjustment of the census totals. The U.S. Circuit Courts of Appeals, however, reversed the decision on procedural grounds, leaving undetermined the issue of the size of the undercount. The census data have therefore been published without adjustment. Both cities, however, have announced their intention to appeal the appellate-court decision.

There probably has never been a national census in a large country that achieved a complete count of the population. More complete registers of population, better census-taking techniques, more sophisticated procedures for sampling and new methods of information processing have reduced the discrepancy between a nation's population and the census count; some small proportion, however, usually goes uncounted. Analysis of censuses taken recently in industrialized nations suggests that typically a few percent of a population is missed. The undercount in the 1976 census of Australia is estimated to have been 2.7 percent; that in the census of Canada taken the same year is estimated to have been 2.0 percent; that in the 1962 census of France is estimated to have been 1.7 percent.

The proportion of the population that remains uncounted in U.S. censuses has generally been in the same range as that of other industrialized countries. The census of 1950 was the first in which the Bureau of the Census made an attempt to estimate the undercount; demographers in the bureau calculated then that 3.3 percent of the population had been missed. The undercount was estimated to have been reduced to 2.7 percent in 1960 and to 2.5 percent in 1970. Because the size of the population increased over the 20-year period, the absolute number of people who were not counted remained roughly constant at about five million.

Demographers in the bureau have said it is not yet possible to make a precise estimate of the 1980 undercount, partly because of the number of illegal aliens who were included in the census and the difficulty of determining the total illegal-alien population. Nevertheless, they have released a preliminary estimate of the 1980 undercount: about 2 percent of the total population. If the estimate is correct, the absolute number missed has decreased slightly to about 4.5 million.

The measurement of a census undercount is not exact, since it is based on estimates of fertility, mortality and immigration and emigration that are themselves subject to error. Thus data gathered in the 1970's enabled demographers at the bureau to conclude that the 1970 undercount had been 2.3 percent rather than 2.5 percent. In absolute numbers the 1970 undercount was reduced from 5.3 million to 4.7 million.

The undercount is politically troublesome not because it represents a large proportion of the population but because some groups are always counted less completely than others. The likelihood that a given individual will be counted varies with geographic region, sex, race and age. The undercount in the 1970 census is estimated to have been greatest in the South because of the substantial black population there and because much of the region is rural. (Although it is possible to obtain accurate lists of housing units in urban areas to provide the basis for a census, such lists are far less complete for the scattered population of rural districts.)

The census undercount for men is proportionally greater than that for women, and more blacks remain uncounted than whites. In 1970 the undercount of black males was 10.1 percent and that of white males was 2.1 percent, according to the revised estimates. The undercount of females was 5.3 percent among blacks and .9 percent among whites. Measurement of the undercount in 1980 is still subject to revision. It is currently estimated, however, that 7.5 percent of black males went uncounted, compared with .5 percent of white males and males of other races. Among women the undercount of blacks was 2.1 percent and that of white women and women of other races was 1.7 percent. Therefore although census coverage of both whites and blacks has improved, the proportion of blacks who are not counted remains substantially higher than that of whites.

Even within the black population the undercount is by no means uniform. The young, who make up the most mobile segment of the population, are the most likely to be missed. As an extreme example, it is estimated by the bureau that 22 percent of black males 28 years old were not counted in 1970. Hence for the purpose of apportionment of representation a black man of 28 counted as only 78 percent of a person between 1970 and 1980.

The data on a possible undercount of people of Spanish origin are not as complete as those for blacks. What evidence does exist suggests that Hispanics are

	1980 CENSUS DATA (THOUSANDS)	CHANGE 1970–80 (PERCENT)	CENSUS DATA (THOUSANDS) 1970	1960	CHANGE 1960–70 (PERCENT)	REPRESENTATIVES IN CONGRESS BASED ON 1980 CENSUS	CHANGE 1970–80
U.S. TOTAL	**226,505**	**11.4**	**203,302**	**179,323**	**13.4**	**435**	**—**
NORTHEASTERN STATES	**49,137**	**0.2**	**49,061**	**44,678**	**9.8**	**95**	**−9**
NORTH CENTRAL STATES	**58,854**	**4.0**	**56,590**	**51,619**	**9.6**	**113**	**−8**
SOUTHERN STATES	**75,349**	**20.0**	**62,813**	**54,973**	**14.3**	**142**	**+8**
WESTERN STATES	**43,165**	**23.9**	**34,838**	**28,053**	**24.2**	**85**	**+9**
NORTHEASTERN STATES							
NEW ENGLAND	12,349	4.2	11,847	10,509	12.7	24	−1
Maine	1,125	13.2	994	969	2.6	2	—
New Hampshire	921	24.8	738	607	21.6	2	—
Vermont	511	15.0	445	390	14.1	1	—
Massachusetts	5,737	0.8	5,689	5,149	10.5	11	−1
Rhode Island	947	−0.3	950	859	10.6	2	—
Connecticut	3,108	2.5	3,032	2,535	19.6	6	—
MIDDLE ATLANTIC	36,788	−1.1	37,213	34,169	8.9	71	−8
New York	17,557	−3.8	18,241	16,782	8.7	34	−5
New Jersey	7,364	2.7	7,171	6,067	18.2	14	−1
Pennsylvania	11,867	0.6	11,801	11,320	4.2	23	−2
NORTH CENTRAL STATES							
EAST NORTH CENTRAL	41,669	3.5	40,263	36,225	11.1	80	−6
Ohio	10,797	1.3	10,657	9,706	9.8	21	−2
Indiana	5,490	5.7	5,195	4,663	11.4	10	−1
Illinois	11,418	2.8	11,110	10,081	10.2	22	−2
Michigan	9,258	4.2	8,882	7,823	13.5	18	−1
Wisconsin	4,706	6.5	4,418	3,952	11.8	9	
WEST NORTH CENTRAL	17,184	5.2	16,328	15,394	6.0	33	−2
Minnesota	4,077	7.1	3,806	3,414	11.5	8	—
Iowa	2,913	3.1	2,825	2,758	2.4	6	—
Missouri	4,917	5.1	4,678	4,320	8.3	9	−1
North Dakota	653	5.6	618	632	−2.2	1	−1
South Dakota	690	3.6	666	680	−2.0	1	—
Nebraska	1,570	5.7	1,485	1,411	5.2	3	—
Kansas	2,363	5.1	2,249	2,179	3.2	5	—
SOUTHERN STATES							
SOUTH ATLANTIC	36,943	20.4	30,679	25,972	18.1	69	+4
Delaware	595	8.6	548	446	22.9	1	—
Maryland	4,217	7.5	3,924	3,101	26.5	8	—
District of Columbia	638	−15.7	757	764	−0.9	—	—
Virginia	5,346	14.9	4,651	3,967	17.2	10	—
West Virginia	1,950	11.8	1,744	1,860	−6.2	4	—
North Carolina	5,874	15.5	5,084	4,556	11.6	11	—
South Carolina	3,119	20.4	2,591	2,383	8.8	6	—
Georgia	5,464	19.1	4,588	3,943	16.4	10	—
Florida	9,740	43.4	6,791	4,952	37.1	19	+4
EAST SOUTH CENTRAL	14,663	14.5	12,808	12,050	6.3	28	+1
Kentucky	3,661	13.7	3,221	3,038	6.0	7	+1
Tennessee	4,591	16.9	3,926	3,567	10.1	9	—
Alabama	3,890	12.9	3,444	3,267	5.4	7	—
Mississippi	2,521	13.7	2,217	2,178	1.8	5	—
WEST SOUTH CENTRAL	23,743	22.8	19,327	16,951	14.0	45	+3
Arkansas	2,286	18.8	1,923	1,786	7.7	4	—
Louisiana	4,204	15.3	3,645	3,257	11.9	8	—
Oklahoma	3,025	18.2	2,559	2,328	10.0	6	—
Texas	14,228	27.1	11,199	9,580	16.9	27	+3
WESTERN STATES							
MOUNTAIN	11,368	37.1	8,290	6,855	20.9	24	+5
Montana	787	13.3	694	675	2.8	2	—
Idaho	944	32.4	713	667	6.9	2	—
Wyoming	471	41.6	332	330	0.6	1	—
Colorado	2,889	30.7	2,210	1,754	26.0	6	+1
New Mexico	1,300	27.8	1,017	951	6.9	3	+1
Arizona	2,718	53.1	1,775	1,302	36.3	5	+1
Utah	1,461	37.9	1,059	891	18.8	3	+1
Nevada	799	63.5	489	285	71.6	2	+1
PACIFIC	31,797	19.8	26,548	21,198	25.2	61	+4
Washington	4,130	21.0	3,413	2,853	19.6	8	+1
Oregon	2,633	25.9	2,092	1,769	18.2	5	+1
California	23,669	18.5	19,971	15,717	27.1	45	+2
Alaska	400	32.4	303	226	34.1	1	—
Hawaii	965	25.3	770	633	21.6	2	—

CHANGES IN THE POPULATION OF THE STATES between censuses may have important political consequences. The population of the U.S. grew to about 226,500,000 between 1970 and 1980, an increase of 11.4 percent. The growth was not evenly distributed, however. The population in the Northeast and North Central census regions grew at a rate lower than that of the nation as a whole, whereas the population in the Southern and Western states grew faster than the national average. The census provides the basis for apportioning seats in the House of Representatives. If the census returns are not further adjusted, the Northeast will lose nine House seats and the North Central region eight; the South will gain eight seats and the West nine. In addition the distribution of billions of dollars in Federal revenue shared with the states, which is also allocated on the basis of data from the census, may shift toward the South and the West.

missed in a proportion between that of whites and that of blacks but closer to the level among the black population. A substantial proportion of both groups live in the inner zones of central cities. In these areas the difficulties of enumeration are compounded by the large number of people without permanent residence and by attitudes of distrust or hostility toward governmental authority.

In the controversy issuing from the census undercount an important technical point has sometimes been overlooked. Since the census is the most complete enumeration of the residents of a country, how is it possible to determine that some part of the population has not been counted? Beyond that, how is it possible to estimate the size of the undercount?

Demographers rely mainly on two methods to make such determinations. In the demographic method records of births, deaths, immigration and emigration, among other records, are analyzed to estimate the increase in the population since the previous census. The additional population is combined with the count of the population from the preceding census to yield some idea of how large the population is at the time of the current count. The demographic method is the one that has most often been employed in estimates of the undercount made by demographers within the Bureau of the Census.

In the matching method census returns are compared with records of a small sample of the population to determine the proportion of that sample missed in the census. One sample that has often been selected for matching studies is the one that provides the basis for the Surveys of Current Population conducted by the bureau. The sample is made up of about 65,000 households; the number of people in each household is known with greater precision in the sample than in the census. By matching names on the census list with names from the Surveys of Current Population it is possible to identify people in the sample who were missed in the census. It is also possible to match census returns with Social Security or Internal Revenue Service records. From the proportion missed one can then estimate the size of the undercount in the national population.

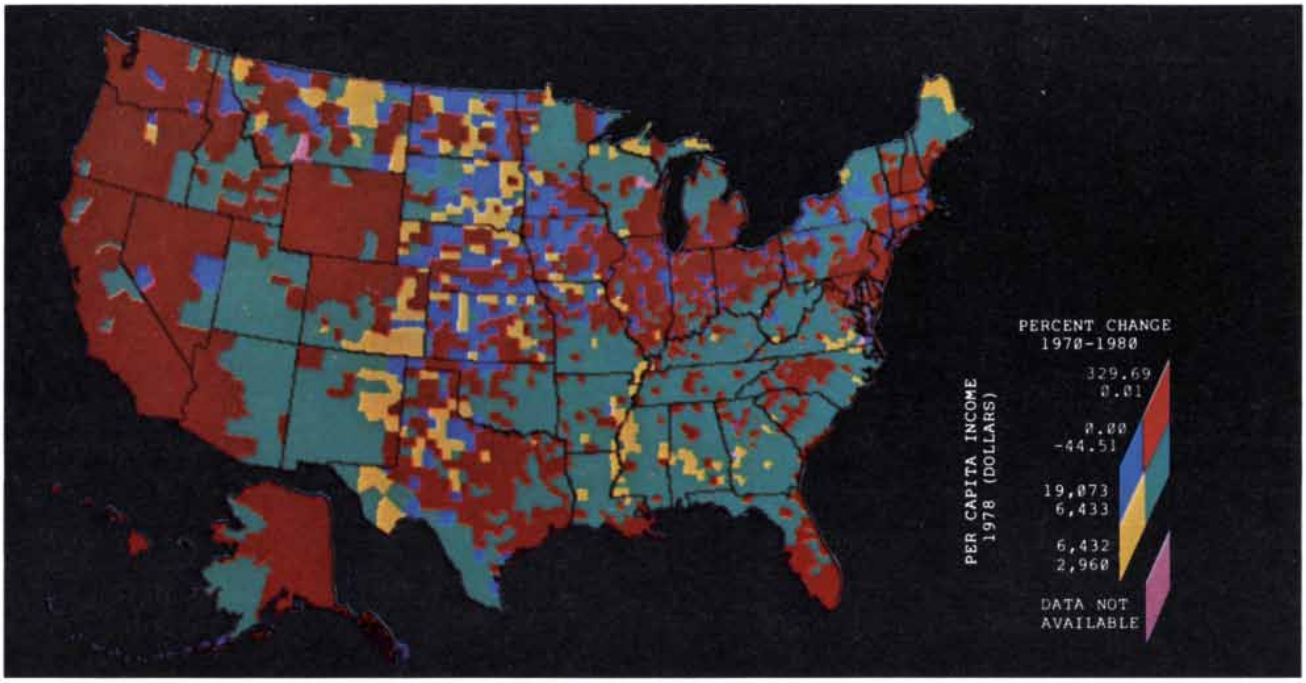
In previous censuses the demographic and the matching methods together provided a reasonably accurate estimate of the undercount; in those censuses, however, there were far fewer illegal aliens in the country, and the issue of illegal aliens had attracted far less political attention than it has now. In order to determine the undercount of the 1980 census with a degree of precision that is acceptable more intensive studies and

	1980 POPULATION (THOUSANDS)			CHANGE 1970–80 (PERCENT)		
	TOTAL	CENTRAL CITY	SUBURBS	TOTAL	CENTRAL CITY	SUBURBS
U.S. TOTAL	**102,816**	**39,639**	**63,177**	**+6.9**	**−4.9**	**+16.0**
NORTHEAST	**27,294**	**11,173**	**16,120**	**−4.9**	**−12.3**	**+1.0**
NORTH CENTRAL	**24,937**	**8,764**	**16,173**	**+0.3**	**−14.1**	**+10.5**
SOUTH	**25,388**	**10,200**	**15,189**	**+19.8**	**+2.4**	**+35.3**
WEST	**25,197**	**9,502**	**15,695**	**+17.7**	**+8.1**	**+24.3**
NORTHEAST						
New York	9,081	7,035	2,045	−9.0	−10.9	−1.6
Philadelphia	4,701	1,681	3,020	−2.6	−13.8	+5.1
Boston	2,760	563	2,197	−4.8	−12.2	−2.7
Nassau–Suffolk	2,604		2,604	+1.9		+1.9
Pittsburgh	2,261	424	1,837	−5.8	−18.4	−2.4
NORTH CENTRAL						
Chicago	7,058	2,986	4,071	+1.2	−11.4	+12.9
Detroit	4,344	1,197	3,147	−2.0	−20.0	+7.7
St. Louis	2,345	451	1,894	−2.7	−27.6	+5.9
Minneapolis–St. Paul	2,109	639	1,471	+7.3	−14.2	+20.4
Cleveland	1,896	573	1,323	−8.1	−23.7	+0.8
SOUTH						
Washington, D.C.	3,045	635	2,410	+4.6	−16.0	+11.9
Dallas–Fort Worth	2,964	1,284	1,680	+24.7	+3.7	+47.4
Houston	2,891	1,574	1,317	+44.6	+27.6	+72.0
Baltimore	2,166	785	1,382	+4.6	−13.4	+18.6
Atlanta	2,010	422	1,588	+26.0	−14.7	+44.3
WEST						
Los Angeles–Long Beach	7,446	3,309	4,136	+5.7	+4.4	+6.8
San Francisco–Oakland	3,227	1,013	2,214	+3.8	−6.0	+9.0
Anaheim–Santa Ana–Garden Grove	1,926	550	1,376	+35.5	+24.0	+40.7
San Diego	1,860	875	985	+37.0	+25.4	+49.1
Denver–Boulder	1,615	566	1,050	+30.3	−2.8	+59.6

LARGE CITIES of the South and the West census regions grew rapidly during the decade. The populations of SMSA's in the North Central region did not change appreciably. Populous SMSA's in the Northeast region lost residents. Data in the upper panel of this table are for the 50 largest SMSA's; in the lower panel information is given for the five largest SMSA's in each region. Although suburbs in the Northeast and North Central regions gained residents, the central cities in those regions lost substantial portions of their populations. Central cities of the South and the West grew somewhat, but not nearly as fast as suburban areas in those regions.

AREA (AS OF JUNE 19, 1981)	1980 POPULATION (MILLIONS)	1970 POPULATION (MILLIONS)	CHANGE 1970–80 (PERCENT)	PERCENT OF U.S. TOTAL
U.S.	226.5	203.3	11.4	100.0
SMSAS (318)	169.4	153.7	10.2	74.8
CENTRAL CITIES (429)	67.9	67.8	0.1	30.0
SUBURBS	101.5	85.6	18.2	44.8
NONMETROPOLITAN AREAS	57.1	49.6	15.1	25.2
SCSA'S (13)	67.8	65.0	4.3	30.0
SMSA'S AND SCSA'S OF MORE THAN 1 MILLION (32)	100.2	93.8	6.7	44.2
SMSA'S AND SCSA'S OF MORE THAN 2.5 MILLION (12)	68.1	65.2	4.5	30.0

TRENDS IN METROPOLITAN POPULATION during the 1970's varied greatly according to the size of the city and whether an area was part of the central city or of the suburbs. The largest cities and the central cities were the most likely to grow slowly or to lose population during the decade. Overall the population of the 318 Standard Metropolitan Statistical Areas (SMSA's) grew by 10 percent, a rate lower than that of the nation or of the nonmetropolitan areas. The largest metropolitan census units are the Standard Consolidated Statistical Areas (SCSA's), which are made up of several SMSA's; the largest SCSA is composed of New York, Newark and Jersey City. Although the population of the suburban areas of metropolitan centers increased by 18 percent, the population of the central cities remained roughly constant.



AVERAGE PER CAPITA INCOME in each county of the U.S., displayed simultaneously with the change in population, indicates that some areas of the West and the Southwest that gained residents were also characterized by high income. Much of the population growth was the result of migration from the older industrial regions of the North; the availability of jobs was a powerful attraction to those who were moving. New technology applied to industrial processes and improved communications and transportation have made it much less disadvantageous for factories to be outside metropolitan industrial centers. The result has been the rapid development of rural and suburban industrial complexes, such as "Silicon Valley," near Palo Alto, Calif., where many electronics and information-processing companies have plants. Businesses that supply corporate or personal services have followed industry. There was no single relation, however, between income and population growth that prevailed in all parts of the country. Some areas of the Northeast where per capita income is high lost population; the South, which is the region of lowest per capita income, recorded some sharp population gains in some areas.

additional methods will be required.

Before the census began workers in the bureau utilized the demographic method to estimate that the actual population (as distinct from the expected census count) was about 226 million. This number is quite close to the actual 1980 census count, with no allowance made for the undercount. The coincidence is probably the result of the large count of illegal aliens whose presence was not anticipated by the bureau.

Even though only limited census results are now available, the returns have confirmed some broad trends suggested by smaller demographic studies. Perhaps the most significant trend is a change—the magnitude of which is yet to be precisely measured—in longstanding patterns of migration. Between 1900 and 1970 there were several consistent migratory flows in the U.S. There was a movement from the heavily settled East to the more sparsely settled West. There were substantial migrations of blacks from the South to Northern and Western cities, and there were large movements from rural areas to urban centers as farm employment declined and that in manufacturing increased.

The results of these simultaneous population movements were quite different at national and local levels. In the nation as a whole there was a tendency toward a more dispersed distribution of the population. Within smaller geographic units such as states, however, there was a tendency toward a higher degree of concentration as the number of metropolitan areas increased and the population of existing metropolitan areas grew. The movement of residents of central cities to the suburbs after World War II did not reverse the growth of the metropolitan areas; it merely redistributed the population within them.

At least one aspect of this well-established pattern was fundamentally altered in the 1970's. For the first time since the taking of the census began, the population of nonmetropolitan counties, made up of rural areas and small towns, grew faster than the population of the metropolitan areas. At the same time the growth rates of the Northeast and North Central regions declined steeply; growth rates there during the 1970's were much lower than the national average.

These changes in growth rates may represent the beginning of a new phase of demographic organization in the U.S. The economies of scale and of proximity that favored the concentration of business and industry in the metropolitan centers may be undergoing modification by new demographic and economic forces. Japan and the industrial countries of Europe are showing similar demographic trends as their metropolitan centers, like those in the U.S., reach what may be the limits of urban growth.

The populations of the census regions of the West and South grew by 24 percent and 20 percent during the decade, about twice the rate of growth of the U.S. population as a whole. The population of the Northeast census region increased by less than 1 percent and that of the North Central region by 4 percent. The shift of population to the West and South can be explained to some extent by flows of illegal immigrants into those areas. There can be little doubt, however, that the older industrial regions have become areas of net outward migration.

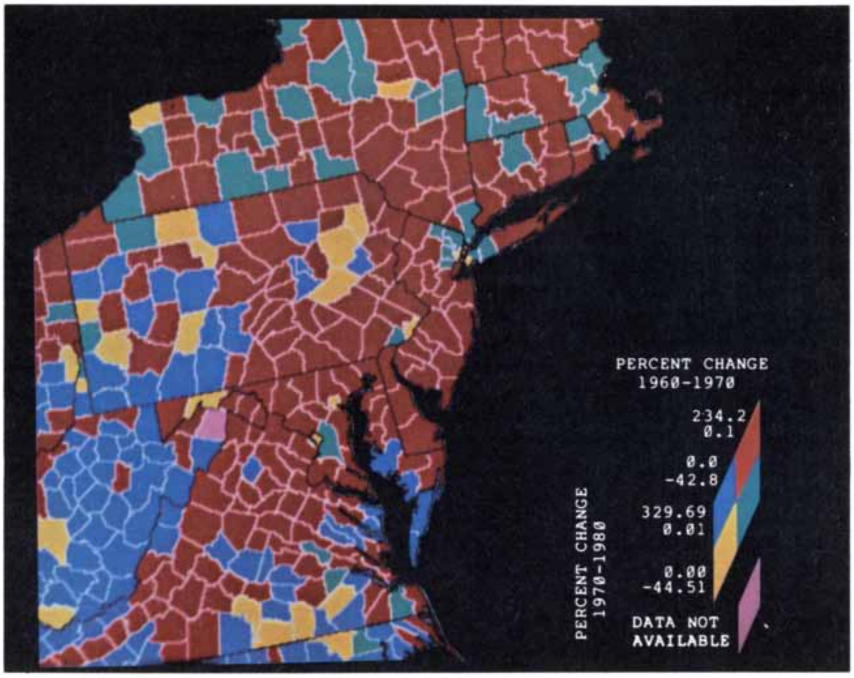
The political consequences of this change may be considerable. If the published census enumeration is not revised, 17 seats in the House of Representatives will shift. The Northeast will lose nine seats and the North Central region eight; the South will gain eight seats and the West nine. Such a redistribution of representation may lead to changes in the political composition of the Congress as the elected representatives of regions with younger populations and more vital economies come to outnumber the representatives from the older sections of the country.

The population of rural areas and small towns increased by 15 percent between the 1970 and the 1980 censuses. The metropolitan centers grew by 10 percent. The distinction between population centers and more sparsely settled areas is formalized in the census by the designation of Standard Metropolitan Statistical Areas, or SMSA's. An SMSA is a county or a group of counties encompassing at least one city with a population of 50,000. There were 284 SMSA's prior to the 1980 census. On the basis of the 1980 census returns the Office of Management and Budget (OMB), which designates SMSA's, added 35 new SMSA's and dropped one, for a total of 318. The OMB also added four Standard Consolidated Statistical Areas (SCSA's) for a total of 13. An SCSA is composed of two or more SMSA's; the largest is made up of the New York, Newark and Jersey City SMSA's. Counties not included in an SMSA are considered nonmetropolitan.

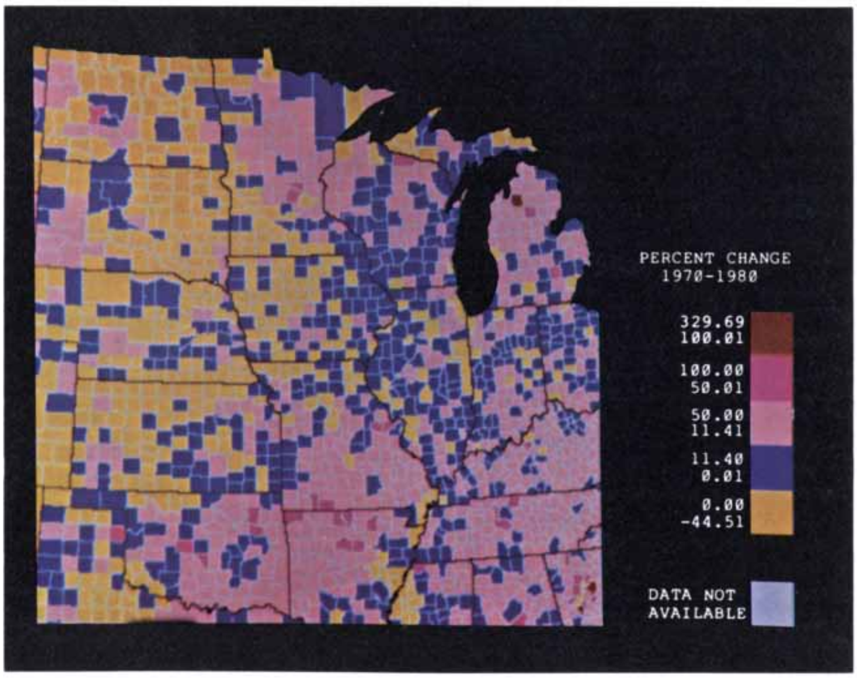
The faster growth of population in rural areas and small towns was all but universal: in 49 states the nonmetropolitan counties had the higher rate of population increase. Although nonmetropolitan counties increased more rapidly than SMSA's in the 1970's, the proportion of the total population living in SMSA's actually increased because of the designation of new SMSA's and the continued growth of the other SMSA's. Thus in 1980 the 318 SMSA's had 75 percent of the population whereas a decade earlier 243 SMSA's had had 69 percent. Between 1970 and 1980 the population of the 13 SCSA's increased by only a little more than 4 percent; nevertheless, by 1980 they included 30 percent of the total population.

The deceleration of the growth of the metropolitan areas varied according to the region and the size of the city; within metropolitan areas there were also variations. The reduction in growth (and in some cases the loss of population) was most acute in the central cities, the largest cities and those of the Northeast and North Central regions. Whereas the overall increase in the population of the SMSA's was 10 percent, the 32 largest metropolitan areas, which have more than a million residents each, grew by only 7 percent; of these the 12 largest areas, each of which has more than 2.5 million residents, grew by less than 5 percent.

Of the 50 largest metropolitan areas 10 are in the Northeast census region; six of these lost residents. The largest loss, both as a percentage of population and in absolute numbers, was that of the New York SMSA: 9 percent. The



COUNTIES OF THE NORTHEASTERN U.S. reflected several of the demographic trends that prevailed in the nation during the 1970's. This map is a detail of the bottom map on page 54; it shows population growth in 1960–70 and 1970–80. The New York and Boston metropolitan areas, which gained population during the 1960's, lost residents during the next decade. (The New York SMSA lost 9 percent of its population in the course of the 10-year period.) Nassau and Westchester counties, suburban areas adjoining New York whose populations include large numbers of people who work in the city, showed the same pattern. Some rural counties lost population in both decades. Most counties that are made up of rural areas and small towns, however, lost population in the 1960's and then gained residents in the next decade.



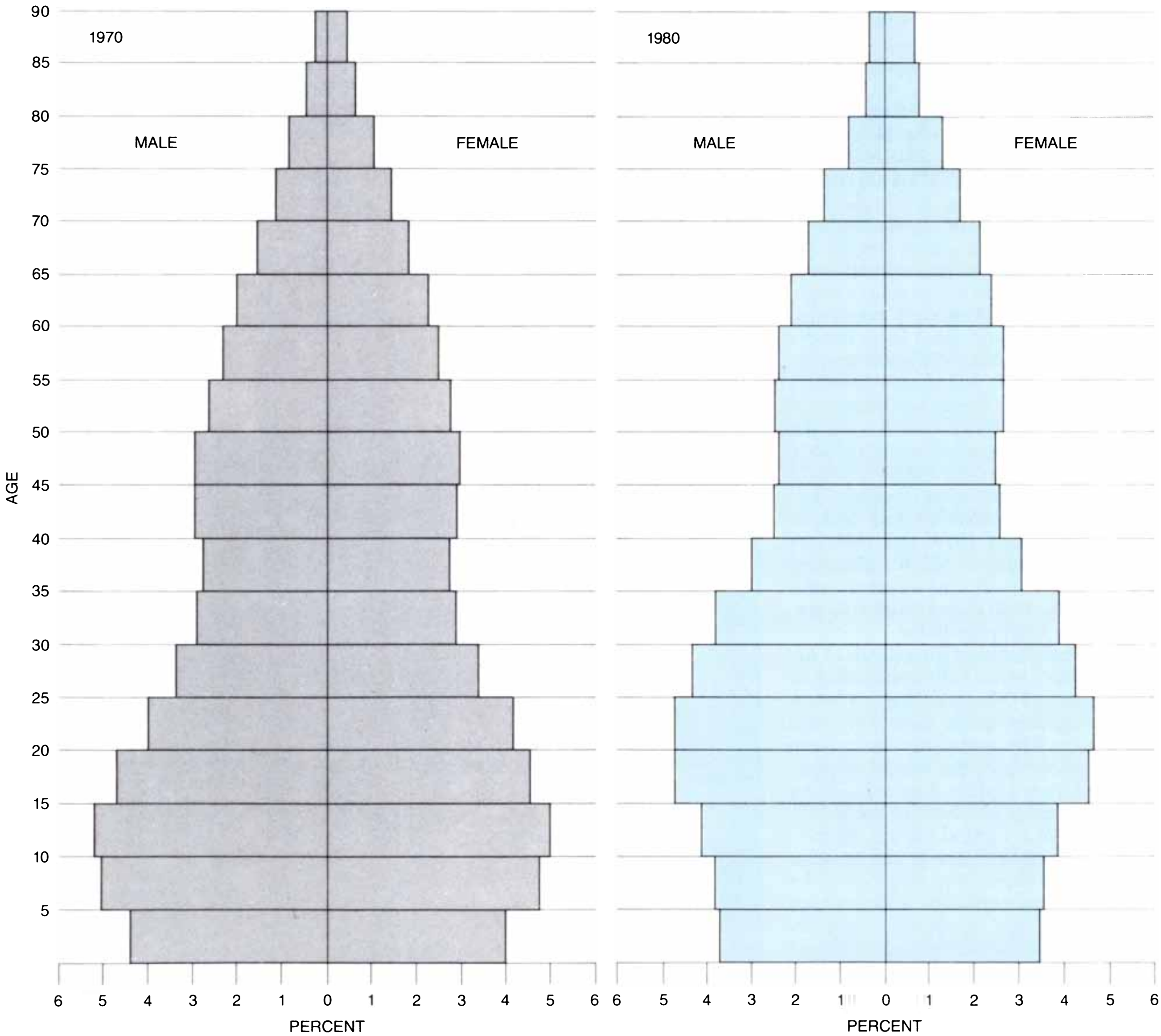
COUNTIES OF THE MIDDLE WEST were the site of much demographic change between the 1970 and 1980 censuses. This map shows alterations in the population of individual counties; it is a detail of the top map on page 54. Farm counties in Nebraska, South Dakota and Iowa, where farm holdings are being consolidated, were areas of population loss. Several of the central cities also lost population. The population of many suburban areas grew, but at a rate lower than the national average. Rural areas and small towns showed much more rapid growth.

remaining four large SMSA's in the Northeast had minimal gains of population. In the North Central region five of the 11 largest metropolitan areas lost population; in the Cleveland SMSA the loss was 8 percent. In striking contrast, none of the metropolitan centers of the West or South lost population, and some gained substantially. The SMSA that includes Fort Lauderdale and Hollywood, Fla., for example, grew by 62 percent, and the population of the Phoenix SMSA increased by 56 percent.

Within the SMSA's there were sharp variations in population growth according to whether a county is within the city limits or is part of the suburban ring. The population of the 429 central cities increased by only .1 percent while the population of the suburbs was increasing by more than 18 percent. These shifts too were influenced by the regional trends. Among the central cities of the 50 largest SMSA's those in the Northeast and North Central regions declined in population by 12 percent and 14 percent; suburbs in these areas grew by 1 percent and 10 percent. In the South and West the central cities of the largest SMSA's grew by 2 percent and 8 percent; the populations of the suburbs grew by 35 percent and 24 percent.

That these findings reflect a new phase of population growth and distribution and are not merely an extension of older patterns has been shown by several demographers. Daniel R. Vining and Anne Strauss of the University of Pennsylvania have demonstrated that in the 1970's the population of the U.S. for the first time became less concentrated when its distribution was measured in geographic units of all sizes, from counties to major census regions.

The concentration of a population may well diminish for some geographic units and rise for others in the same period. From 1900 to 1970 the population



AGE STRUCTURE of the U.S. population changed appreciably between censuses, as is shown in this diagram in which the age distribution of 1980 is compared with that of 1970. During the decade the median age of the population rose to 30 years, a level it has reached only once before, in 1950. The change was due largely to the aging of the exceptionally large cohorts born during the "baby boom" that followed World War II. The members of the baby-boom cohorts are now between 25 and 34, and there was a sharp rise in the proportion of the population in this age group. Subsequent cohorts were much smaller, and there has been a decrease in the proportion of the population under 15 years old. Past fluctuations in fertility and increases in longevity resulted in the growth of all age groups older than 50.

of the U.S., measured in large regions and states, became steadily less concentrated, reflecting the overall movement from east to west. In counties and in economic regions of states, however, the population became progressively more concentrated, as a result of the movement to the cities. Since 1970, however, as Vining and Strauss have demonstrated, the population has become less concentrated in geographic units of all sizes.

Coming after a long period of urbanization, reductions in population concentration may indicate a new pattern of settlement in the U.S. The dispersal of population that is central to this hypothesis was confirmed in the 1980 census returns. The proportion of counties that lost population was relatively small: about 20 percent.

One possible explanation of the growth in nonmetropolitan areas is that it is merely a result of the further geographic expansion of the suburbs. This hypothesis has been tested and found inadequate by Calvin L. Beale of the Department of Agriculture. In an analysis of the preliminary census data Beale showed that counties adjacent to metropolitan areas did grow fairly rapidly during the decade; their average population increase was 17 percent. Counties not adjacent to SMSA's, however, grew almost as fast, at an average rate of 14 percent. These counties are presumably beyond the commuting range of cities; their growth is not part of the continuing expansion of the urban centers.

What is responsible for the dispersal of the American population that began in the 1970's? Several forces seem to have contributed to this tendency. Beale has suggested that the major factors in the accelerated growth of the nonmetropolitan areas were "energy and other mining developments, resort activities, retirement, urban flight and comparatively high birthrates."

Daniel Garnick of the Department of Commerce has pointed out that the changing character of industrial production may be particularly important in the recent shifts in population distribution. The movement of industry to the "sun belt" areas, and the consequent rapid growth of employment there, has certainly been a factor in the rapid population growth of the South and West. It must also be kept in mind, however, that the flow of illegal immigrants has contributed to the population growth in these regions.

Although the sharp break with the patterns of population movement that had prevailed for much of the previous 70 years is the most remarkable aspect of the 1980 census results that are now available, several more gradual and predictable trends are worthy of note. The median age of the population rose to 30 for the second time since the census was begun; as of 1980 the median was exactly 30 years. The change in the age structure of the population reflected not only increased longevity but also decreased fertility and the aging of cohorts born during the "baby boom" of some 35 years ago. The groups born immediately after World War II were exceptionally large; they were succeeded after 1964 by much smaller cohorts. As a result the proportion of the population aged 25 to 34 accounted for about half of the increase in total population during the decade; the same will be true of those aged 35 to 44 in the 1980's. There has been a sharp decrease in the population younger than 15, a trend that will also persist in the 1980's.

	UNDER 15 (PERCENT)	65 AND OVER (PERCENT)	MEDIAN AGE
U.S.	22.6	11.3	30.0
White	21.3	12.2	31.3
Black	28.7	7.9	23.2
Spanish Origin	32.0	4.9	23.2

AGE STRUCTURE OF RACIAL AND ETHNIC GROUPS varies widely, partly as a result of differences in fertility during the past decade. The white population had the lowest fertility of the three groups between 1970 and 1980. Whites now have the largest proportion of older people. The higher fertility of the black population has resulted in a median age of about 23. Hispanics have the highest fertility and the youngest age distribution of the three groups.

Partly as a result of a changing age structure, the numerical balance between the sexes continued to shift toward a larger proportion of women during the decade. In 1980 there were 94.5 males for every 100 American females, a slight decrease from the ratio of 94.8 in 1970.

The fertility of American women reached its lowest level since the end of World War II in 1976; it had risen slightly by the end of the decade. In 1980 American women were having an average of 1.9 children by the end of their reproductive years.

Fertility still varies a good deal according to ethnic and racial group, however. Changes in the definition and reporting of race and Spanish origin make it difficult to compare the growth rates of racial and ethnic groups in the past decade. It is clear, however, that the growth rates of blacks and Hispanics exceeded the growth rate of whites. Accurate comparisons must await Bureau of the Census refinements and corrections of the returns.

In 1980, 83 percent of the population described themselves as white, 12 percent as black and 5 percent as members of other races. Six percent identified themselves as being of Spanish origin (which is an ethnic rather than a racial grouping). As a result of immigration and recent higher fertility both the black and the Hispanic populations have younger age distributions than the white population has.

The geographic distribution of blacks and Hispanics changed somewhat during the decade. By 1980 black migration into the central cities had become minimal; increases in the black population of those areas were mainly due to births. The black population of the suburbs, however, grew rapidly (from a small base) in the late 1970's. The black population of rural areas and small towns remained roughly constant. Data from the census indicate that blacks who migrate from nonmetropolitan areas are now going to the metropolitan centers of the South and West rather than to those of the urban North as they had in earlier decades.

The 1980 census returns indicate that of the 226.5 million people in the nation 220.8 million lived in households and 5.7 million lived in group quarters. The number of households was 80.4 million; thus the average number in each household was 2.75. During the decade the population living in households increased by 12 percent. The number of households increased by 27 percent, and the average number of people in each household, continuing a long-term trend, declined by about 13 percent. Social and economic factors responsible for the reduction in household size include the decreased birthrate, the increasing frequency of divorce and the growing number of people who choose to live alone. The continued decrease in the number of extended families is also a factor.

The Microchannel Image Intensifier

It is a glass wafer, perforated by millions of electron-multiplying tubes, resembling a compound eye. It can transform a dim pattern of electromagnetic radiation into a brightened, pointillist image

by Michael Lampton

An amplifier is a device that boosts the power of a signal while otherwise changing it as little as possible. Amplification is usually thought to apply only to signals that vary in time, such as the electrical output of a microphone. A visual scene constitutes a signal that varies in space as well as in time; in order to amplify such a signal it has often been necessary to convert it into a purely temporal pattern. This can be accomplished by scanning the scene in some systematic way, as in a television camera. In the past 25 years, however, several advances in glassworking technology have made it possible to assemble millions of microscopic amplifiers into a geometric array. Each amplifier in the array is a tube or channel about 15 micrometers in diameter that can brighten a small, well-defined portion of the scene. Hence the entire array of amplifiers, operating simultaneously and in parallel, functions as an image intensifier, making faint images brighter without destroying the spatial information of the input signal. Such an array of microscopic amplifiers is called a microchannel plate.

Much of the economic motivation for the development of the microchannel plate has been military. A compact image intensifier that incorporates a microchannel plate can increase the brightness of a scene by a factor of up to 10,000, thereby making possible surveillance at night under darker conditions than ever before. Goggles, binoculars and the like incorporating microchannel intensifiers have become essential in nighttime military operations where, for example, vehicles must be driven without headlights. The U.S. Army Night Vision Laboratory at Fort Belvoir, Va., has been chiefly responsible for the continuing development of these devices. At present such applications account for the production of about 10,000 microchannel intensifiers per year.

The microchannel intensifier has nonmilitary applications as well. In astronomy it can enhance the brightness of an image to the point where individual photons can be detected. As a result an image that would otherwise require a prohibitively long photographic exposure can be recorded in minutes. Image intensifiers have also been made available to people suffering from the form of night blindness called retinitis pigmentosa, although the cost (about $2,700) remains a barrier to their wider distribution.

The microchannel plate is not limited to the intensification of visible images. When it is coupled with suitable electrical and optical systems, it can convert a two-dimensional signal in the near-infrared or the ultraviolet region of the electromagnetic spectrum into a visible image. Moreover, it can directly detect X rays, ions and electrons, making possible a wealth of new applications in such areas as oscilloscope design, electron microscopy and the study of the chemical composition of materials by photometric methods. A microchannel plate is employed in the X-ray telescope aboard the Einstein Observatory satellite and in the ultraviolet spectrometers carried aboard the Voyager missions to Jupiter and Saturn. The European X-ray Observatory Satellite (*EXOSAT*), to be launched next year by the European Space Agency, will carry microchannel plates as part of its complement of X-ray detectors.

What are the principles that underlie the operation of a microchannel plate? It is useful to begin the answer with a description of an image intensifier that does not incorporate a microchannel plate. Such a device depends on an interaction of electromagnetic energy with matter that was first explained by Einstein in 1905. The interaction is the photoelectric effect, and it was one of the first physical processes to be accounted for by the idea that energy comes in quanta, or discrete packets.

When a photocathode, or negatively charged electrode, is exposed to electromagnetic radiation whose wavelength is shorter than some critical value, the photocathode emits a current of electrons that can be collected by a nearby anode, or positively charged electrode. For many metals the critical wavelength lies in the ultraviolet region of the spectrum, although for some substances it lies in the visible region or in the near-infrared region. The rate at which the charge is transferred is proportional to the intensity of the radiation, but whether or not the charge is transferred at all depends only on the wavelength.

Einstein explained this phenomenon by supposing that an electron bound in matter requires some definite minimum energy to liberate it. Moreover, when the liberating energy is supplied by electromagnetic radiation, it can be absorbed only in the discrete quanta called photons. From the earlier work of Max Planck it was known that the energy of a photon varies inversely with the wavelength of the radiation. Hence a photon with a wavelength longer than the critical wavelength has insufficient energy to knock an electron out of the metal.

An image intensifier exploits the photoelectric effect to convert the stream of photons in electromagnetic radiation into a stream of electrons. Each photon can stimulate the emission of at most one electron, and so it might seem that nothing would be gained by this strategy. What makes possible the amplification of the image is the electric charge of the electron. In an applied electric field the electrons can be accelerated, so that their kinetic energy is increased. In this way the total energy available for forming an image can be multiplied many times.

In its simplest form an image intensifier is a vacuum tube with two electrodes that are connected to an external power supply. No microchannel plate is incorporated into such a device. An image is focused onto the photocathode, and each area of the photocathode emits electrons in numbers determined by the brightness of the image in that area. Each electron is accelerated by an electric field generated by a bias potential of several thousand volts applied between the two electrodes. The accelerated electrons travel across the evacuated gap to the anode, following paths that are approximately straight lines. In this way the flux of electrons that reaches the anode corresponds to the brightness variations of the image on the photocathode. The spatial resolution of the image is well preserved because the photocathode and the anode are close together.

An image tube has a visible output because its anode is made of a fluorescent material. When an electron strikes the anode, it excites and ionizes atoms in the material. A portion of the energy is then released in the form of photons of visible light as the electrons from the excited atoms return to their usual energy levels. The emitted photons constitute the image that is visible on the fluorescent anode.

I pointed out above that each photon striking the cathode can give rise to one electron at most; actually the efficiency of the photocathode is only about 10 percent, that is, only about one photon in every 10 stimulates the emission of an electron. Moreover, at the anode only about 30 percent of the energy conveyed by the accelerated electrons is converted into light. In spite of these inefficiencies there is a net gain, or amplification of the luminosity of the image, because the increase in the energy of each emitted

PHOTOGRAPH OF A NIGHTTIME SCENE illuminated by starlight was made with a 35-millimeter camera fitted with a microchannel image intensifier. The brightness of the photograph approximates the brightness of the scene as it would appear to the eye through the intensifier. The green color of the sky and the treetops in the photograph is the color of the fluorescent light emitted by the screen on which the amplified image can be viewed. The color corresponds to the wavelength of light to which the eye is most sensitive, namely about 540 nanometers. The photograph was made with a daylight color film having an ASA film-speed rating of 400, exposed for one-fourth of a second at a lens-aperture setting of $f/2.8$. Without the image intensifier such a photograph would require from four to eight minutes of exposure. The intensifier was provided by Paul Lighty of the ITT Electro-Optical Products Division in Roanoke, Va.

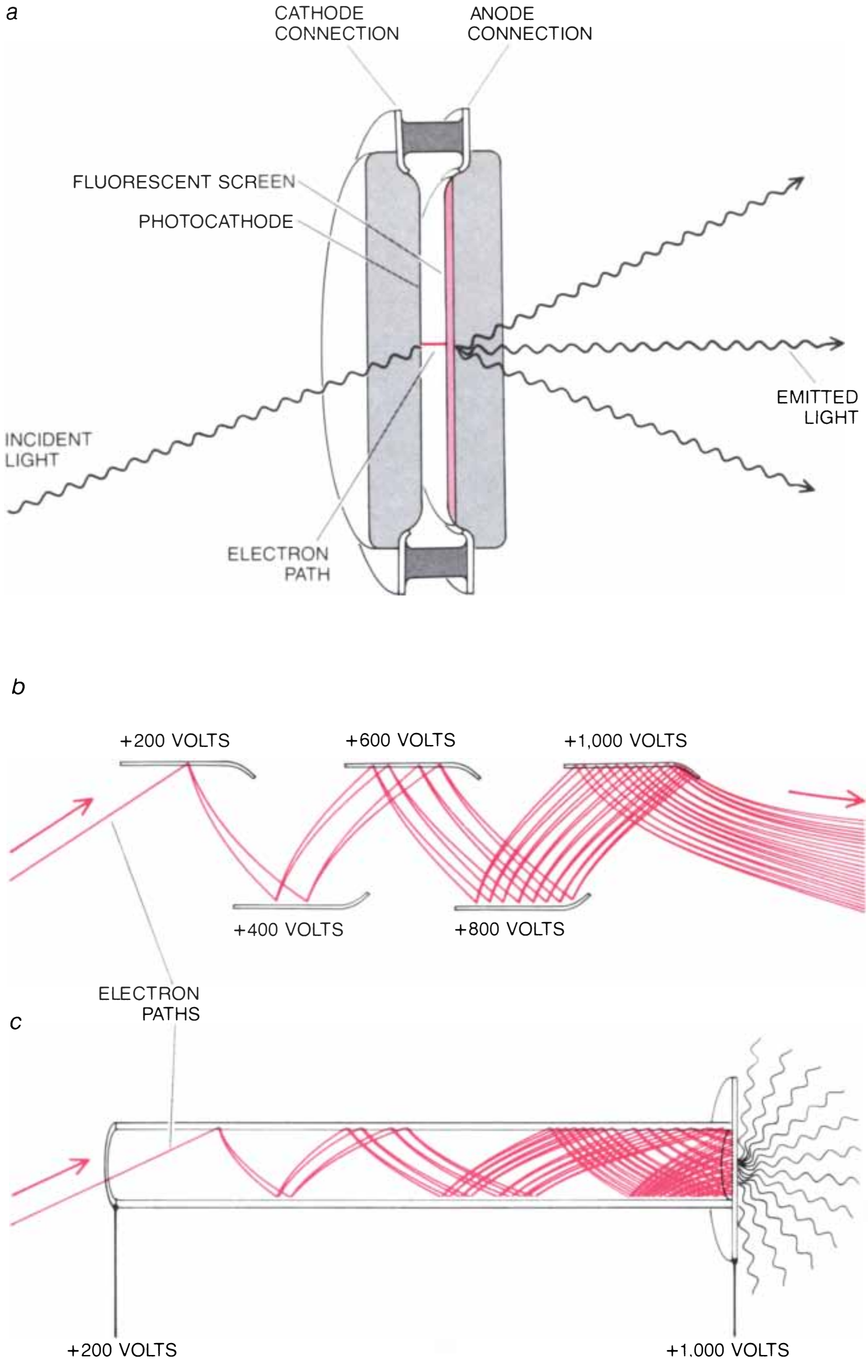


IMAGE INTENSIFIERS exploit the photoelectric effect, whereby the energy of incident photons, or quanta of electromagnetic radiation, is converted into the energy of moving electrons. Because electrons, unlike photons, are electrically charged particles their energy can be increased by acceleration in an electric field. When photons strike a photosensitive metal plate, electrons are ejected into an evacuated region. The electrons are accelerated toward a positively charged fluorescent plate, where their energy is converted again into the energy of visible light. In a two-electrode image tube (*a*) at most one electron is accelerated across the gap for each incident photon, and the electron must travel directly across the gap if the image is not to be blurred. Hence to preserve the sharpness of the image the gap must be narrow. The achievable gain is limited by the voltage that can be applied across the narrow gap without sparking. Electrons that strike certain materials can cause the ejection of additional electrons, and a cascade of electrons can thus be induced by maintaining several electrodes at successively higher voltages (*b*). Although such an electron multiplier can yield high gain, the dispersion of the cascading electrons destroys information about the spatial distribution of the incident photons. A way to achieve both high gain and spatial resolution is to provide one electron multiplier for each pixel, or picture element, in an image, but then the multipliers must be made very small. This can be done with the continuous-channel electron multiplier (*c*), which functions much like the multiple-electrode one but confines the electron cascade to the bore of the channel. Techniques for drawing glass fibers can be employed to make such multipliers that are only a few micrometers in diameter; arrays of the multipliers can then be installed in image tubes.

electron can be several thousandfold.

It may seem that an image signal could be amplified to any level desired by increasing the bias potential that accelerates the electrons. Unfortunately a two-electrode image tube cannot be operated beyond a certain maximum applied voltage. When the voltage exceeds this level, charge begins to flow spontaneously across the small gap between the photocathode and the anode and overwhelms the smaller current of electrons that transmits the image. The effect could be overcome by redesigning the tube to have a larger gap, but then the spatial resolution of the image would deteriorate because of small, random deviations in the path taken by the electrons. By focusing the electrons with electric and magnetic fields it is possible to increase the gap, to operate at a higher voltage and so to increase the gain in image brightness. The electric and magnetic fields must be carefully controlled, however, and such focusing devices can cause geometric distortion of the image or nonuniformities in its brightness. Moreover, for many purposes in physics and astronomy even the more complex intensifiers that focus the electrons have insufficient gain.

The fact that each incident photon can cause at most one electron to be transferred to the anode is a fundamental limitation of such image tubes. In certain materials, however, electrons can be knocked free by an incoming electron just as they can by an incident photon. The effect is called secondary surface emission, and it can be exploited in an image tube to multiply the number of electrons that reach the anode.

Electron multipliers have been employed since the early 1950's as the amplifying components of photomultiplier tubes. Metal electrodes called dynodes are given progressively higher positive potentials and are arranged so that the electrons leaving one dynode are directed toward the next. An incident electron that strikes the first dynode knocks several low-energy electrons out of the metal. The secondary electrons are accelerated to the second dynode, where they are multiplied again. Successive multiplications lead to an exponential growth in the electric charge liberated by the initial electron. Because all the electrons are accelerated by the field, the combined energy of the electrons that finally strike the anode can be enormously greater than it is when a single electron is accelerated.

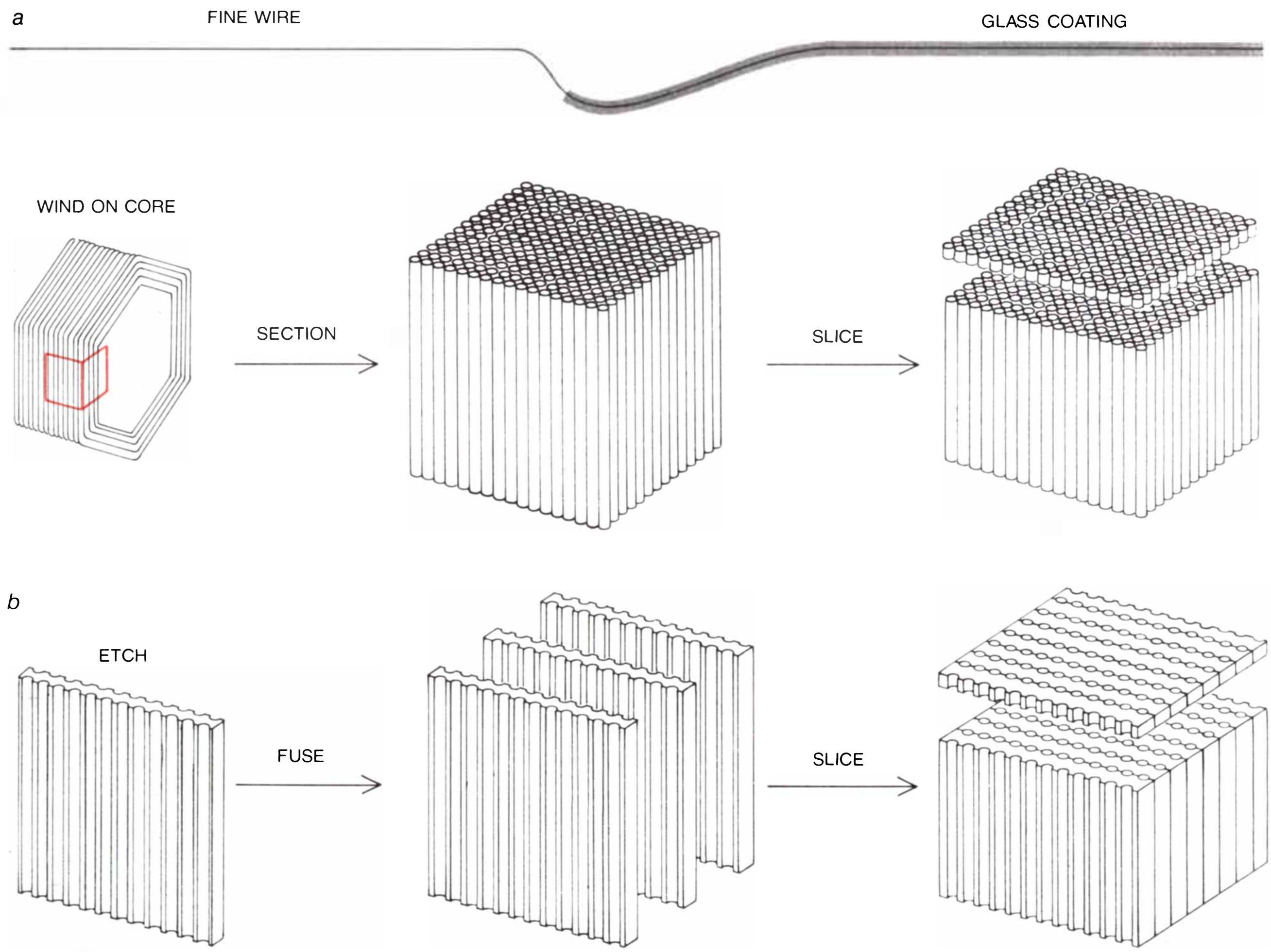
High amplification can be attained in a dynode multiplier structure, but only at the expense of spatial resolution. The geometry of a dynode does not ensure that the electrons travel in a straight line; as a result electrons emitted from a given region of the photocathode do not necessarily land in the corresponding region of the anode. The microchannel plate is a device that combines the gain of an electron multiplier with the spatial resolution of an image intensifier. It consists of millions of independent electron multipliers assembled in a two-dimensional array. There is no attempt to maintain spatial resolution within any one multiplier; instead each multiplier corresponds to a single pixel, or picture element. The intensified image is a pattern of fine dots of varying brightness.

The key step in the development of the microchannel plate was the invention of the continuous-channel electron multiplier. It is a glass tube connected at each end to a source of electric power, supplied with a bias potential of about 1,000 volts and placed in a vacuum. It acts, however, much like a multiplier assembled from dynodes. An incident electron can initiate a cascade of secondary electrons that grows exponentially as it progresses along the inside of the tube. The cascade is entirely confined to the interior of the tube or channel. Such continuous-channel multipliers were first investigated at the Bendix Research Laboratories in Southfield, Mich., by George W. Goodrich, James R. Ignatowski and William C. Wiley from 1959 to 1961.

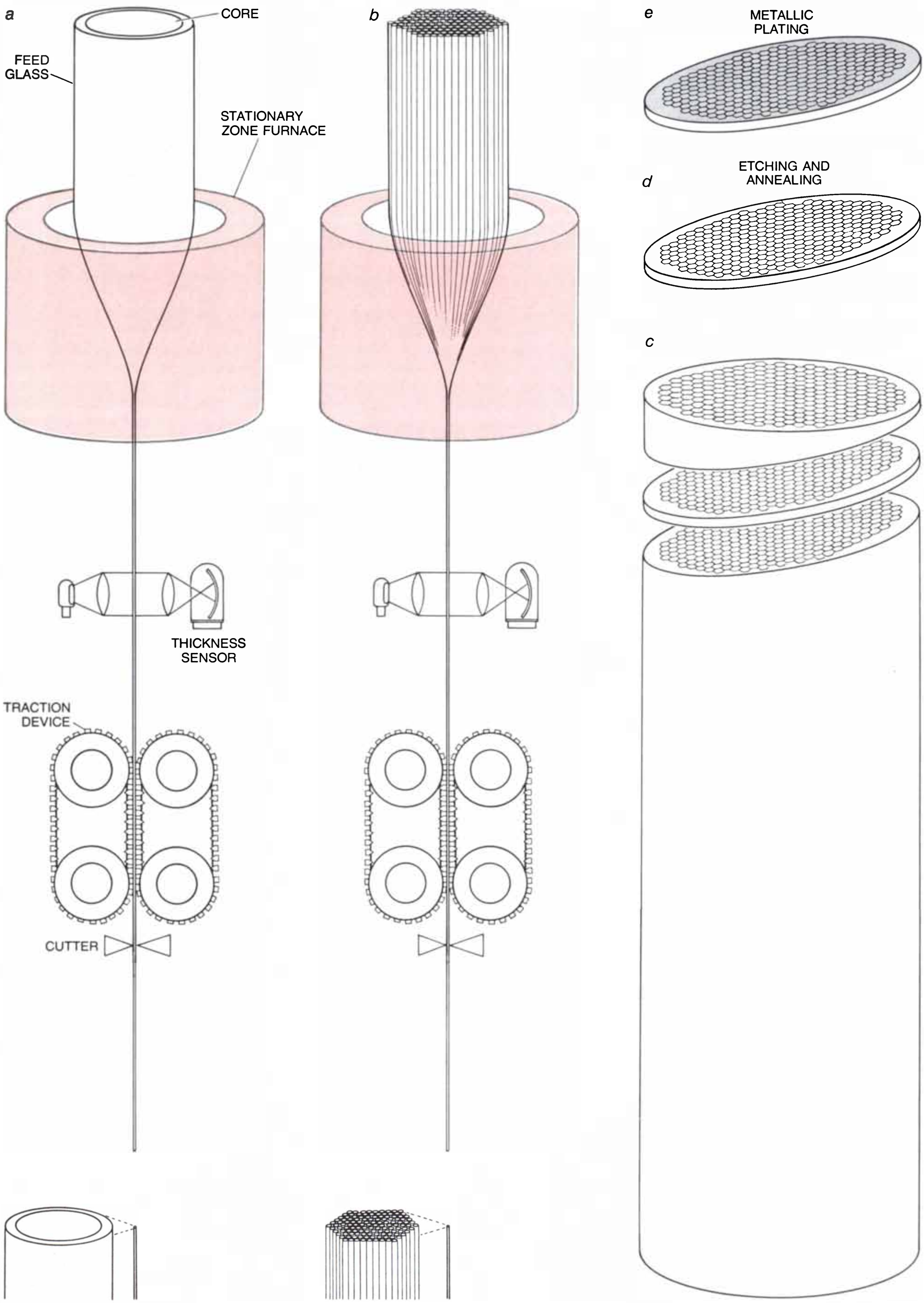
Materials for making a channel multiplier must satisfy two requirements. First, the wall of the channel must be able to emit more electrons than it absorbs. In the practical range of electron energies a variety of materials, including many glasses, emit an average of about two electrons for each incident electron. Second, the electrical conductivity of the material must be predictable and controllable, so that the charge removed from the channel wall can be replenished and the uniform electric field can be reestablished. The second requirement presented a considerable challenge to early experimenters.

In a glass an electric current can be conducted by the diffusion either of free electrons or of ions (charged atoms). To obtain controlled conduction the ionic conductivity of the glass must be minimized. The diffusion of ions is a kind of electrolysis that leads to unwanted chemical activity at the electrodes, such as the depletion of alkali metals, and the polarization of electric charges within the glass. Hence conduction of electricity solely through the motion of free electrons must be promoted.

Although many glass recipes have been employed in making channel multipliers, by far the most successful has been a mixture of about 50 percent lead oxide, 40 percent silicon dioxide and smaller quantities of several alkali oxides. In the absence of further treatment glass of this composition has a high electrical resistivity, or in other words a low conductivity. The lead atoms in the glass are stable either as neutral atoms or as positive ions that have given up electrons in forming an ionic chemical



SEVERAL UNSATISFACTORY TECHNIQUES were proposed for fabricating fine tubes of glass and assembling them into a large array before a practical method evolved. In one technique (*a*) a wire was coated with molten glass; after cooling the resulting fiber was wound on a spool. A section of the densely packed fibers was cut from the spool and the section was sliced into wafers. Finally the wire cores were etched away in acid, leaving a perforated glass wafer. It proved difficult, however, to maintain the uniformity of the fibers during the winding, and the need to manipulate microscopic fibers from the first step of the process made it economically unsound. A second technique (*b*) employed photolithography to etch thin plates of glass, which were then stacked, fused and sliced into wafers. The depth and the width of the channels, however, could not be controlled during the etching process to within the required tolerances.


a
CORE
FEED GLASS
STATIONARY ZONE FURNACE
THICKNESS SENSOR
TRACTION DEVICE
CUTTER
b
e
METALLIC PLATING
ETCHING AND ANNEALING
d
c

bond with oxygen. The conductivity can be improved by removing the oxygen from the lead in the chemical reaction called reduction.

Reduction can be carried out at the surface of the glass by heating it in an atmosphere of hydrogen gas. After several hours at 400 degrees Celsius the reduction of the lead oxide has penetrated to a depth of several tenths of a micrometer and most of the reduced lead has evaporated from the surface. The lead that remains on the surface coalesces into metallic clusters, giving the glass a characteristic black color. The electrons that once formed the ionic bond of the lead oxide are freed to carry electric current, and the surface layer of the glass has become a semiconductor. Its resistivity is from 10^8 to 10^{14} ohms per square, a range suitable for the manufacture of channel multipliers and microchannel plates. (Surface resistivity is expressed in ohms per square, without specifying the dimensions of the square, because the resistivity of a square region on a surface of uniform composition does not depend on the size of the square.)

The first experimental arrays of continuous-channel multipliers were assembled by binding together a few dozen channel multipliers with glass that has a low melting point. Each channel was about a millimeter in diameter, so that the dot pattern that could be resolved with the early instruments was rather coarse. Moreover, to reproduce a high-quality image many more amplifying elements are needed. One array, laboriously assembled by hand at the Bendix Research Laboratories, was made up of about 5,000 channels hexagonally packed with .15-millimeter spacing between the channel centers. Although the dot resolution of this array exceeds the quality usually attained in the reproduction of photographs in newspapers and magazines, the area of the image is only about a square centimeter.

It might seem that the outcome of experiments with devices on the scale of the early instruments could not be extrapolated to microscopic arrays of channel multipliers capable of transmitting high-quality images. The extrapolation is possible in part because for a given applied potential the gain of a channel multiplier does not depend on its size. In a longer channel each electron falls farther under the influence of the force exerted by the electric field, so that it has more time to gain energy. In a shorter channel the electric field is more intense: the force acting on the electron varies inversely with the distance between the electrodes. The two effects cancel each other exactly.

The handmade arrays established that the individual channels in an array respond independently of one another. It was found, however, that the quality of the image is quite sensitive to variations in the diameter of the channels and also somewhat sensitive to nonuniformities in their spacing. In the early 1960's the main challenge to the development of microchannel plates was to devise manufacturing techniques that could maintain high uniformity in the diameter of the channels.

One technique invented at the Mullard Research Laboratories in London and developed at the Laboratory of Electronics and Applied Physics near Paris was the metal-core process. A fine, uniform wire is coated with heat-softened channel-multiplier glass and wound onto a polygonal drum. When a thick layer has been built up, it is cut into blocks. The blocks are then stacked side by side and fused into a boule of multiplier glass in which many thousands of wires are embedded. The boule is thinly sliced and etched in dilute acid to remove the metal wires. The resulting glass wafers are penetrated by thousands of channels of highly uniform diameter.

The metal-core process has two major disadvantages. First, although the diameter of the channels is uniform, their spacing is not. It is difficult to wind the fragile, glass-coated wire onto the core with perfect uniformity. The second disadvantage is that in all stages of the manufacture of the glass fiber it has its final, microscopic diameter. Hence the coating and winding machines, which represent a sizable capital investment, can process only a small amount of material per unit of time. This has proved to be a serious economic obstacle.

Another technique, called the grooved-plate process, was investigated in the U.S. and Switzerland. A large number of fine, parallel grooves are etched by photolithography into each of many thin glass plates. The plates are then stacked and fused to form the block from which the microchannel wafers are cut. The method seemed promising at first because the spacing of the grooves can be precisely controlled in photolithographic etching. Moreover, curved or zigzag channels can easily be made, and such shapes overcome certain limitations of gain inherent in the geometry of straight channels. Controlling the width and the depth of the grooves during etching and fusing proved to be difficult, however, and resulting nonuniformities of gain led to the eventual abandonment of the grooved-plate process.

The technique that has proved to be economically suited to manufacturing microchannel plates in quantity is called the two-draw process. It is based on methods for drawing or stretching glass into microscopically fine fibers; such methods began to evolve in Egyptian workshops during the Eighteenth Dynasty (1570–1320 B.C.). Certain glass compositions that can be worked over a wide range of temperatures have the property of preserving their cross section when they are heated and drawn. The Bendix investigators first exploited this property for the fabrication of microchannel plates in 1960, and it has since been adopted by other manufacturers.

In the two-draw process channel-multiplier glass is cast or extruded to form a cylindrical ingot, then cooled and ground into a uniform rod several centimeters in diameter. The rod is bored along its axis, and another rod, made of glass with a different composition, is usually fitted into the bore. At the end of the process the core is removed by etching it away in a bath of hot dilute acid, but during the intervening stages of manufacture it supports the outer cylinder so that the finished microchannels are more nearly uniform in size and shape.

During the first draw the glass cylinder is suspended vertically in a zone furnace, where the temperature can be controlled from point to point. The bottom of the cylinder is heated to about 500 degrees C. and a gob of soft glass descends from the furnace, suspended by a drawn-glass fiber whose diameter is about a millimeter. By the time the fiber has descended several meters below the furnace it is cool enough to be handled by a traction machine that regulates the speed of the draw. Regulating the speed in turn gives precise control of the diam-

TWO-DRAW PROCESS creates microchannel plates by drawing fibers from a heated boule of glass. The fibers retain the cross-sectional geometry of the boule, although on a smaller scale. A cylindrical tube of glass about 50 millimeters in diameter is fitted with a core made of a glass that can later be etched away in acid. The tube is heated and drawn to a thickness of about a millimeter (*a*). Several thousand 15-centimeter lengths of drawn tubing are assembled by hand into a hexagonal bundle about 50 millimeters across. The block formed in this way is again heated and drawn to a thickness of about a millimeter (*b*). The resulting fiber is a geometrically similar copy of the 50-millimeter hexagon, complete with the fine structure of the assembled glass tubes. The hexagonal fibers are then fused together and sliced at an oblique angle to create wafers about a millimeter thick (*c*). The wafers are etched in acid to remove the glass cores and heated in an atmosphere of hydrogen gas to develop the desired electrical conductivity on the inner walls of the microchannels (*d*). Metal is evaporated onto both surfaces of the wafer so that electrical connections can be made (*e*). The sizes of the tubes packed into the hexagon and the hexagons assembled on the finished wafer have been exaggerated for clarity.

eter of the fiber. Below the traction machine the fiber is cut into segments about 15 centimeters long. Several thousand segments are assembled into a hexagonal bundle.

The second draw is similar to the first. The hexagonal bundle of fibers is suspended and heated in a zone furnace and drawn into a hexagonal compound fiber about a millimeter across. The spacing between the individual cylinders is thereby reduced to a few hundredths of a millimeter, their final spacing in the microchannel plate. Again the drawn glass is cut into segments, which are packed together and fused in a vacuum to form a boule of hexagonal solids. The boule can be as large as 125 millimeters in diameter and can incorporate millions of microchannels.

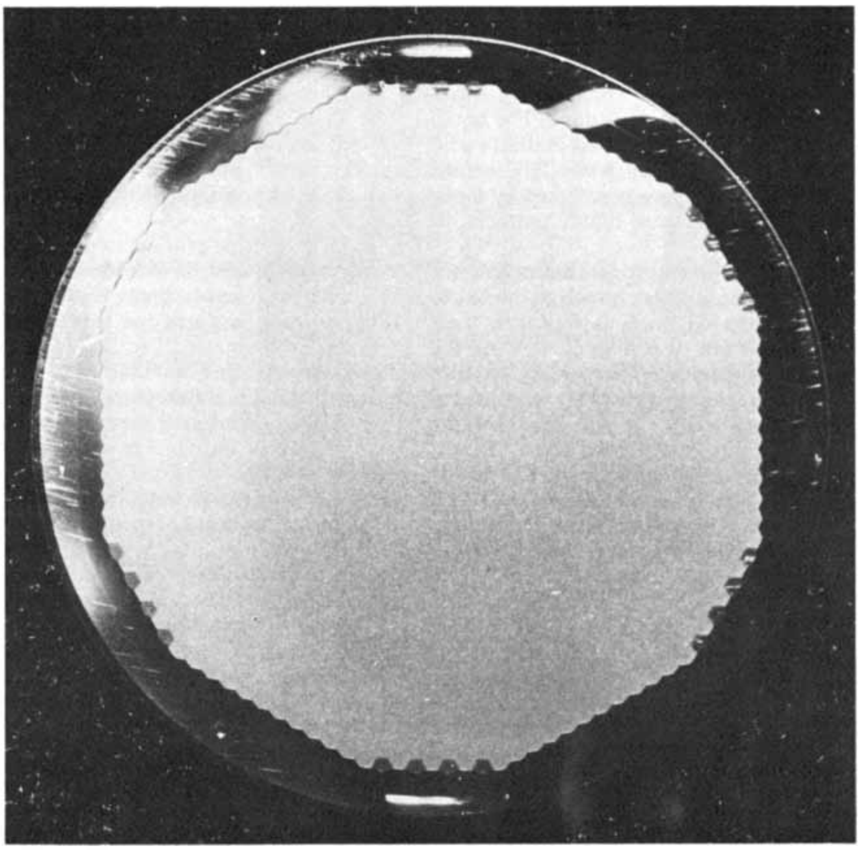
Microchannel plates are made by slicing the boule into wafers about a millimeter thick and polishing the faces of each wafer. The slice is usually made at an oblique angle to the axis of the microchannels so that in the finished plate electrons will collide with the channel wall near its input end rather than flying straight through to the anode. Grinding, drilling and hot-working can alter the wafer to meet special needs. In the final steps the core glass is dissolved in acid and the plate is heated in hydrogen gas to reduce the lead oxide and give the glass surface the required electrical conductivity. The wafer is finished by evaporating a metal film onto both faces so that electrical connections can be made to all the channels.

There are now two basic kinds of image intensifier that incorporate a microchannel plate. The proximity-focused intensifier resembles a simple two-electrode intensifier into which a microchannel plate has been inserted. The incident image is focused onto the photocathode, the emitted electrons are accelerated across a small gap into the microchannel plate and the more numerous electrons emerging from the plate are accelerated across a somewhat wider gap onto a fluorescent screen. Such a tube is compact, free from distortion and unaffected by magnetic fields. Moreover, like an image formed by contact printing on photographic paper, the output image is flat and always in focus.

The proximity-focused intensifier is particularly well suited for making photographs of high-speed phenomena. The microchannel plate intensifies the faint image that results from the brief photographic exposure time. In addition the device can serve as a fast camera shutter: the externally controlled voltage difference between the photocathode and the front of the plate provides a means of rapidly switching the intensifier on and off. A proximity-focused intensifier can respond to an electrical shutter pulse lasting for only a few nanoseconds. Hence an unblurred photograph can be made of an extremely short-lived event.

The second kind of image intensifier employs a nonuniform electrostatic field to focus the electrons emitted by the photocathode onto the microchannel plate. Like an optical lens, the tube inverts the image and can also magnify it; in some cases the device can even act as a zoom lens, that is, one that simultaneously adjusts the magnification and the focus. Because of the inversion of the image the electrostatically focused intensifier is particularly useful in conjunction with optical devices that also invert an image. The microchannel plate can replace a complex optical system, such as the one in binoculars, that rights the inverted image formed by an ordinary lens.

Although the microchannel plate has been applied chiefly to image intensification, it can amplify any form of radiation capable of initiating a cascade of electrons in the channels. Ions and electrons as well as photons at wavelengths in the ultraviolet and X-ray regions of the spectrum can all initiate electron cascades in the microchannels. Such forms of radiation can therefore be sensed without a photocathode. The radiation is brought to a focus on the front face of the plate and the cascading electrons are detected at the rear face. If the plate is equipped with a fluorescent anode screen, it can function as an image converter, making visible the spatial pattern of radiation outside the visible part of the spectrum.



FINISHED MICROCHANNEL PLATE is a thin wafer that can be installed in image tubes of various designs. The standard 25-millimeter microchannel plate shown incorporates some three million microchannels that can brighten an image by a factor of about 10,000. The microchannel plate was fabricated by the Galileo Electro-Optics Corporation in Sturbridge, Mass.

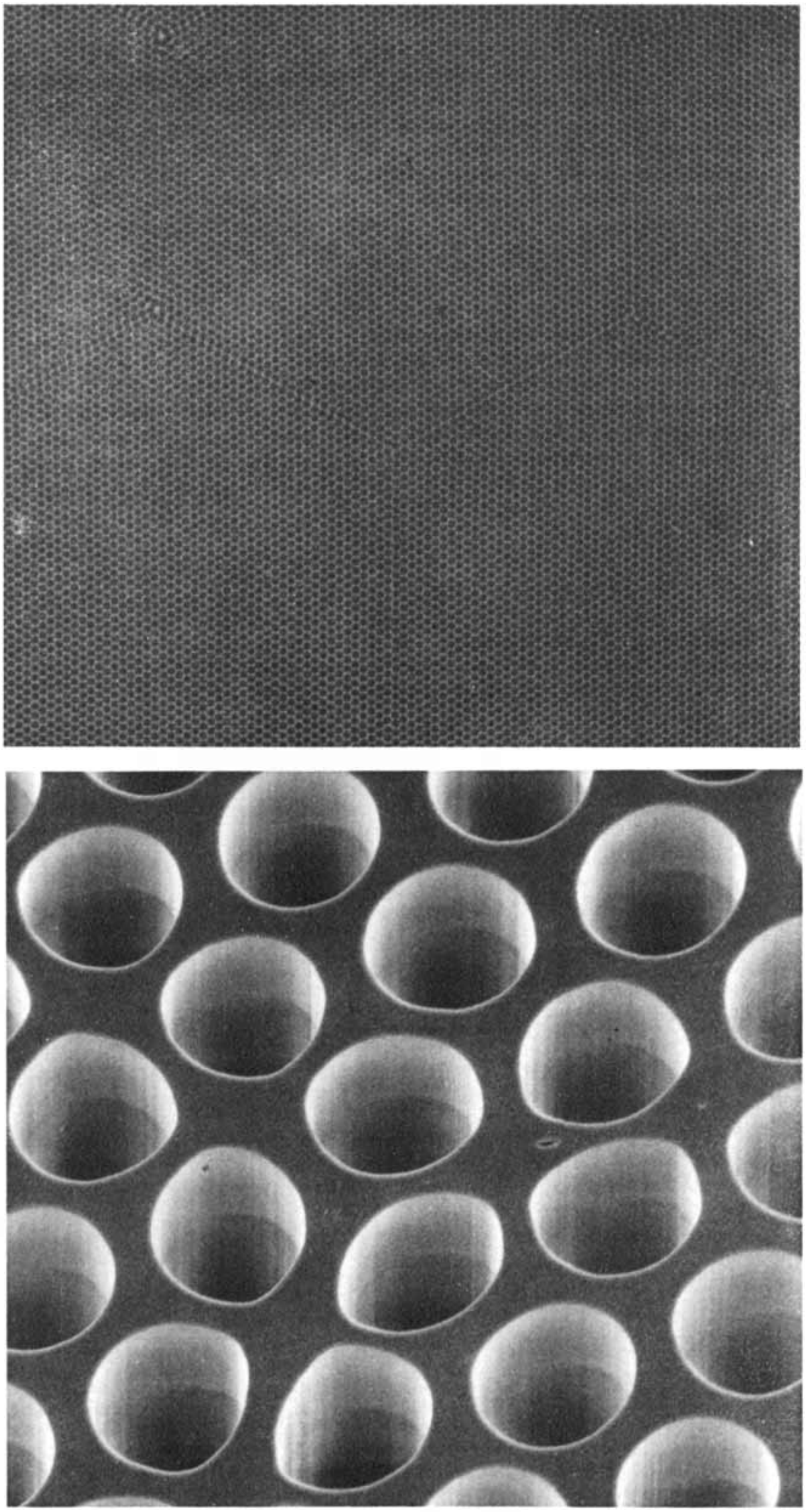
Microchannel plates that act as converters as well as intensifiers have been incorporated into laboratory instruments such as high-speed oscilloscopes, transmission electron microscopes, field-ion microscopes and mass spectrometers. In these applications the gain of the microchannel plate makes possible the recording of extremely brief, faint or low-contrast features. Moreover, it allows the instruments to be operated with a lower beam current, which is sometimes an important advantage. In an electron microscope, for example, damage to the material being examined that is caused by the electron bombardment can be reduced.

Four major physical constraints limit the performance of current microchannel plates. First, the average output signal cannot exceed the maximum current that can be sustained in the walls of the microchannel. When the flux of electrons is too great, electric charge removed from the glass is not replaced immediately and the field is modified; as a result the overall gain of the channel is reduced. Bright parts of an image can thereby become saturated and the contrast in such regions of the image is lost. Although the effect is undesirable in astronomy, where a well-calibrated dynamic range is required, it can be beneficial in night-vision devices, where high contrast in bright parts of an image might diminish the sensitivity of the observer's eye to darker regions.

A second limitation of channel multipliers is a phenomenon called ion feedback. When the channel is operating at high gain, gas atoms in the channel can be ionized by collisions with the cascading electrons. (At least a few such atoms are present even in the best attainable vacuum.) A positively charged ion formed in this way is accelerated by the electric field toward the input end of the tube, where it may strike the channel wall and initiate a new cascade. The new cascade can cause further ionization and the additional electrons tend to mask the signal. Moreover, ions that strike the photocathode can shorten its life, although damage of this kind is now routinely prevented by depositing a thin film of aluminum oxide, permeable to photoelectrons but not to ions, over the entrance face of the plate.

One way to minimize ion feedback is to make the channels curve or zigzag. Then the electrons easily cascade toward the anode, but the ions are inhibited from moving toward the photocathode by collisions with the walls. Although the simplest way to cut such channels is by photolithographic etching, the process, as I have indicated, is not technically satisfactory. In applications where high gain is needed two or more plates are often stacked and rotated with respect to one another. Since the wafer faces are not cut at a right angle to the channel bore, each resulting microchannel effectively turns a corner partway along its length. Successive cascades of electrons readily follow the path around the corner but ions do not because they do not initiate cascades of other ions. By this strategy the applied voltage (and hence the gain) can be substantially increased before the threshold of regenerative ion feedback is reached.

In an effort to form curved microchannels and so minimize ion feedback groups at several laboratories are exploiting the predictable variation of glass viscosity with temperature. If a glass microchannel plate is sandwiched between metal blocks that are at different temperatures, a uniform tempera-



UNIFORMITY OF SIZE AND SPACING of the microchannels is critical to the transmission of high-quality images. A microchannel plate made from drawn-glass fibers exhibits good uniformity, as is shown in the two photomicrographs. In the upper photomicrograph, made by Adolf R. Asam of the ITT Electro-Optical Products Division, a microchannel plate is enlarged 110 diameters; its hexagonal secondary structure is clearly visible. In the lower photomicrograph, made by the author, a second microchannel plate is enlarged 650 diameters. The latter microchannels are 40 micrometers in diameter and are spaced 50 micrometers apart.

ture gradient is set up in the plate. Close to the hotter block the glass is more fluid. If the plate is "ironed" by displacing one of the blocks sideways, the microchannels tilt more near the hotter block than they do near the colder one. Hence each microchannel is bent into a curve. Although curved-channel plates are still experimental devices, they can multiply secondary electrons by a factor of about a million without ion feedback, and they exhibit the same resolution of spatial features as conventional microchannel plates.

A third limitation of a channel multiplier is the charge density of the electrons in the channel, which is called the space charge. When the electron cascade reaches a linear density of about 10 million electrons per millimeter, mutual electrostatic repulsion tends to return additional secondary electrons to the surface of the channel before the field can significantly accelerate them. The result is to limit further growth of the cascade. In many applications the effect is actually beneficial because it makes the output pulses of the multiplier more nearly identical. Random fluctuations in the total charge carried by each cascade are thereby suppressed.

The fourth limitation of the microchannel plate governs its efficiency. The surface area of the entrances to the channels is less than the surface area of the entire plate. If circular channels are employed, the ratio of the total channel cross section to the plate area must be less than 91 percent owing to geometric constraints; because of the thickness of the channel walls the ratio in most microchannel plates is only about 55 percent. Thus about half of the input flux strikes the metal-plated web area between the channel entrances.

Certain strategies can be adopted to reduce this loss. If a strong electrostatic field is applied to the front of the plate, the electrons emitted from the web areas can be pulled into an adjoining channel and so initiate a cascade. The entrances to the channels can also be made funnel-shaped by etching the wafer, raising the ratio of channel area to total surface area. Several laboratories are experimenting with square or hexagonal channels, which pack together more efficiently than channels with a circular cross section.

The output of a microchannel plate need not be registered on a fluorescent screen. In many applications, such as the remote sensing of photometric data from a spacecraft, there is a need for an electrical rather than a visual output. The simplest such sensor is the microchannel photomultiplier; it resembles a proximity-focused microchannel intensifier in which a simple metal anode replaces the fluorescent screen. An electrical output pulse is obtained from the anode each time electrons cascade through the microchannel plate. Because of the extremely brief delay between the stimulus and the amplified output pulse the microchannel photomultiplier has been employed for counting individual photons and for measuring the flight time of subatomic particles produced in accelerators.

A conventional photomultiplier has only one connection for its electrical output, and so it is not suited to the sensing of images. A microchannel photomultiplier, on the other hand, can be designed so that the position of an output pulse on the anode can be decoded electrically. Conceptually the simplest way to accomplish this is to provide each pixel with its own anode and detecting circuit. The technique is limited to the output of rather coarse images, however, because the number of anode wires and trigger circuits (which must be equal to the number of pixels) cannot be increased indefinitely.

In order to reduce the number of trigger circuits each pulse can be detected simultaneously on overlapping grids of vertical and horizontal wires. If the pulse triggers the nearest vertical and the nearest horizontal wire in the grids, the number of circuits required varies as the square root of the number of pixels. More complex arrangements can yield

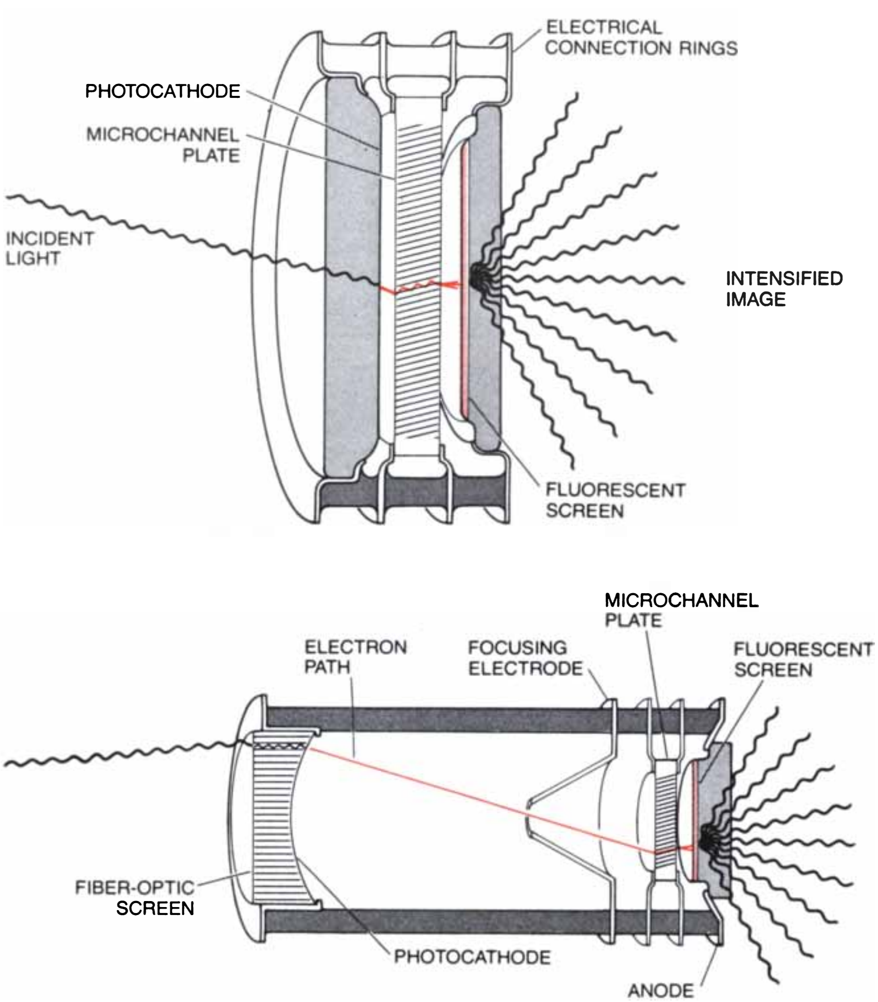
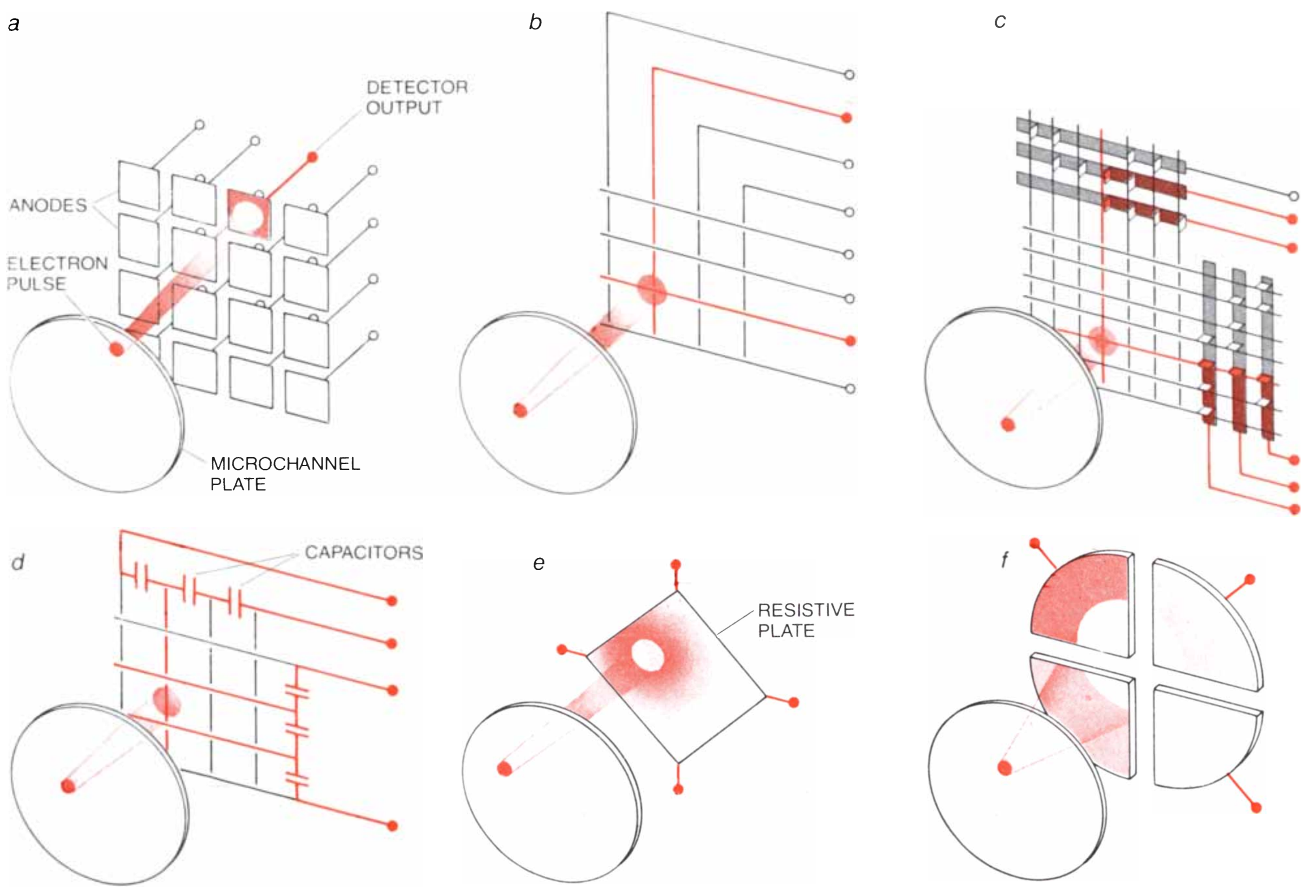
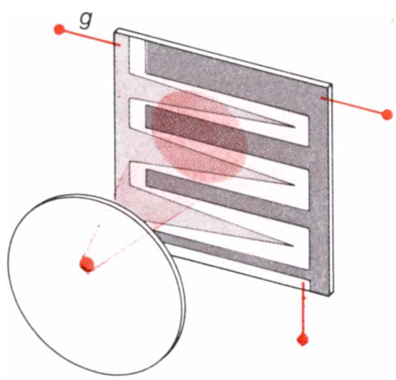


IMAGE TUBES incorporate microchannel plates in two principal configurations. In each configuration the plate is sandwiched between a photocathode, or negatively charged electrode, and a fluorescent anode, or positively charged electrode. The electric potentials of the electrodes on the microchannel plate are intermediate to those of the photocathode and the anode, so that the potentials of the four electrodes become more positive from left to right in the diagrams. The proximity-focused intensifier (*upper diagram*) is similar to a two-electrode intensifier. The electrons emitted by the photocathode cross a short gap to the microchannel plate, where they are accelerated and multiplied. The electrons emerging from the plate strike the anode, again across a short gap. The spatial organization of the input image is preserved because the two gaps are small enough to minimize electron dispersion. Photons enter an electrostatically focused intensifier (*lower diagram*) through a thick fiber-optic window that seals the evacuated interior of the image tube. On the inside of the window is the photocathode; the electrons it emits are focused by an electric field onto the microchannel plate. In this way the brightened image is inverted, so that the image tube can be conveniently mated to additional optical systems that also invert an image. Hence the output image will not be inverted. By changing the focus of the electron beam the image can also be magnified or reduced in size.





ANODES OF DIVERSE DESIGN register the position of a pulse of electrons emerging from a microchannel plate so that image data can be stored and manipulated. Coarse resolution can be obtained by detecting the position of the pulses on an array of separate anodes (*a*). For higher resolution the number of anodes required makes the circuitry prohibitively complex, and other designs are employed. Some of them measure the position of a pulse by depositing charge on a pair of wires in a horizontal and a vertical grid. Each wire in the grid can have its own detection circuit (*b*), or the position of the pulse can be encoded as a binary number (*c*). The latter configuration requires fewer detection circuits for a given image resolution. Other designs detect a cloud of charge on various portions of the anode and determine the position of the pulse by comparing the charge in one circuit with that in another. Charge ratios must be calculated in both the horizontal and the vertical directions so that the spatial coordinates of the pulse can be determined. The charge can be measured at both ends of two capacitor strings (*d*), on the four corners of an electrically resistive plate (*e*), on the four quadrants of a disk (*f*) or on more complex geometrical arrangements of electrically insulated anodes (*g*). In each case the ratios of the charge detected by the various circuits can frequently give the position of a pulse on the microchannel plate to within a few thousandths of the diameter of the field of view.

still more efficient encoding of the positional information. In particular, if the pulse can trigger any combination of wires in the grids, every possible combination of wires can code for some position of the pulse. With such a configuration the number of circuits varies only as the logarithm of the number of pixels.

Another kind of position-sensing anode encodes the coordinates of the electron pulse in an analogue, or continuous, manner. I shall describe one recently developed arrangement that employs only three detecting circuits. The anode is made up of three metal electrodes placed well away from the output face of the microchannel wafer, so that a portion of the electron cloud associated with every electron cascade is intercepted by each of the electrodes. One of the electrodes has a zigzag shape and divides the plane of the anode into two parts [*see illustration above*]. On one side of the zigzag the area of the charge-collecting electrode varies linearly with the *x* coordinate and is independent of the *y* coordinate. The area of the electrode on the other side of the zigzag varies linearly with the *y* coordinate and is independent of the *x* coordinate. By amplifying and measuring the charge pulses on all three electrodes independently and then calculating the ratios of the pulses received by the *x*-dependent and *y*-dependent electrodes to the sum of all three pulses, the position of the output pulse can be accurately determined. The design was originally suggested by Hal O. Anger of the University of California at Berkeley.

My colleagues and I in the Space Sciences Laboratory at Berkeley have adopted Anger's scheme for two spaceflight applications. Four other groups are also seeking to incorporate the detector into photomultipliers that count photons at visible wavelengths. Such a position-sensitive detector can function as the input device for a computer system that can then record and analyze the data. Hence the microchannel plate can serve as an intermediary between incoming radiation and outgoing data of almost limitless variety. Systems of this kind may lead to significant advances in physics and astronomy.

Inducible Repair of DNA

Damage to the hereditary material switches on the "SOS response," in which several enzyme systems function in concert. Two proteins, RecA and LexA, play multiple roles in repairing genetic damage

by Paul Howard-Flanders

The life of every organism and its continuity from generation to generation depend on the long-term stability of hereditary information encoded in the double helix of DNA. That stability is not the result of invulnerability to damage: DNA, whether it is in a bacterial cell or in the millions of cells of a human being, is quite sensitive to damage by harmful radiations and chemical agents in the environment. Its stability and the precision of its replication are preserved by enzymes that continually repair genetic lesions. In particular, significant damage to DNA molecules can induce an emergency response in which increased quantities of the repair enzymes are synthesized and are caused to function as an orchestrated ensemble. Various DNA-repair systems have been observed and studied for many years, but it is only in the past two years that the nature of certain of the induced responses and the remarkable regulatory systems that control them have begun to be understood in detail.

Both DNA's vulnerability to damage and its susceptibility to repair are implicit in its architecture. Each strand of the double helix is a chain of sugar molecules and phosphate groups. From this backbone there project a series of the chemical groups called bases. The actual building blocks of DNA are nucleotides, each one consisting of a base, a sugar and a phosphate group. The bases are of four types: two purines, adenine (*A*) and guanine (*G*), and two pyrimidines, cytosine (*C*) and thymine (*T*). The bases on one strand are linked by hydrogen bonds to the bases on the other strand to form double-strand DNA. The bases are complementary: *A* pairs only with *T*, and *G* pairs only with *C*. As a result the two strands are complementary and each is in effect a template for the other: where one strand has the sequence *TGACCGCTTA*, the other strand must have *ACTGGCGAAT*. Complementarity is the basis of DNA's replication, its expression in the form of protein and its recombination. In replication each strand serves as a template along which the daughter strand is laid down. In expression the hereditary information encoded in the sequence of bases along one strand of a gene is transcribed into the complementary sequence on a strand of RNA and then is translated into protein; each three-base "codon" specifies one of the amino acids that constitute the protein chain. In genetic recombination strands of DNA are cut and homologous regions are joined to make new combinations of genes.

Clearly the incorporation into DNA of an incorrect or altered base, or the presence of any lesion that distorts the double helix or prevents accurate base pairing, will tend to interfere with replication, protein synthesis and recombination. Unless damage to DNA is successfully repaired it can cause the death of a cell; in the cells of a mammal such damage seems sometimes to initiate a process that leads to cancer.

Although damage to DNA can take many forms, the lesion whose repair has been most studied is the pyrimidine dimer, a product of exposure to ultraviolet radiation. A dimer is formed when two adjacent pyrimidine bases (either thymines or cytosines) on a strand are linked together by a four-carbon ring. The two bases of a dimer are pulled out of alignment, so that their hydrogen bonds to the complementary bases are broken and the DNA backbone is distorted, which in turn prevents the correct pairing of two bases on each side of the dimer. The presence of a single pyrimidine dimer is enough to interrupt both transcription and replication.

Basic knowledge of DNA repair processes has come primarily from genetic experiments with bacterial cells exposed to ultraviolet radiation and from biochemical studies of the course of DNA replication and repair and of the enzymes that participate in those processes. In this article I shall trace the story of some of the discoveries chronologically, in the hope of reminding the reader that complex biological systems are not often elucidated all at once or even in logical progression. Rather it is usually the case that numerous observations and bits of data accumulate, often without obvious relation or immediate significance; they acquire significance from one another, and gradually the pieces fall together and begin to make sense.

By the late 1950's it was known that whereas the common intestinal bacterium *Escherichia coli* generally survives moderate exposure to ultraviolet radiation, a mutant strain of the bacterium is particularly sensitive, or vulnerable, to such radiation. Why is that so? One possibility was that the normal bacteria had an enzyme or a system of enzymes able to repair ultraviolet damage to DNA and so increase the rate of survival, whereas the sensitive bacteria lacked that ability. Enzymes, like other proteins, are encoded by genes, and in 1961 my colleagues and I at the Yale University School of Medicine undertook genetic experiments to find genes that might specify repair enzymes. We took advantage of the fact that *E. coli* cells can exchange genes by conjugation, a primitive form of mating between male and female cells. We worked with the standard laboratory strain *E. coli* K12 because it was known to be relatively resistant to radiation and because its genetic map was already well known, that is, the location of many of the genes along the circular double-strand DNA molecule constituting the bacterial chromosome had been determined.

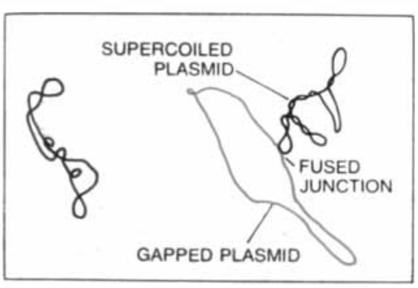
On isolating *E. coli* mutants that were sensitive to ultraviolet radiation, to X rays or to both, we found their sensitivity to be a stable, heritable property that could be transmitted between mating cells. By standard genetic methods we were able to identify the mutant genes and to map their positions on the *E. coli* chromosome. One class of ultraviolet-sensitive mutants was shown to have mutations in genes we designated *uvrA*, *uvrB* and *uvrC*; other mutants, sensitive both to ultraviolet and to X rays, carried mutations in a gene we called *lexA*. The *uvr* genes, it turned out, code for repair enzymes. The *lexA* gene codes for a different kind of protein, but that was not to become clear until much later.

What biochemical mechanism is mediated by the *uvr* genes? Richard B. Setlow and William L. Carrier of the Oak Ridge National Laboratory and Richard P. Boyce, Lee Theriot and I at Yale studied the fate of pyrimidine dimers in the DNA of normal *E. coli* and ultraviolet-sensitive mutants that were *uvr*⁻, or deficient in one or another of the *uvr* genes. To do so we exposed the cells to ultraviolet radiation and then examined their DNA for pyrimidine dimers. We found that whereas the dimers remained in the *uvr*⁻ cells' intact chromosomal DNA, as we expected, they were not present in the chromosomal DNA of the resistant bacteria. The dimers had not disappeared from the resistant cells, however. They could be found in short single strands of DNA only a few base pairs long. What we had discovered was "excision repair." The single strands of DNA incorporating dimers had been excised, presumably by an enzyme or enzymes encoded by one or more of the *uvr* genes. We reasoned that the repair of DNA might then be completed by the enzymes DNA polymerase and DNA ligase, which respectively would replace the excised nucleotides and ligate the sugar-phosphate backbone; the intact complementary strand would serve as a template. Soon direct evidence of just such repair synthesis was reported by David E. Pettijohn and Philip C. Hanawalt of Stanford University.

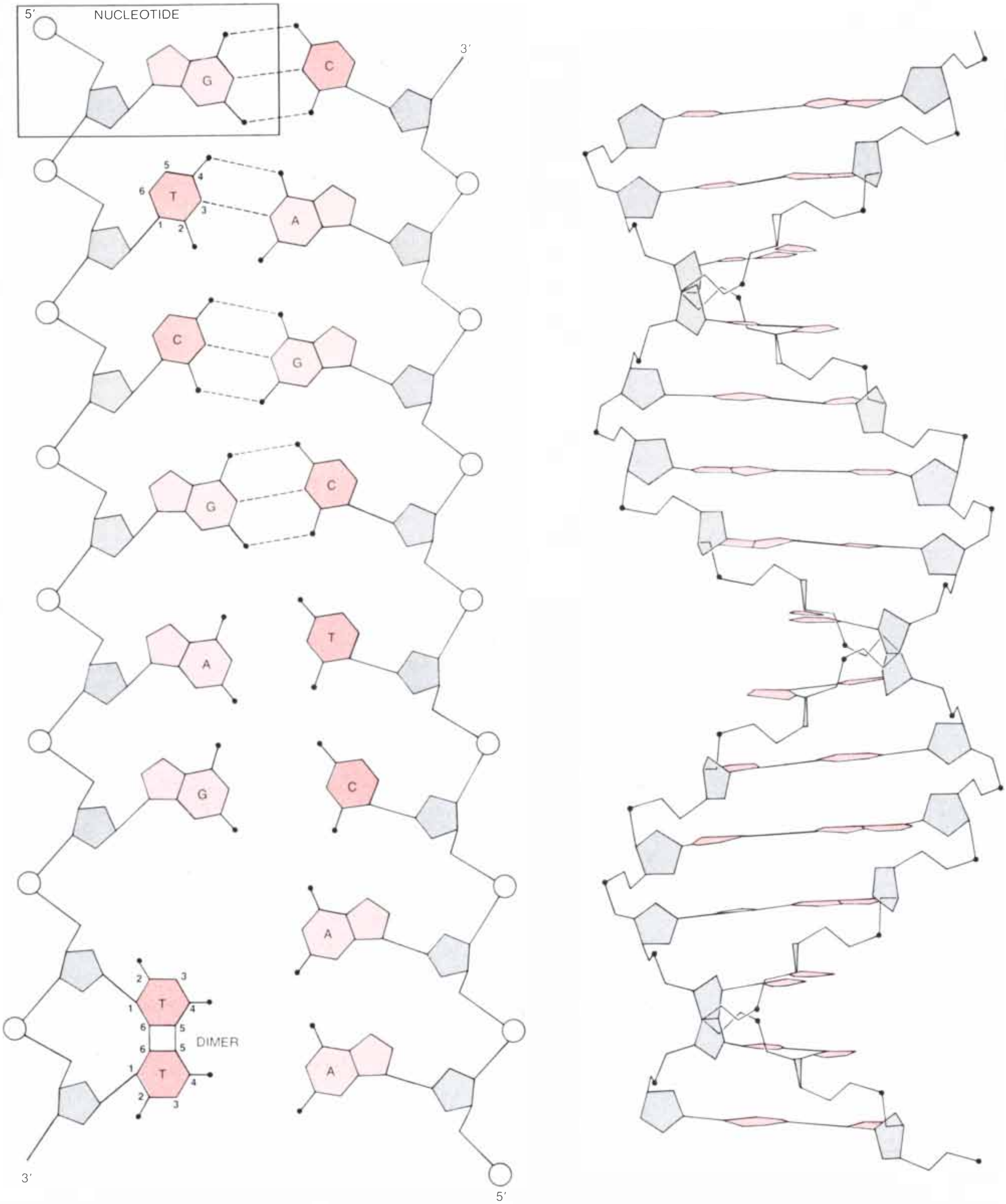
Excision repair has been found to be one of the fundamental DNA-repair processes, not only in bacteria but also in mammalian (including human) cells, and it is effective in eliminating not only pyrimidine dimers but also other kinds of damage to DNA. Boyce and I quickly found, for example, that it repairs damage caused by mitomycin *C*, which acts primarily by cross-linking the two strands of the double helix. The fact that mitomycin *C* is an antitumor drug (because it damages DNA) and also is carcinogenic in animal cells (for the same reason) was a reminder that knowledge gained from studies of DNA repair in bacteria might be broadly applicable.

The discovery of a second major DNA-repair process, very different from excision repair, came in a roundabout way. In 1965 A. John Clark and Anne D. Margulies of the University of California at Berkeley were studying genetic recombination, the process in which genes on different chromosomes become linked, as for example in the course of bacterial mating. Clark isolated *E. coli* mutants whose DNA failed to recombine as expected after mating, and he traced the deficiency in recombination to a mutation in a particular gene, which he called *recA*. When he tested the *recA*⁻ cells for resistance to radiation, he found to his surprise that they were extremely sensitive to damage by ultraviolet radiation and X rays.

W. Dean Rupp and I were puzzled by this apparent relation between a deficiency in recombination and sensitivity to radiation. Perhaps it meant that the protein product of *recA* mediated some repair process as yet unknown. We compared the ultraviolet sensitivity of cells that were deficient in either *uvrA* or *recA* with the sensitivity of double mutants, which were deficient in both genes. The double mutants were some 50 times more sensitive than either the *uvrA*⁻ or the *recA*⁻ cells, a clear indication that two quite different repair processes



RECOMBINATION mediated by the RecA protein is the basis of one mechanism of DNA repair. A step in the recombination process is demonstrated in an electron micrograph made by Era Cassuto and Stephen C. West in the author's laboratory. As is indicated in the map (*left*) of the micrograph, two copies of a DNA plasmid (a small, circular double-strand molecule) are shown pairing at a "fused junction," an early event in recombination. One plasmid is an intact circular molecule of double-strand DNA, as is shown by its twisted "supercoiled" form. It is paired with a plasmid that is identical except for the fact that a short segment of one strand of its DNA has been removed by enzymes, leaving a gap (which "relaxes" the plasmid). Such fused pairs were found only when intact plasmids and similar ones with a single-strand gap were incubated with RecA and ATP (the cell's major energy source); pairs always included at least one relaxed molecule. Presumably RecA bound to the single-strand region opposite the gap and aligned it with a homologous region on one strand of the intact plasmid.



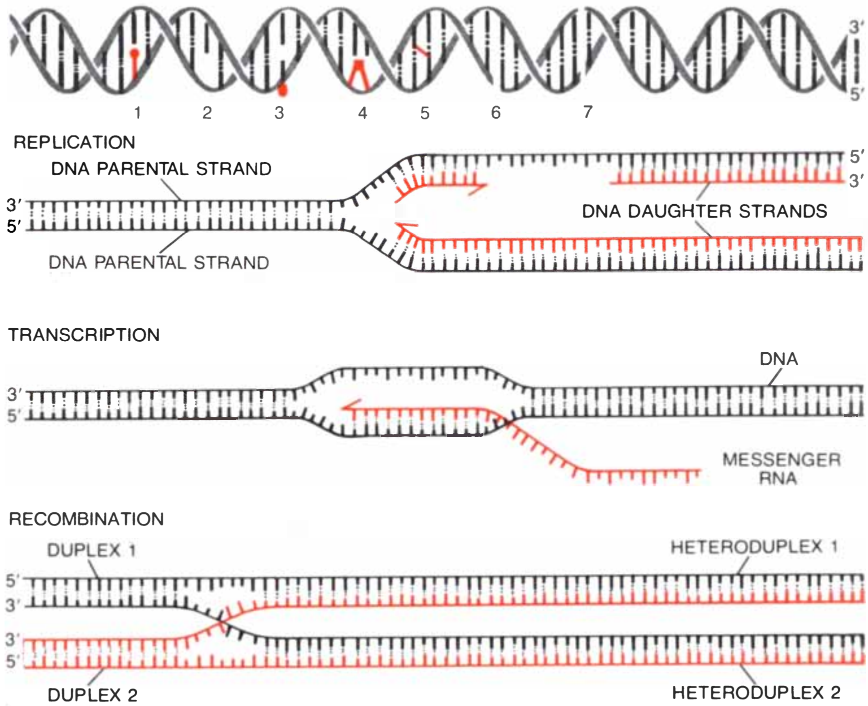
DNA MOLECULE (*left*) **consists of two strands of nucleotides, each comprising a sugar** (*gray*), **a phosphate group** (*open circle*) **and one of four bases** (*color*): **the purines adenine** (*A*) **and guanine** (*G*) **and the pyrimidines thymine** (*T*) **and cytosine** (*C*). **The bases project from the sugar-phosphate backbone, and the two strands are linked by hydrogen bonds** (*broken lines*) **between complementary bases:** ***A*** **can pair only with** ***T***, **and** ***G*** **can pair only with** ***C***. **Genetic information is encoded in the sequence of the bases along one strand. The two strands have opposite polarity; the base sequence is by convention read in the 5′-to-3′ direction. Here the double helix is shown untwisted; the plane of the bases, which is actually perpendicular to the axis of the helix, is rotated 90 degrees in order to display the chemical structure. Pyrimidine dimers, the lesions that are most studied in investigations of DNA repair, are caused by exposure to ultraviolet radiation. A dimer** (*bottom*) **is formed when the double bonds between the 5′ and 6′ carbon atoms of two adjacent pyrimidines open and form a four-carbon ring, pulling the two bases together and thus disrupting their hydrogen bonding to the complementary bases on the opposite strand; the backbone is distorted and correct hydrogen bonding is impossible also for two bases on each side of the dimer. Dimers are most frequently formed between adjacent thymines on a strand, but they can also be formed between two cytosines or between a thymine and a cytosine. A skeletal model of DNA** (*right*) **shows how the two sugar-phosphate chains are twisted into a double-helical structure, but here the details of the paired bases are obscured because the bases are almost perpendicular to the plane of the page.**

were implicated. To simplify matters we studied the synthesis of DNA in *uvrA*⁻ bacteria, which could not conduct excision repair but presumably retained whatever system was based on the *recA* gene product. In 1968 we discovered postreplication repair, which operates only after the replication of the ultraviolet-damaged DNA.

In a cell exposed to ultraviolet radiation the DNA is replicated normally along undamaged stretches of the parental template strands. When the replication fork (the point at which the two parental strands are unwound and exposed to the replicating enzymes) reaches an unexcised pyrimidine dimer, correct base pairing is impossible along the damaged strand. The synthesis of one new strand is interrupted, to be reinitiated at some point in the undamaged region of the template beyond the dimer. As a result there are "postreplication gaps" in the daughter strands. Such gaps cannot be handled by excision repair, which requires an intact complementary strand; they must be repaired by a different process. We reasoned that the base sequence is effectively lost from both strands of one of the two newly formed sister duplexes: one strand has a gap and the other has a dimer. The missing information must somehow be recovered if the gap is to be repaired correctly. We realized that the required base sequence is available nearby, in the sister duplex (double strand) formed at the same replication fork. Might a recombination event substitute one strand of the sister duplex for the missing bases in the postreplication gap, thereby supplying a template with the appropriate sequence of bases opposite the dimer?

In an effort to detect such recombinational repair Rupp and I labeled the DNA with isotopes that made it possible to distinguish parental and daughter strands from each other on the basis of their density. We were able to count the number of postreplication gaps and the number of sister-strand exchanges. Sure enough, there were about as many sister exchanges as there were gaps. Confirmation of the postreplication recombinational repair came from the finding by Ann K. Ganesan of Stanford that pyrimidine dimers do not remain exclusively in ultraviolet-irradiated strands but are distributed among new strands as well as old ones as the DNA replicates.

Postreplication repair complements excision repair nicely. The gaps left in daughter strands by dimers (or other forms of damage) that reach the replication fork before being excised are filled by recombinational repair; the dimer (or other damage) left in one parental strand is handled by excision repair.



DAMAGE TO DNA can take many forms. As is indicated along the double helix (*top*), a base may be altered chemically by alkylation or hydroxylation or in other ways (*1*), or it may be deleted (*2*). A chemical adduct may be formed by the addition of a bulky molecule (*3*). Adjacent pyrimidines may be linked in a dimer (*4*), the two strands may be chemically cross-linked (*5*) or one strand may be linked to a protein. One strand (*6*) or both strands (*7*) may be broken. Most such lesions disrupt base pairing or the continuity of the template and thereby interfere with correct functioning of DNA. Replication and transcription depend on base pairing for the synthesis of two new daughter strands of DNA or a strand of messenger RNA; most recombination processes depend on base pairing to bring homologous stretches of DNA into alignment.

A fuller understanding of the role of *recA* in postreplication repair emerged from studies of a seemingly unrelated phenomenon: lysogenic induction. A bacterium is said to be lysogenic when the DNA of a bacteriophage (bacterial virus) that infects it becomes integrated into the bacterial chromosome as a prophage instead of multiplying. (Such multiplication leads to the formation of new phage particles and kills the cell.) One such "temperate" virus, which infects lysogenic strains of *E. coli,* is phage lambda. Its DNA can remain dormant in the bacterium, replicating with the bacterial chromosome, for many generations. When the bacterium is exposed to ultraviolet radiation or some other DNA-damaging agent, the prophage is "induced." Its DNA loops out of the chromosome, replicates and directs the synthesis of viral proteins; hundreds of new virus particles are produced and burst out of the cell. In 1967 Clark, who had discovered *recA,* and Israel Hertman and Salvador E. Luria of the Massachusetts Institute of Technology found that *recA*⁻ lysogenic cells are not induced, that is, lambda remains a dormant prophage even when the cell is exposed to ultraviolet. The *recA* gene, in other words, was shown to be required for lysogenic induction.

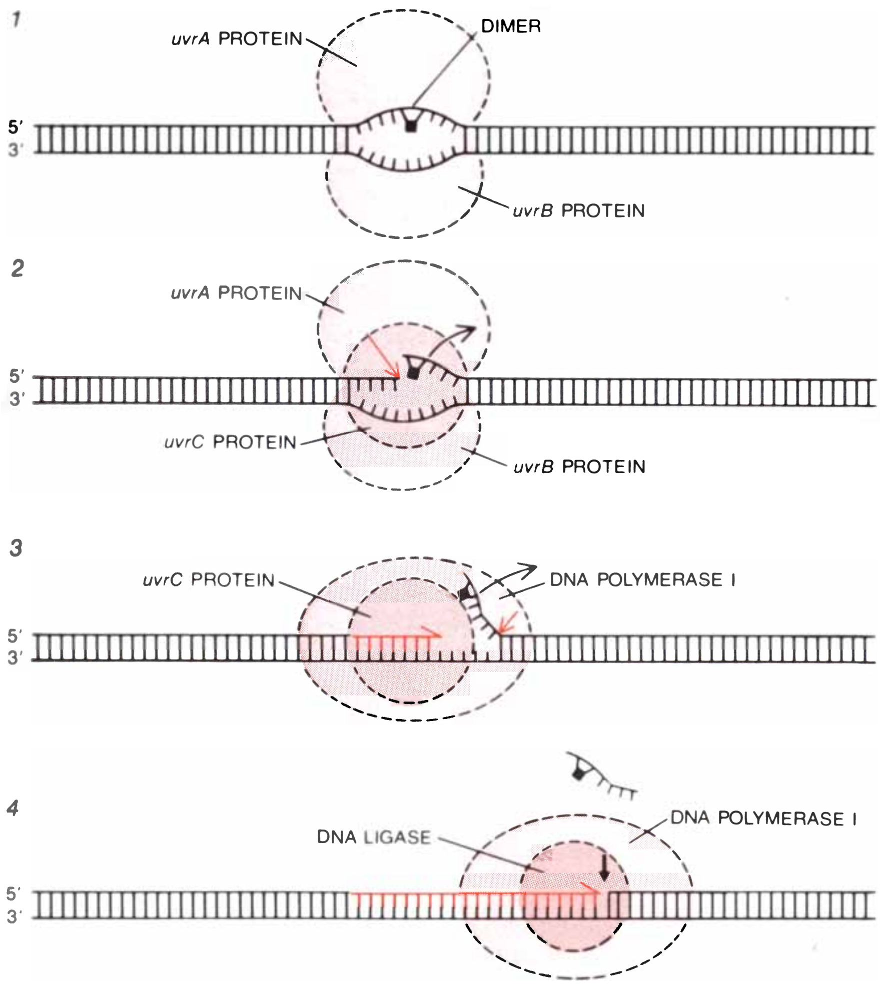
The *recA* gene is implicated in a great deal more than that. By the mid-1970's it was known that the same events within an irradiated cell that precipitate lysogenic induction switch on the synthesis of DNA-repair enzymes; they also interrupt DNA replication and cell division, so that the cell becomes elongated instead of partitioning to form two daughter cells, and they increase the rate of mutagenesis and promote the synthesis of large amounts of a particular protein, which for a time was known as protein *X.* These responses (and several others I have not listed) all appear to be reactions to distress; Miroslav Radman of the Free University of Brussels gave them the collective name "*SOS* response."

In 1975 Jacqueline George and Marc Castellazzi of the Institute for Research in Molecular Biology in Paris found that a particular *recA* mutation acts as a switch that turns on the *SOS* response even in the absence of ultraviolet irradiation. They worked with a strain of lysogenic *E. coli* carrying a temperature-sensitive mutation, *recA441,* that was known to be lethal to the bacteria at 42 degrees Celsius. They found that cell death could be prevented by a second mutation in either of two other genes: *sulA* or *sulB.* Now that *recA441* cells could be kept alive at 42 degrees, George and Castellazzi showed that even in the absence of ultraviolet radiation all the *SOS* functions were switched on at that temperature. If a mutation in *recA* could switch on the *SOS* response, the response must depend (at least in part) on the *recA* gene.

At about the same time Jeffrey W. Roberts of Cornell University was investigating the mechanism by which ultraviolet radiation brings about prophage induction. It was known that the lambda prophage is kept in its dormant state by a regulatory protein called a repressor, which keeps all the phage genes (except for the one that specifies the repressor itself) turned off. The repressor binds to "operator" regions on the phage DNA and prevents the binding of the RNA polymerase that would otherwise transcribe the phage genes into RNA. Seeking the mechanism whereby ultraviolet irradiation releases the lambda repressor, Roberts and his co-workers found that in the course of induction the repressor seemed to disappear. It turned out that the repressor is released from the prophage DNA because it has been cut in two. The Cornell workers identified and purified the enzyme responsible for the cleavage: a protease (protein-cleaving enzyme) with a molecular weight of 38,000. The protease was detected in increased amounts after irradiation in wild-type cells but not in *recA*⁻ mutants. It therefore might be the product of the *recA* gene. (Increased amounts of the protease were not detected in *lexA*⁻ cells either, for reasons that will be made clear below.)

Roberts and his colleagues found that the protease they were able to purify from *recA441* mutants (the one in which the *SOS* functions are always turned on at a temperature of 42 degrees) was more active than the protease from wild-type cells; on the other hand, a mutant in which prophage induction does not take place even after irradiation showed no repressor-cleaving activity. Further analysis showed that the protease is an unusual one. It is very slow-acting, its specific targets are repressors (such as lambda repressor) and its activity is dependent on the presence not only of ATP (the primary source of biochemical energy) but also of single-strand DNA. Although Roberts suspected his protease was the *recA* gene product, it was hard to prove that it was not just a minor component of his preparation until a higher degree of purification could be attained.

The final identification of the *recA* gene product and its purification in quantity represented an important advance, which was achieved by Roberts and his collaborators and by a number of other investigators: Peter T. Emmerson and Stephen C. West at the University of Newcastle upon Tyne, Lorraine J. Gudas and David W. Mount at the University of Arizona, John W. Little and D. G. Kleid at the Stanford Research Institute, Kevin McEntee at the University of Chicago and Hideyuki Ogawa and Tomoko Ogawa at the University of Osaka. By 1977 it was established that protein *X*, which had been observed in irradiated cells in 1971 by Masayori Inouye and Arthur B. Pardee at Princeton University, and which Gudas had found to be missing or altered in *lexA*⁻ and in *recA*⁻ mutants, was in fact the RecA protein. Now the way was opened to biochemical studies of the protein and of the mechanism of post-replication repair.



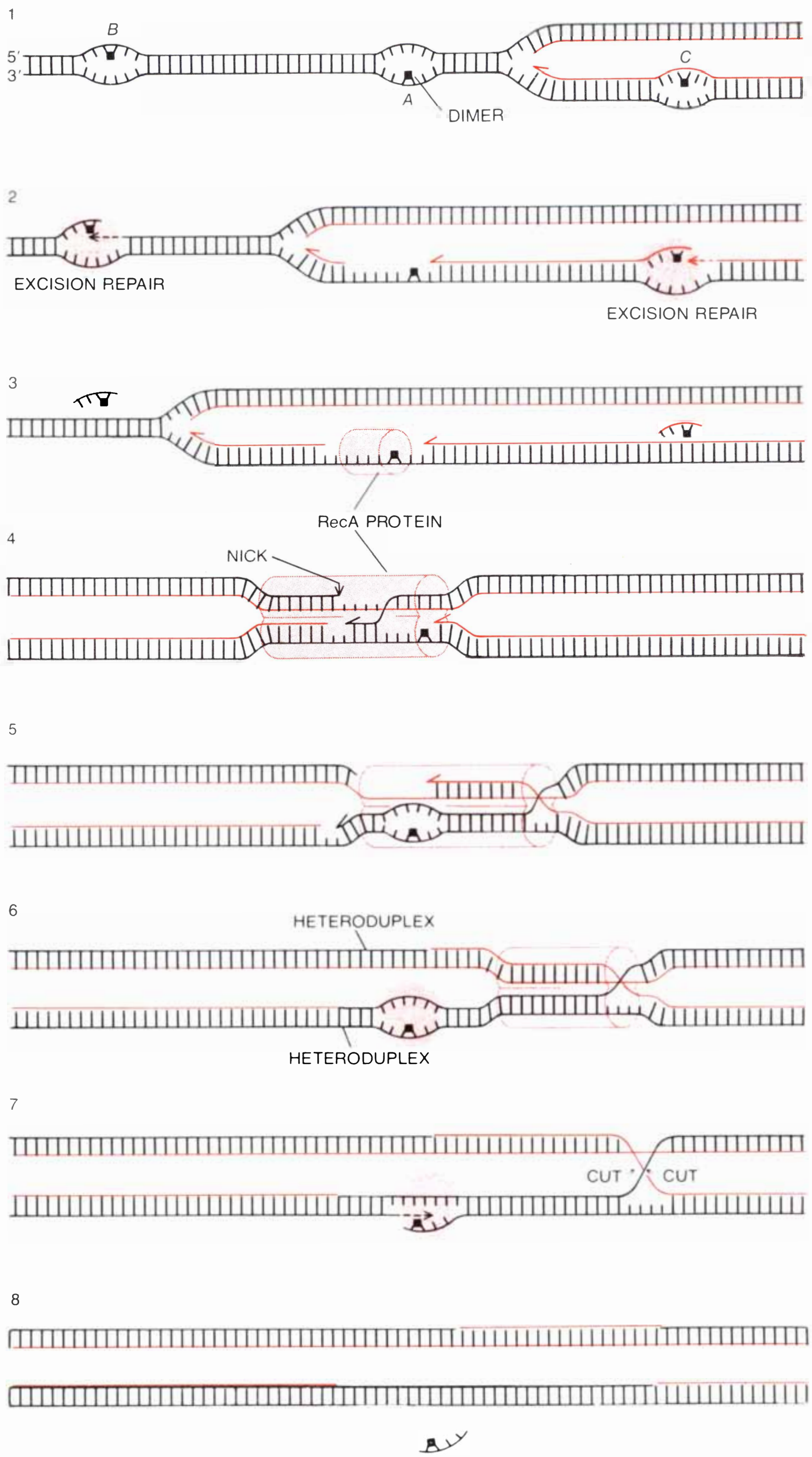
EXCISION REPAIR removes pyrimidine dimers and some other base lesions from DNA in a cut-and-patch process. Three special enzymes encoded by the genes *uvrA*, *uvrB* and *uvrC* take part along with the enzymes DNA polymerase I and DNA ligase. First the *uvrA* and *uvrB* proteins bind to the damaged site (*1*). Perhaps with the help of the *uvrC* protein they make a nick (*colored arrow*) at the 5′ end of the damaged region (*2*). In the presence of the *uvrC* protein, DNA polymerase I binds at the nick and adds nucleotides to the 3′ end according to the base-pairing rules. The polymerase makes a second nick in the chain to release the damaged region (*3*); it may continue to digest nucleotides and replace them one at a time, translating the nick to the right (*4*). Finally, nick (*heavy arrow*) is closed by DNA ligase, completing repair.

Meanwhile the other half of what was coming to be known as the "*rec-lex* phenomenon" was under investigation. In 1973 Mount at Arizona and K. Brooks Low at Yale examined a partial diploid strain of *E. coli:* cells carrying a short extra segment of the chromosomal DNA, so that two versions of some of the genes on the chromosome are present. They observed that *lexA*⁻ mutants are genetically dominant, that is, a cell carrying both the wild-type *lexA* gene and a mutant *lexA*⁻ gene displays the characteristics of the mutant cell, including sensitivity to X-ray damage. Now dominance in a mutant gene is often a sign that the gene has to do with genetic regulation. For example, certain mutations in the *E. coli* gene *lacI,* which specifies the repressor controlling expression of the genes involved in lactose metabolism, are dominant. It seemed possible by analogy that *lexA* might encode a repressor protein that regulates certain DNA-repair enzymes.

Mount reasoned that *lexA*⁻ mutants might be sensitive to radiation because they make a faulty repressor: one that is not released following irradiation. (That would explain the dominance of *lexA*⁻. Even if normal repressor that can be released were present in a cell, the faulty *lexA*⁻ protein would continue to bind and repress the repair enzymes.) Working with a strain carrying a particular *lexA*⁻ mutation called *lexA3,* Mount looked for cells that had reverted to ultraviolet resistance. He found one that still carried the original mutation and in addition carried a new mutation designated *lexA51*, which mapped within or near to the *lexA* site. Apparently this mutation affected *lexA* function in such a way that the genes formerly repressed by the *lexA* protein were released, repair enzymes were synthesized and ultraviolet damage was repaired. It developed that *lexA51* mutants make repressor that fails to bind correctly to DNA. As a result not only repair-enzyme synthesis but also all the other *SOS* functions, including the enhanced synthesis of RecA, are switched on all the time, even in the absence of ultraviolet radiation or other DNA-damaging agents. And so the *SOS* response was shown to depend not only on *recA* but also on *lexA*, and it seemed likely that the product of the latter was a repressor protein controlling all the *SOS*-response genes.

The next step was to identify the *lexA* protein. Little (who was by then at Arizona) and Mount did so by means of recombinant-DNA techniques. They isolated the *lexA* gene, spliced it into a plasmid (a small, circular piece of nonchromosomal DNA that replicates autonomously in the bacterial cell), inserted the recombinant plasmid into *E. coli* and were able to identify and purify LexA: a protein with a molecular weight of 24,000. In view of LexA's possible role as a repressor, they tested it for susceptibility to cleavage by the *recA* protease and made the important discovery that RecA protein cleaves LexA. It does so faster than it cleaves lambda repressor, suggesting that LexA is the protease's primary substrate. Mark Ptashne and Roger Brent of Harvard University went on to show that the *lexA* gene is itself repressed by LexA protein: the gene is repressed by its own product, as in the case of the lambda repressor gene and some other genes.

If LexA is a repressor, what genes does it repress in addition to its own? Cynthia J. Kenyon and Graham C. Walker of M.I.T. devised an ingenious experiment to learn which *E. coli* genes are switched on in response to damage to their DNA. Working with a strain from which all the genes governing the metabolism of lactose had been deleted, they applied a method devised by Malcolm J. Casadaban and Stanley N. Cohen of the Stanford University School of Medicine to insert the *lacZ* gene at random positions in the chromosome. Then they looked for cells that were able to metabolize lactose only if a DNA-damaging agent (such as mitomycin *C*) was present. In such a cell, they could assume, some gene that was normally repressed had been switched on in response to DNA damage. The gene's normal function had been disturbed by the *lacZ* insertion, however, and so it was *lacZ* that was expressed. When the genes thus switched on were identified (by their failure to perform their normal function), they turned out to include the



POSTREPLICATION REPAIR deals with a pyrimidine dimer (*A*) that interferes with replication, during which the parental strands (*black*) unwind at the replication fork and two daughter strands (*color*) are synthesized (*1*); other dimers (*B, C*) are handled by excision repair. Dimer *A* prevents base pairing along a stretch of one parental strand, producing a postreplication gap opposite a stretch of single-strand DNA (*2*). RecA protein (*light color*) binds to the single-strand region (*3*) and aligns it with a homologous region of the sister duplex; when homologous pairing is achieved, an enzyme nicks the duplex (*4*). RecA protein switches the free end of the duplex's parental strand into the gap, producing a crossed-strand exchange (*5*). The upper heteroduplex can now be repaired by DNA polymerase. With the correct sequence in place opposite dimer *A* and with RecA released, dimer *A* is dealt with by excision-repair enzymes (*6*). Finally, two cuts are made by an enzyme at the site of the crossed-strand exchange (*7*), resolving the recombination process and producing two intact heteroduplex molecules (*8*).

repair-enzyme genes *uvrA* and *uvrB* and also *umuC,* which is implicated in the increased mutagenesis that follows ultraviolet irradiation.

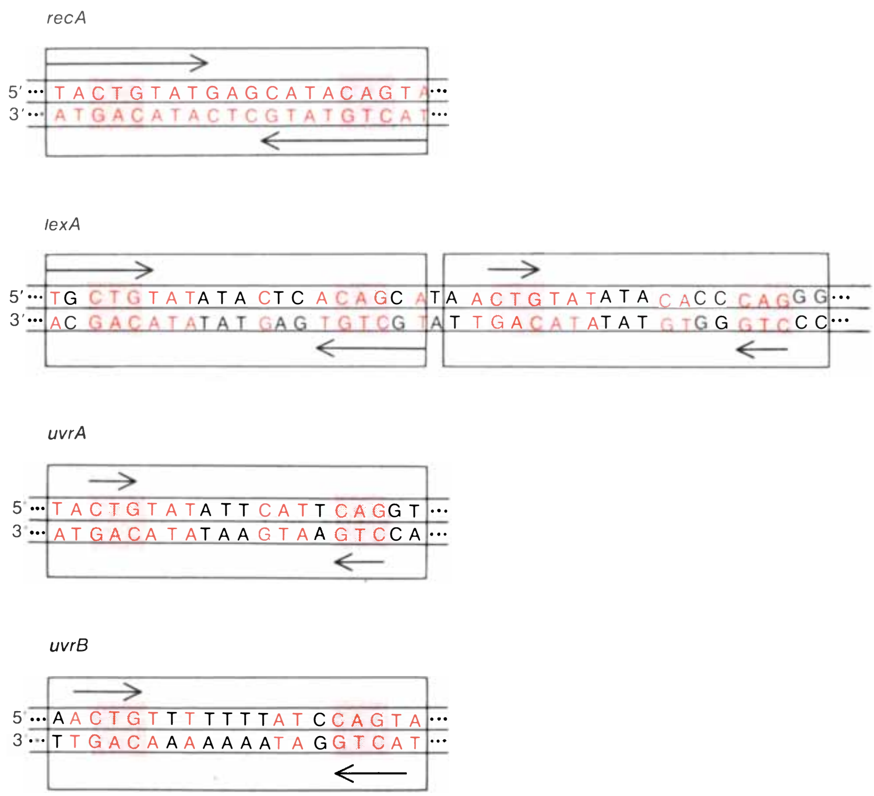
The M.I.T. experiment had identified certain genes that apparently were repressed by the LexA protein. Barry M. Kacinski, Aziz Sancar and Rupp sought direct evidence of such regulation. They found that Mount's *lexA51* mutant, in which the *SOS* functions are always switched on because there is no repression by LexA, displays far more *uvrA* and *uvrB* activity in incising damaged DNA than normal cells do. The increased activity enhances DNA repair: Jeffrey Auerbach in my laboratory detected a higher level of cell survival, which could be traced to the greater *uvrA* and *uvrB* activity, in *lexA51* mutants than in wild-type cells.

If LexA protein is indeed a repressor of a group of genes involved in DNA repair, it must bind to operator regions near the beginning of each of those genes, and one would expect to find similar nucleotide sequences in those regions. By now the sequence of the *recA* gene has been determined by Toshihiro Horii and the Ogawas at Osaka and by Sancar and Rupp; the sequence of the operator region of *lexA* has been worked out by the Osaka group and by Little and Mount, and parts of the *uvrA* and *uvrB* genes have been sequenced by Sancar and Rupp. A very similar sequence about 20 nucleotides long is indeed found in the operator regions of each gene, and this "*SOS* box" has (to varying degrees in different genes) the inverted-repeat form that is characteristic of repressor-binding sites. In the case of the *lexA* gene itself there are actually two nearly identical, adjacent sequences, which are analogous to the multiple binding sites in the operator of phage lambda. Recently the LexA protein has been purified by Little and Mount and by Brent and Ptashne and has been shown to bind specifically to the *SOS* boxes of *recA* and *lexA.*

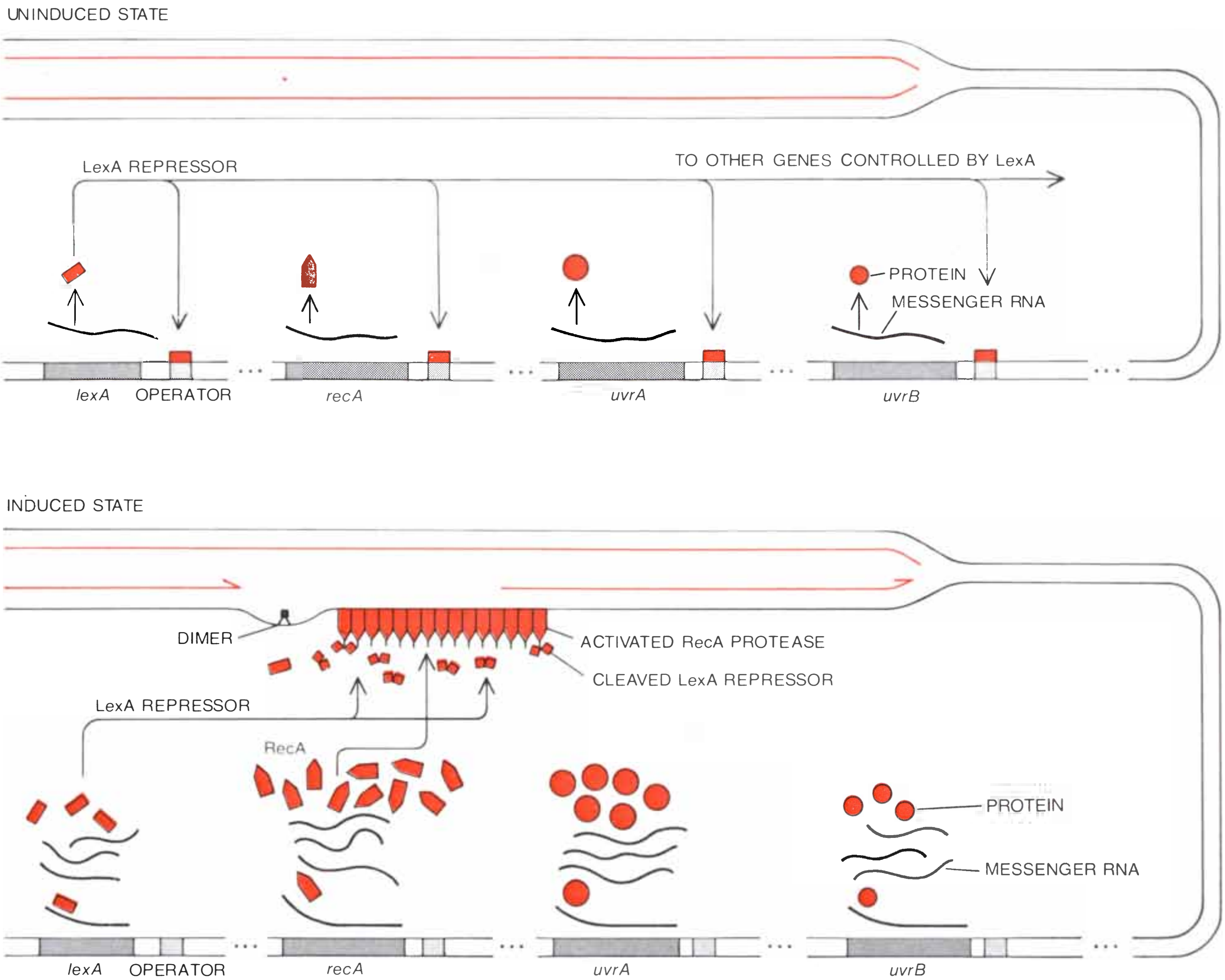
The interrelated events of the *SOS* response and their regulation can now be described in broad outline. In an undamaged, growing cell the DNA-repair genes are almost completely repressed by LexA; enough of the repressor is synthesized, however, to occupy the operator regions, and enough of the *uvr* proteins is synthesized to handle sporadic excision repairs. When the cells are exposed to a significant dose of ultraviolet radiation, pyrimidine dimers that cannot be handled by the few excision-enzyme molecules present in the uninduced state give rise to postreplication gaps. Any RecA protein already present in the cell binds to the single-strand DNA opposite the gaps; the DNA stimulates the RecA's protein-cleaving activity and the *SOS* response is induced. The RecA protein cleaves LexA repressors, releasing *recA* and the other regulated genes from repression; large amounts of RecA protein are synthesized, and the additional protein binds to single-strand DNA that is still free and then to adjacent double-strand regions. Under these conditions the LexA protein is cleaved as soon as it is synthesized and the regulated genes remain switched on—as long as single-strand DNA is present in the cell to stimulate RecA's protease activity. After completion of postreplication repair, with all the DNA now in double-strand form, RecA is no longer activated as a protease. Newly synthesized LexA protein is no longer cleaved and can function as a repressor; the cell returns to the uninduced state.

Clearly the *SOS* response promotes the efficient repair of DNA by increasing the availability of excision-repair enzymes as well as by opening the way to postreplication repair. In excision repair the enzymes specified by *uvrA, uvrB* and *uvrC* act together to recognize the damaged bases and start the excision process. Since cells survive many thousands of such lesions even though the process is slow, large quantities of the enzymes must be required. Rupp and his colleagues have shown that an undamaged cell normally has only 10 or 20 molecules of the enzyme specified by *uvrA,* it is the increased synthesis following *SOS* induction that makes enough enzyme available for the efficient excision of dimers in large numbers. The precise mechanism of action of the three enzymes is now being studied in several laboratories. The genes for the enzymes have been cloned on multiple-copy plasmids to increase the level of synthesis. Moreover, the powerful "maxicell" method developed by Sancar and Rupp makes it possible to identify and purify the enzymes even though convenient functional assays for the proteins are not available.

RecA, the reader will have noted, has two very different functions: it is both a protease in its regulatory role and a recombination enzyme that manipulates strands of DNA. In recombination, such as occurs in postreplication repair, its first function is to promote homologous pairing between two DNA molecules, that is, to bring regions having a sequence of complementary bases into precise alignment in preparation for an exchange of strands. In 1979 McEntee at Chicago, George M. Weinstock and I. Robert Lehman at the Stanford School of Medicine and Takahiko Shibata, Chanchal DasGupta, Richard P. Cunningham and Charles M. Radding at



OPERATOR REGIONS near the beginning of four genes controlled by the repressor protein LexA have similar base sequences about 20 base pairs long; they are the binding sites for the repressor. (Binding has been demonstrated so far only for the *recA* and the *lexA* sequences.) Here base pairs in the *lexA, uvrA* and *uvrB* operators that are identical with those in the *recA* operator are shown in color, with base pairs that differ shown in black. All the binding sites, called *SOS* boxes, include inverted-repeat sequences (*arrows*), which read the same in the 5'-to-3' direction on both strands. They include the common sequences *CTG* and *CAG* (*colored boxes*) separated by 10 bases. The *lexA* gene has two nearly identical *SOS* boxes rather than one.



REGULATORY SYSTEM based on *lexA* and *recA* is quiescent during normal growth in the absence of damage to DNA (*top*). LexA repressor binds to the operators of *lexA, recA, uvrA, uvrB* and some other genes, keeping the synthesis of messenger RNA and protein at the low level characteristic of uninduced cells. Damage to DNA sufficient to produce a postreplication gap activates the *SOS* response (*bottom*). RecA protein binds to the single-strand DNA opposite a gap; its protein-cleaving activity is thereby activated and LexA repressor is cleaved. In the absence of functional repressor the LexA-controlled genes are switched on and protein is synthesized at an increased rate. (Although more LexA is synthesized, it is cleaved as long as RecA is activated by single-strand DNA.) When DNA repair is completed, RecA no longer cleaves repressor; newly synthesized LexA once again binds to operators and cell returns to uninduced state.

Yale showed that RecA protein can promote homologous pairing between a single DNA strand and a duplex. Radding's group and Era Cassuto and West in my laboratory found that the protein can promote homologous pairing also between two circular duplex molecules if one of the duplexes has a single-strand region equivalent to a postreplication gap, but that such pairing is not promoted between two intact duplexes.

The exact mechanism of the pairing is still to be determined, but it seems likely that the duplex with a single-strand gap and the intact duplex are first brought into nonspecific contact by RecA and then are moved in relation to each other until complementary bases are aligned. Some insight into the process is gained through study of its energy requirements. The energy is apparently supplied by ATP; we find that the breakdown of ATP (to extract its energy) by RecA is stimulated by both the single-strand regions and adjoining duplexes with a single-strand gap but not by intact duplexes. Evidently the protein adopts an active configuration on the duplexes with a single-strand gap. Large amounts of RecA protein (one molecule per five base pairs) are needed for homologous pairing, enough to give the strand with a gap in it a supportive fibrous structure that brings the DNA molecules into contact and moves them past each other until complementary contacts are established. The relative movement is in caterpillar fashion, the gapped DNA elongating and then contracting as RecA protein first binds to it and then is released as ATP is broken down.

Having brought DNA molecules into alignment in a particular region, RecA goes on to promote the actual exchange of strands. Working with duplex molecules that have the appropriate single-strand regions, West has shown that RecA effects a reciprocal exchange between two double-strand regions; the result is two heteroduplexes in which the four strands have changed partners. This comes about by way of a single-strand crossover, which was proposed as an intermediate structure in recombination by Robin Holliday of the National Institute for Medical Research in London in 1965. In the presence of ATP the RecA protein drives the actual exchange of strands as the crossover site advances along the molecules, a process called branch migration. Enzymes other than RecA are then needed to complete the repair: some kind of nuclease for cutting the strands at various stages of the exchange, DNA polymerase to insert the missing bases to fill the gaps, ligase to seal the sugar-phosphate backbone and perhaps still other enzymes.

I have mentioned the enhanced mutagenesis that seems to be an aspect of the *SOS* response. Some years ago Jean J. Weigle of the California Institute of Technology, studying ultraviolet-induced mutagenesis in phage lambda, discovered that both the frequency of mutation and the survival of irradiated phages were increased when they infected host cells that had themselves been irradiated about 30 minutes beforehand. Evelyn M. Witkin of Rutgers University and Jun-ichi Tomizawa, who is now at the National Institute of Arthritis, Metabolism, and Digestive Diseases, found that mutation is not induced by radiation in *recA*⁻ or *lexA*⁻ bacteria.

Witkin has suggested that a special "error-prone repair" is responsible for the increased survival associated with mutagenesis in irradiated wild-type bacteria. She and Radman have proposed that the *SOS* response promotes the appearance of a special DNA polymerase that is more than ordinarily tolerant of mutations in the template DNA and is able to replicate DNA by observing the base-pairing rules less strictly than the usual polymerase. An error-prone polymerase might increase the level of survival at the expense of an increased frequency of mutation. Just recently Stuart M. Linn, David Lackey and Sharon W. Krauss of Berkeley have detected very small amounts of an error-prone DNA polymerase in cells where increased synthesis of RecA has been induced. Mutagenesis in general and ultraviolet mutagenesis in particular involve complex mechanisms that are only now beginning to be understood.

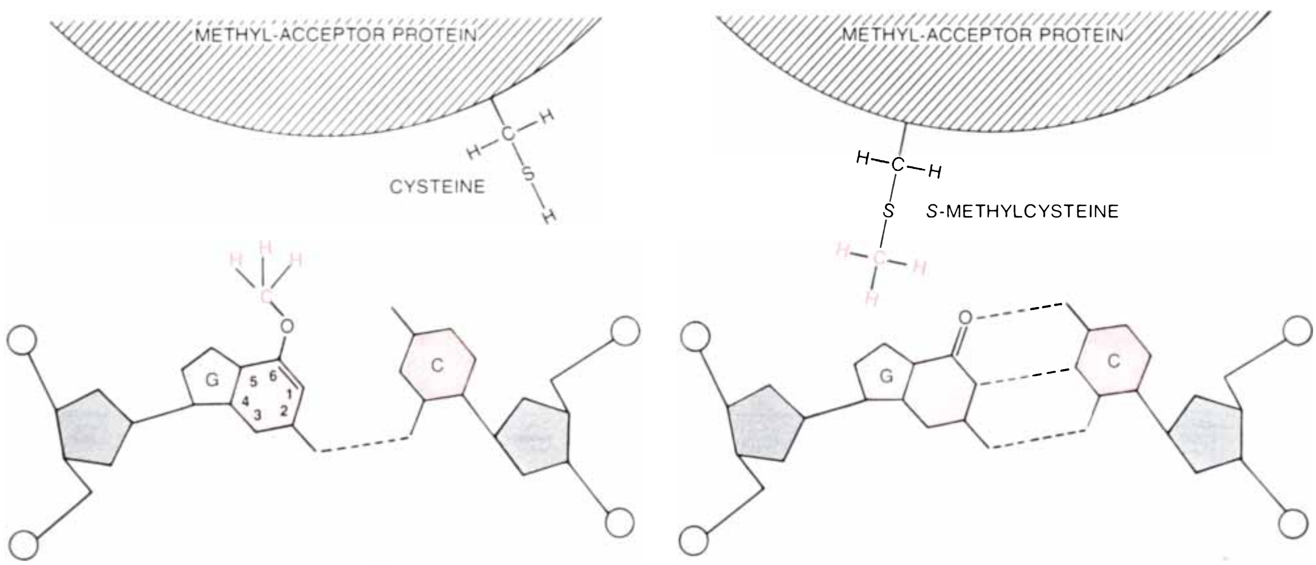
Excision repair and postreplication repair seem to be basic mechanisms for dealing with damage to DNA, but there are more specific remedies for particular lesions. For example, individual incorrect or altered bases are removed by a number of enzymes, called glycosylases, that are highly specific for particular kinds of damaged bases.

One recently discovered repair system is particularly interesting both because of its unusual mechanism and because it seems to represent a second kind of induced response to genetic damage, quite distinct from the *SOS* response. It is called the adaptive response, and it mitigates injury to DNA resulting from the methylation (addition of a CH_3 group) of a guanine base.

The adaptive response was discovered in 1977 by Leona Samson and John Cairns, who were then at the Imperial Cancer Research Fund laboratories in London and were concerned about the possible effects of long exposure to very low levels of potentially carcinogenic chemicals. They investigated the effect on *E. coli* of prolonged exposure to low concentrations of a powerful laboratory mutagen called *N*-methyl-*N'*-nitro-nitrosoguanidine (*MNNG*). To their surprise two hours of incubation in only one microgram of *MNNG* per milliliter of culture medium made the bacteria extremely resistant to a subsequent exposure to the drug at a concentration hundreds of times greater. This adaptive response, which was traced to a gene designated *ada*, lasts for an hour or two. During that time the level of both mutagenesis and cell death resulting from methylation is far lower than the level in cells that have not undergone adaptation. The adapted cells' susceptibility to damage by ultraviolet radiation is not similarly affected, and so the adaptive response cannot be simply an aspect of the *SOS* response.

Thomas Lindahl and his colleagues at the University of Göteborg explored the biochemical mechanism underlying the adaptive response. *MNNG* methylates bases at several sites. The adaptive response appears to be induced when an oxygen atom (designated O^6) on guanine is affected; the inserted methyl group protrudes into what is called the major groove of the DNA double helix. Whereas most altered bases are released or excised from DNA by the appropriate repair enzymes and then replaced, Lindahl found that in this instance the guanine stays in place and the offending methyl group is removed. The job is done by a special protein, synthesized in abundance only during the adaptive response, that acts as a methyl acceptor. A distinctive feature of the process is its finite capacity. When one amino acid unit (a cysteine) of a molecule of the repair protein has taken up one methyl group (to form an *S*-methyl cysteine), the molecule's repair capacity is exhausted.

The full ramifications of the adaptive response remain to be determined. Many aspects of the enzymology of the *SOS* response and the full story of its control by LexA are still being worked out in detail. Further induced-repair systems will surely be discovered. The study of such systems in bacteria is still far from complete, and it is being pursued in many laboratories. At the same time much current research is focused on DNA-repair enzymes and the response to genetic damage in higher organisms, including mammals, motivated in part by increasing evidence that DNA damage and its repair are significant factors in human cancer, and perhaps in other diseases and in the process of aging.



METHYLATION of one site in a guanine by the chemical agent *MNNG* is repaired by a novel process that appears to exemplify an adaptive response to DNA damage. The *MNNG* adds a methyl group (CH_3) to various sites in DNA, including an oxygen atom at position No. 6 of guanine (*left*), disrupting the guanine's hydrogen bonding to a cytosine. Here the repair is accomplished by a protein that is synthesized in the cell during a period of adaptation when the cell is exposed to a low level of *MNNG*. A cysteine (one of the 20 amino acids) on the protein acts as a methyl acceptor: it binds the CH_3 group, thereby restoring the guanine to its original state (*right*).

Molecule-size electronic devices? Why not? Nature does it.

Nevertheless, it won't be easy.

In scarcely the time of a single generation, the human environment in this and similar countries has changed in that so many of us must make and keep our places in society by use of our wits instead of our muscles. Technology, whether worshipped or scorned, did it. To scorn the "computer age" becomes a bit pathetic. Too late. We can't make a phone call, or engineer a gene, without the computer.

The computer age was made possible by refining the crude technology of lithography, the drawing of pictures on stone with a greasy ink enabling multiple copies to be created. The refinements involved more facile media than "stone" and "grease," the ability of photons to draw details finer than the human hand, and finally, more complex ideas about what to draw. Further development of these refinements continues in thousands of minds and laboratories. Results of progress take virtually invisible form—so incredibly small have those silicon chips become.

But so what? Is there meaning to the seeing of a single molecule, in the sense that we see an elephant at the zoo? No, the laws of physics being what they are, the seeing must be more intellectual than sensory.

At the Kodak Research Laboratories, where chemistry made a considerable contribution to launching the art of microelectronics through photolithography, a crew works to explore the potential for finer and finer detail, one of those refinements mentioned above. This crew finds it necessary to work as scientists rather than technologists. As scientists, they publish.* Scientists need the stimulation of other scientists they don't lunch with every day. Here some members of the crew enunciate basic truths on which discourse in this particular field rests.

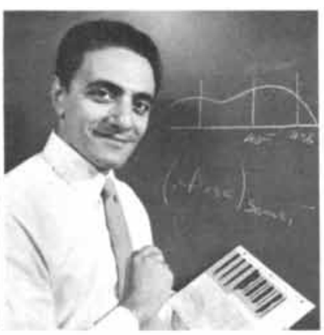
Kodak

---

*e.g.:
Pure & Appl. Chem. 49: 523-538 (1977)
Pure & Appl. Chem. 51: 241-259 (1979)
Prog. Polym. Sci. 5: 61-93 (1977)

J. L. R. Williams: *"The mobility of a solute in a polymer matrix controls both its photophysics and its photochemistry."*

S. Farid: *"The photographic response of light-sensitive polymers depends on the sensitizer photophysics, particularly intersystem-crossing efficiency."*



R. C. Daly: *"Syntheses of photopolymers require careful tailoring of compositions, molecular weights, and glass transition temperatures."*

D. P. Specht: *"The structural features of optical sensitizers are designed to match the physical and photochemical requirements of the polymers."*


For 1982
THE CHRYSLER LE BARON SERIES
They are like no other cars in America, Europe or Japan.
Sedan as shown $8,800.**
Standard equipment includes padded vinyl roof, rack-and-pinion power steering, power front disc brakes/rear drum brakes, cloth bench seat with folding center arm rest,
No other cars combine their luxury, high mileage, six-passenger room and unexpected price.


Coupe as shown $8,706.**

SCIENCE AND THE CITIZEN

Energy Czar

The U.S.S.R. has been the world's largest oil producer since it overtook the U.S. in 1974. Its present output is 12.2 million barrels per day, and production is increasing. Current U.S. output is 10.2 million barrels per day, down from a peak of 11.3 million barrels in 1970. The Russian target for 1985 is between 12.4 and 12.9 million barrels per day, a goal the U.S. Central Intelligence Agency scoffed at a few years ago but one that other observers, now including the Defense Intelligence Agency (DIA), consider attainable. In 1977 the CIA predicted that Russian output would peak at no more than 12 million barrels per day in the early 1980's and then would fall rapidly to between eight and 10 million barrels. The CIA's prediction of a Russian energy shortage is one factor that has been cited by those who believe there is a serious risk of Russian aggression in the Middle East. The CIA now concedes, however, that Russian output in 1985 could remain as high as 11 million barrels per day.

Although Russian success in meeting its oil production goals is still subject to some controversy, the success of the U.S.S.R. in finding and exploiting natural-gas resources is unquestioned. Whereas Russian oil production increased at the rate of 7.3 percent per year between 1960 and 1980, production of natural gas increased at an annual rate of 12 percent. By 1980 the volume had reached some 15.4 trillion cubic feet, equal to about three-fourths of the U.S. production. The Russian target for 1985 is about 22 trillion cubic feet, which will probably be enough to surpass the U.S. in that year.

The U.S.S.R. is now the world's second-largest exporter of oil (3.2 million barrels per day) and the largest exporter of natural gas (1.6 trillion cubic feet per year). As energy prices have increased in recent years the exports have paid off handsomely. According to David Wilson of the University of Leeds the U.S.S.R. had a surplus of $5.2 billion in 1979 in its trade with "hard currency" countries (Western Europe, Japan and the U.S.). Wilson's conclusions appear in the report "Soviet Oil & Gas to 1990," published by The Economist Intelligence Unit Ltd.

The DIA estimates that exports of oil and gas alone could raise Russian hard-currency income to about $22.5 billion per year in the mid-1980's. The DIA suggests that Russian oil production could level off in the late 1980's but predicts that it will rise again in the 1990's.

Coal has not been neglected in the Russian fuel economy, but its contribution has been declining steadily. From 1960 through 1980, while oil and gas production were growing by 311 percent and 863 percent respectively, coal output increased by only 33 percent. The production goal for 1980 was 550 million standard tons. (A standard ton of coal is one with a heat value of 27.8 million B.t.u.) Wilson estimates that the actual output was just under 500 million standard tons, a shortfall of about 10 percent. U.S. production in 1980 was 675 million standard tons (or 830 million short tons), up 91 percent over 1960. Ninety percent of the increase in the U.S. tonnage came from surface mines, which now yield about 60 percent of the total output. The U.S.S.R. has been slowly expanding surface mining, but 60 percent of its coal still comes from underground mines.

In 1960 coal accounted for 54 percent of all fuel production in the U.S.S.R., oil for 30 percent, natural gas for 8 percent and miscellaneous low-grade fuels for the remaining 8 percent. In 1980, according to the Wilson report, the probable distribution was coal 26 percent, oil 45 percent, natural gas 27 percent and other sources 2 percent. In the U.S. in 1980 the comparable amounts (excluding hydroelectric and nuclear power) were coal 31 percent, oil 35 percent and natural gas 34 percent. Total U.S. production of the three fuels in 1980, expressed in quads (units of 10^{15} B.t.u.), was 60.8. The corresponding Russian total was 52.4 quads. In 1960 the U.S. produced 39.9 quads and the U.S.S.R. 17.7 quads. Thus over the 20-year period the U.S. increased its production of coal, oil and gas by 52.4 percent while Russian output virtually tripled.

The greatest expansion in Russian production of oil and gas has come in western Siberia, where oil was discovered in 1960. The potential of the Siberian fields was consistently underestimated even by Russian experts. As late as 1969 it was said that Siberian oil would never account for more than a third of the total Russian output. By 1976, however, Siberia was already contributing 35 percent of the total and by 1980 virtually half; Wilson estimates that if the target of from 12.4 to 12.9 million barrels per day is achieved in 1985, at least 75 percent of it will be pumped from western Siberia.

Beginning in 1963 several supergiant natural-gas fields were found north of the Arctic Circle on the Yamal Peninsula, followed in 1966 by the discovery of the world's largest gas deposit at Urengoi, south and east of Yamal. Urengoi is thought to hold more than 280 trillion cubic feet of gas, considerably more than the total U.S. proved reserves of 195 trillion cubic feet.

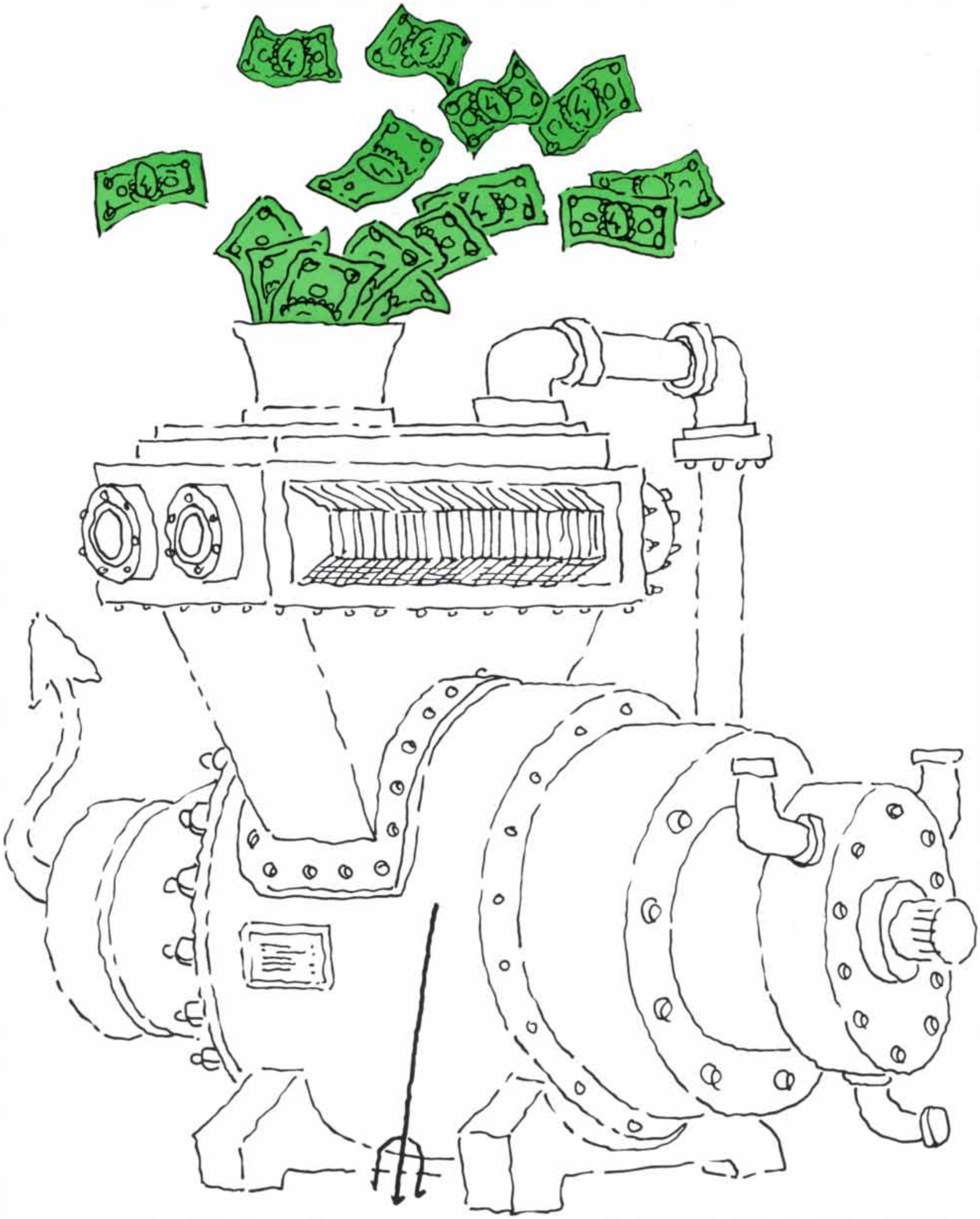
In order to profit from their gas discoveries the Russians have been negotiating for the construction of a $10-billion pipeline 3,500 miles long to convey some 1.4 trillion cubic feet of gas per year from western Siberia to seven European countries. If the pipeline is built, West Germany will get 30 percent of its gas from the U.S.S.R., compared with 17 percent today. At the Ottawa conference of Western leaders last spring President Reagan urged West German Chancellor Helmut Schmidt to reconsider the venture, promising U.S. coal and uranium to replace the Russian gas. Subsequently the Deutsche Bank announced it had reached a new agreement with Russian negotiators. The bank will provide about $2 billion in credits, roughly half the amount originally requested, which will be applied mainly to the purchase of compressor stations and pipe manufactured in West Germany. Instead of two parallel pipelines it now appears only one will be built. The reduced scale of the project probably reflects the current state of the world economy and smaller projections of energy demand rather than a response to the Reagan Administration's disapproval.

Stalking the Wild Weakon

The first glimpse of a weakon, the most highly valued prize in elementary-particle physics today, may come in the next few months, perhaps before the end of the year. The existence of a family of three such particles, which are also known as intermediate vector bosons, has been postulated for more than a decade. In the modern view of the forces of nature the weakons serve as the carriers of the weak nuclear force; they play a role analogous to that of the photon, the carrier of the electromagnetic force. (The other two fundamental forces—gravity and the strong nuclear force—are also thought to be transmitted by intermediary particles, namely the graviton and the eight particles called gluons.)

Assuming that the weakons exist, they are expected to be seen first in the aftermath of violent collisions between high-energy beams of matter and antimatter circulating in opposite directions in a storage ring. The world's first colliding-beam machine with enough energy to create weakons was successfully tested this summer and is now being readied for the search. It is a converted proton-accelerating ring at the European Organization for Nuclear Research (CERN) in Geneva. If the weakons have the expected properties, the CERN investigators should start detecting them soon, at a rate of one weakon every few hours.

Imagine turning hot exhaust into cold cash.



We did.
And lowered fuel costs up to 30%.

You don't become a Garrett engineer by letting energy go to waste. If there's a way to save it, you find it. And one of the things we've found is a way to help make industrial gas turbines work even harder.

By applying Garrett's advanced heat transfer technology, we've designed a new kind of gas turbine regenerator that can reduce a turbine's fuel bill by up to 30%. Simply by making more efficient use of exhaust heat to increase the temperature of the incoming air.

Because greater efficiency goes hand-in-hand with longer service life, Garrett regenerators are designed to last up to 20 years before needing any maintenance. The credit goes to a patented core configuration that solves thermal fatigue problems. And the use of corrosion-resistant stainless steel that easily handles operating temperatures up to 1,200°F. These new regenerators will be on-line and operational in more than 30 installations by the end of 1981.

Equally remarkable is the application of that same heat transfer technology to other areas of industry. By putting high temperature flue gas back to work with a special kind of heat exchanger called a recuperator, we'll be reducing the fuel requirement of large industrial furnaces this year by more than 25% Comparable savings are also available for direct-fired manufacturing processes, like glass-making.

At Garrett, we earned our wings in aerospace, then built ourselves into an industrial giant by keeping our feet on the ground. And our hands on a technology base one step ahead of tomorrow. We're uniquely structured to put that asset to work for you, today. Our engineers work in teams. No one is isolated. They work together. The result has been results.

You'll find us in everything from nuclear powerplants to hospitals. Agriculture to automobiles. But the real frontier for us, is you.

If you'd like to make your endeavors more energy efficient, you'll find us easy to work with. Just write our President, Jack Teske. You'll find him at: The Garrett Corporation, P.O. Box 92248, Los Angeles, CA 90009.

GARRETT The Garrett Corporation
One of The Signal Companies

Garrett

The aerospace company with its feet on the ground.

Although the weakons are closely related to the photon, they differ from it in certain important respects. Whereas the photon has no mass and no electric charge, the three weakons are predicted to be very heavy and two of them are charged particles. The charged weakons (designated W^+ and W^-) are thought to have a mass of about 80 GeV each; the third weakon (designated Z^0) is electrically neutral and should have a mass of some 90 GeV. (In high-energy physics the mass of a particle is often expressed in terms of its equivalent energy in GeV, or billions of electron volts; for comparison the mass of the proton is equivalent to less than 1 GeV.) Finding the three weakons at their predicted masses is considered essential to the confirmation of the prevailing theory linking the weak force and the electromagnetic force.

As recently as five years ago it was generally thought the large mass of the weakons would put them far out of reach of present-day particle accelerators, the largest of which are not colliding-beam devices but instead direct a single beam of particles at a fixed target. The weakons were not expected to be observed until the mid-1980's or later, when a new generation of ultrahigh-energy colliding-beam machines were scheduled to be put into operation. Then an ingenious short cut was proposed. Instead of building a new accelerator, it was suggested, an existing fixed-target accelerator could be converted into a colliding-beam machine by generating a counterrotating beam of antiparticles in the same annular space occupied by the original beam of particles.

Storage rings in which electrons and positrons circulate in a single tube have been operating for more than a decade. Their energy is limited, however, because the electron and the positron have a very low mass. A ring of a given size can attain a much higher energy by storing protons and antiprotons, but until recently it was not possible to accumulate a sufficiently dense "bunch" of high-energy antiprotons.

This problem has now been solved, and two large fixed-target accelerators are being converted into proton-antiproton storage rings. One of the accelerators is the Super Proton Synchrotron at CERN; it has already registered collisions of 270-GeV protons and antiprotons. The other is a proton synchrotron now under construction at the Fermi National Accelerator Laboratory (Fermilab), which will employ a ring of superconducting magnets to reach an energy of 1,000 GeV per beam. The Fermilab conversion is running about two years behind the CERN project; the Fermilab machine is not expected to be fully operational until 1984.

Meanwhile the next generation of colliding-beam accelerators remains a distant prospect. A colliding-beam device called ISABELLE, under construction at the Brookhaven National Laboratory, has run into technical and financial difficulties and faces an uncertain future. In ISABELLE two intersecting storage rings would bring two beams of protons into collision. A still larger device, the LEP, or large electron-positron, storage ring at CERN, has just been approved; this vast machine, which would measure 27 kilometers in circumference, could be in operation by the late 1980's. The main advantage of the bigger accelerators, if they are completed, would be the much greater "luminosity" of their counterrotating beams and thus their higher production rate of unusual particles such as the weakons. In LEP, for example, weakons may be made available for study every 10 seconds or so.

cAMP Site

In organisms higher than the bacteria the molecule cyclic adenosine monophosphate (cAMP) serves as a ubiquitous "second messenger" mediating between hormones and genes. In bacteria such as *Escherichia coli* it has a different role. It forms a complex with a receptor protein (CRP) and the complex serves to activate certain genes, notably those governing the breakdown of lactose and some other sugars to yield glucose. In *E. coli* the cAMP-CRP complex binds to a region of the DNA near the beginning of a regulated gene and enhances transcription of the DNA into RNA. The RNA in turn is translated to make an enzyme that catalyzes the breakdown of the sugars.

The assumption has been that the only function of cAMP in these events is to change the shape of the CRP molecule. The receptor protein in its new conformation could then promote transcription in either of two ways. It might modify the enzyme RNA polymerase, which carries out transcription, or it might destabilize the DNA itself, perhaps by separating the two strands of the double helix so that the polymerase can initiate transcription.

This view has now been challenged by Richard H. Ebright and James R. Wong, who propose a detailed model for gene activation in *E. coli* in which part of the cAMP molecule plays a direct role in destabilizing DNA; CRP serves primarily as a jig to carry cAMP to the correct site on the double helix. Ebright and Wong set forth their model in *Proceedings of the National Academy of Sciences of the United States of America*. When they wrote their article, they were both undergraduate students at Harvard College; they are now continuing with postgraduate work at Harvard. Theirs is the first article written solely by undergraduates to be published in the journal.

The cAMP molecule is a modified form of the nucleotide adenosine monophosphate. The nucleotide consists of the base adenine, a ribose sugar and a

phosphate group; in the cyclic form the phosphate is bonded back to the sugar, forming a six-member ring fused to the five-member sugar ring. Ebright and Wong found that cyclic nucleotides in which the sugar-phosphate portion is modified do not bind to CRP, whereas nucleotides having bases other than adenine bind to CRP as readily as cAMP does. It is the sugar-phosphate portion, then, that interacts most closely with CRP; the adenine remains somewhat exposed. And it is the adenine that is required for regulatory activity. Even though cyclic nucleotides incorporating bases other than adenine bind readily to CRP, they do not form complexes that promote transcription.

A comparison of the structure of adenine with that of the inactive cyclic nucleotides showed that activity depends on the presence of a particular nitrogen atom and a particular amino group (NH_2) that can serve as terminals for hydrogen bonding. The same two sites anchor hydrogen bonds between an adenine and the complementary base thymine in DNA. It is these hydrogen bonds, along with those between the complementary bases guanine and cytosine, that link the two strands of the DNA double helix.

The findings with respect to both binding and activity were confirmed by experiments with simple molecules called indoles, some of which function like cAMP in the activation of *E. coli* genes governing the metabolism of the sugar arabinose (but not lactose). Indoles have a fused-ring structure somewhat similar to the sugar-phosphate ring system of the cyclic nucleotides; various side chains that project from the indole rings are structurally analogous to the base in a nucleotide. Ebright and Wong found that the indole fused-ring structure binds to CRP regardless of its side chain, just as the sugar and phosphate bind regardless of the base in the case of a cyclic nucleotide. The side chains of the indoles are exposed and particular side chains are essential for transcriptional activity. Hence the side chain of an active indole seems to have the same role as adenine has in cAMP. The active side chains are very different structurally from adenine, but they all terminate in an oxygen atom and either an OH or an NH_2 group, and these groups are spatially disposed to form hydrogen bonds to a DNA nucleotide base.

Ebright and Wong conclude that the distinction between the bound and the exposed domains on cAMP (and on indoles) and the requirement for a specific hydrogen-bond pattern in the exposed domain suggest a mechanism for gene activation by the cAMP-CRP complex. They propose that when CRP recognizes and binds to a specific sequence of nucleotides on the DNA, it inserts the cAMP's exposed adenine partway into the helix. When the hydrogen bonds of a thymine-adenine base pair in the DNA break momentarily (they do so periodically as the DNA "breathes"), the thymine forms new hydrogen bonds to the intrusive cAMP adenine. A link between the two DNA strands is thus disrupted, and so the helix is destabilized locally. The destabilization spreads, the two strands open up and RNA polymerase gains access to initiate transcription.

Political Arithmetic

A U.S. congressman may occasionally have divided loyalties, but no congressman is himself divisible. This simple fact has given rise to a recurrent political controversy over the apportionment of representation in the House of Representatives. Although Article I of the Constitution requires that a state be represented in proportion to its population, the framers failed to specify the mathematics to be used in allocating representatives. Because the proportion of the nation's population that lives in a given state rarely corresponds to a whole number of representatives, a rounding procedure must be employed.

It might seem there should be a method that is simple, equitable and universally acceptable. It has been proved, however, that no procedure is free of all the possible inequities. In the history of the House several methods have been employed; their flaws have led to considerable wrangling. This controversy, which was the occasion of the first presidential veto in U.S. history, has been given new significance by the returns of the 1980 census. If the results of the census are not altered, 17 seats in the House will be transferred from the older industrial regions of the Northeast and Middle West to the South and West, reflecting the migration that has taken place in the past decade [see "The Census of 1980," by Philip M. Hauser, page 53].

The mathematics of apportionment has recently been discussed by William Lucas and David Housman of Cornell University; writing in *Engineering: Cornell Quarterly,* they point out that the approaches taken to the problem are of two kinds. The first kind of procedure is based on a number called the quota, which is the fraction of the national population that lives in a state multiplied by the total number of representatives in the House. (The size of the House has been fixed since 1912 at 435.) The quota is ordinarily not a whole number but an integer with a fractional remainder. In one quota procedure, called the method of largest fractions, each state is first allotted a number of representatives equal to the whole number in the quota. The seats that remain are then allotted to the states with the largest remainders.

Alexander Hamilton proposed the method of largest fractions in 1792 to divide 120 seats among the 15 states in the first House of Representatives. Al-

Each plate bearing an original work of art created by Basil Ede exclusively for this collection.

Each decorated with a border of pure 24 karat gold.

Available by subscription only.
Advance subscription deadline:
November 30, 1981

THE ARTIST. Basil Ede, of Sussex, England, has been called "the outstanding wild bird portraitist of his generation, and perhaps of his century." He is widely regarded as Europe's most distinguished bird artist. Indeed, the authority of his style and the distinction of his work have profoundly influenced the way birds are portrayed by artists throughout the world.

After notable exhibitions in London and other major European cities, Basil Ede was honored by a one-man show at the Smithsonian Institution's National Collection of Fine Arts in Washington, D.C. This was followed, during recent years, by important exhibitions at New York's famous Kennedy Galleries.

His paintings have been commissioned by the National Audubon Society and the World Wildlife Fund, among others. And he is represented in many public and private collections, including the Smithsonian Institution and the collection of HRH Prince Philip, Duke of Edinburgh.

Now, at the height of his career, Basil Ede has created an important new series of superlative works of art in fine porcelain — *Water Birds of the World*. Twelve

CANADA GOOSE

collector's plates portraying the wild beauty and majesty of water birds in precise, authentic detail — with Ede's inimitable flair for color and composition.

Each of these twelve plates is, in itself, a masterful work of art. Together, they form an incomparable collection that will be a proud acquisition for every sub-

WOOD DUCK

scriber . . . a focus for conversation and admiration when displayed in the home.

THE WATER BIRDS PLATES. The plates will be large in size — 9 inches in diameter — to provide full scope for Ede's finely detailed portrayals of water birds in their natural habitat.

And they will be crafted under the supervision of Franklin Porcelain in Japan, by craftsmen schooled in the delicate, demanding art of fine porcelain.

Franklin Porcelain has devoted more than two years to meticulous preparation for the issuance of these plates, and every detail will be of the highest quality. For example, each plate will be hand-decorated with a border of pure 24 karat gold. And each plate will incorporate as many as sixteen separate ceramic colors.

Ede's very beautiful and marvelously accurate works of art — created especially for this collection and available only on these fine porcelain plates — make this an acquisition that will be enjoyed and prized by all who love the beauty of birds, of art, and of porcelain.

THE SIGNATURE EDITION. The first edition of this collection will be a special Signature Edition. Each collector's plate in this limited edition will bear Basil Ede's signature on the back in 24 karat gold. This signature will appear *only* on the plates in this first edition.

The plates in this unique edition will be crafted exclusively for individual subscribers. The edition is available only until the end of 1982. And the limit of one collection per subscriber will be enforced without exception. Thus, the total number of sets of the Signature Edition will be forever limited to the exact number of original subscriptions entered during the offering period.

ADVANCE SUBSCRIPTION DEADLINE: November 30, 1981. The Signature Edition will be issued to subscribers at the convenient rate of one plate every other month. The original issue price of just $55 per plate is payable in two convenient monthly installments of $27.50 per month. Each plate will be accompanied by specially written reference material and a special wall mount. And a Certificate of Authenticity will accompany each collection.

There is no need to send any payment now. But your advance subscription application is valid only if postmarked by November 30th.

ADVANCE SUBSCRIPTION APPLICATION

WATER BIRDS OF THE WORLD

Valid only if postmarked by November 30, 1981.

Franklin Porcelain
Franklin Center, Pennsylvania 19091

Please enter my subscription for the *Water Birds of the World* porcelain plate collection, bearing new and original works of art by Basil Ede and his signature in 24 karat gold.

I need send no payment now. The twelve plates are to be sent to me at the rate of one plate every other month, and I will be billed for each plate in two equal monthly installments of $27.50* each, beginning with shipment.

**Plus my state sales tax.*

Signature ____________________
ALL APPLICATIONS ARE SUBJECT TO ACCEPTANCE.

Mr.
Mrs.
Miss ____________________
PLEASE PRINT CLEARLY

Address ____________________

City ____________________

State, Zip ____________________

Limit: One collection per subscriber.

though the bill specifying the method passed Congress, it was vetoed by Washington on the advice of Thomas Jefferson, who had another procedure in mind. (Hamilton's method was revived 60 years later, however, and thereafter served as the basis of apportionment until 1910.)

Hamilton's method has one important mathematical virtue, called the quota property: the number of representatives allotted to each state is either the largest whole number in the quota or the next-largest integer. The method does, however, lead to two mathematical and political problems. If there is a shift in population among the states (as there has been in the past decade), one state may lose a seat to another even though the population of the first state increased proportionally more than that of the second. This effect has been designated the population paradox. Moreover, if the number of representatives in the House is increased, a state may lose a seat even though there has been no change in the population of any state. This "paradox of House size" has been realized several times.

Procedures of the second kind, which are called divisor methods, were formulated in an attempt to avoid these paradoxes. The starting point of a divisor method is a congressional district of a predetermined size: the divisor. The Constitution specifies that the divisor be at least 30,000. The population of each state is divided by the divisor; since the state population is rarely an integral multiple of the divisor, a rule is needed for dealing with the remainder.

Four such rules have been devised; they all avoid the paradoxes, but none of them possesses the quota property. (Indeed, it has been proved that no procedure can have the quota property and always avoid the population paradox.) The first divisor method is the one Jefferson advanced in 1792, and which was adopted for apportioning the first House. Jefferson's method ignores the remainder, so that the number of representatives is equal to the number of times the divisor will go into the state population completely. In addition to lacking the quota property, Jefferson's procedure is biased in favor of states with a large population because the disregarded remainder is likely to be a smaller proportion of such a state's total population. Lucas and Housman note that the bias "probably did not escape Jefferson and his friends from Virginia, the largest state at the time."

In rejoinder to Jefferson, John Quincy Adams of Massachusetts in 1832 proposed an alternative method equally biased in favor of small states. In the Adams method all fractions are rounded up to the next-largest integer. This proposal was never adopted by Congress. In a third divisor method, favored by Daniel Webster, all fractions are rounded to the nearest whole number, whether it is larger or smaller than the fraction. Webster's method is not biased toward either large or small states. It was adopted in the 1840's and was employed again from 1912 to 1940.

The method currently in use is a fourth divisor method formulated by Joseph A. Hill of the Bureau of the Census and Edward V. Huntington of Harvard University. Their procedure is designed to minimize a measure of inequity in the apportionment of representatives. Under the present law the variable that is minimized is the difference among the states in the ratio of population to seats in the House. The measure that is now utilized is the relative rather than the absolute difference, but the method could accommodate any measure of inequity. Like the other divisor methods, the Huntington-Hill formula avoids the population paradox and the paradox of House size, but it does not satisfy the quota condition; the number of representatives can be smaller than the whole number of the quota or larger than the next-largest integer.

Unless there is a change in the law before the elections of 1982, the Huntington-Hill method will determine the apportionment of the next Congress. The procedure was adopted in 1942 by a Congress in which a majority of the representatives were Democrats. It had been determined that the substitution of the new procedure for the Webster method would result in the transfer of a seat from Republican Michigan to Democratic Arkansas. "Except for the Democrats from Michigan," Lucas and Housman observe, "the vote was strictly along party lines."

You Are What You Eat

Primates other than man depend mainly on plants for food. Presumably man's forebears were also herbivorous before they began to scavenge meat from dead animals and to kill game. Even then they probably continued to eat at least some plant foods, and when hunting was largely replaced by farming, the amount of plant material in the diet must have increased again. How can one measure the proportions of meat and vegetables eaten by hominids whose only remains are teeth and bones? A peculiarity of metabolism offers a clue.

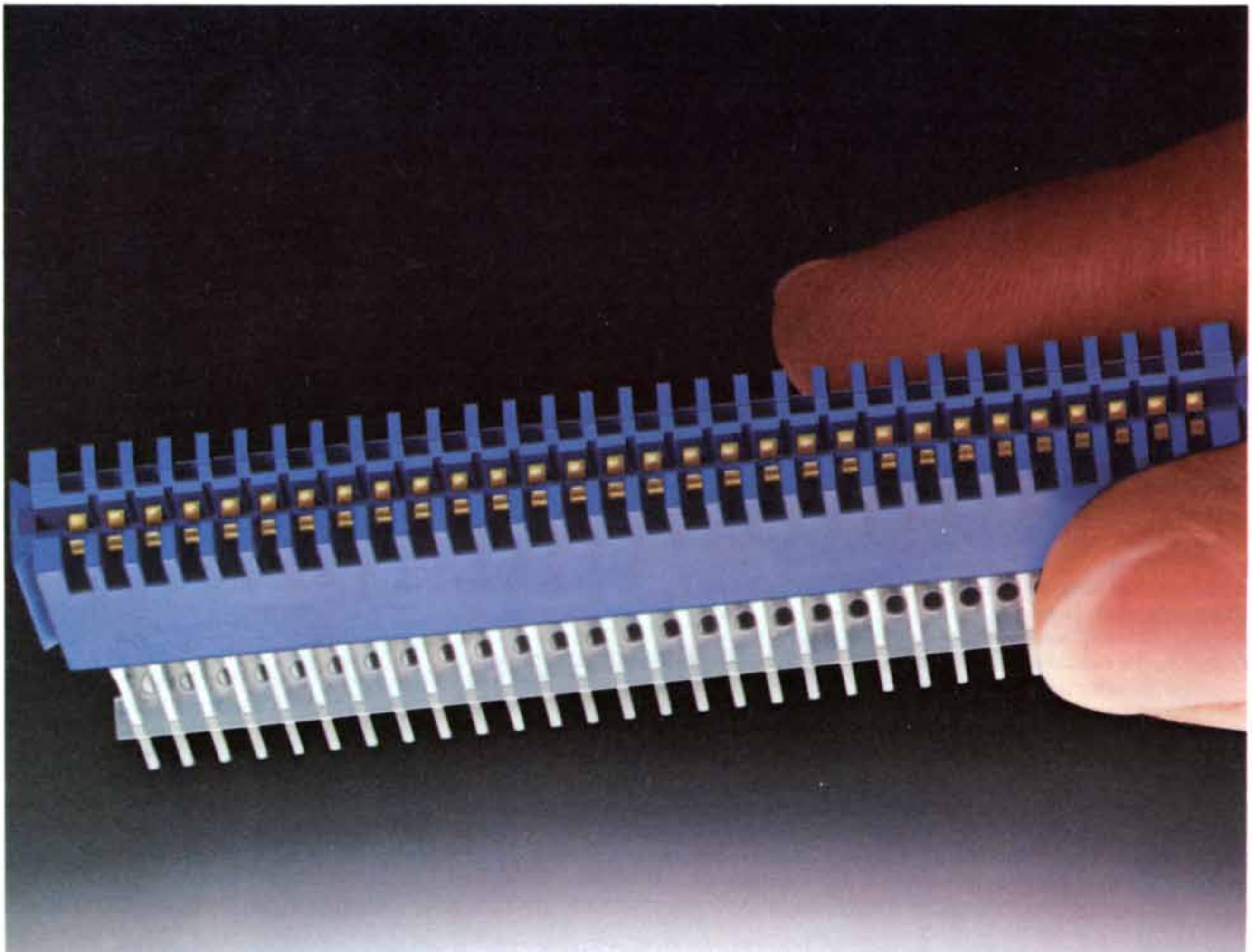
The clue is found in the biochemistry of the elements calcium and strontium. Calcium, of course, is an essential nutrient and a major constituent of both teeth and bones. Strontium is chemically similar to calcium and therefore tends to accumulate in the same tissues. Plants, which take up both elements from ground water, have a relatively high ratio of strontium to calcium in their tissues. Animals that feed on plants discriminate digestively against strontium and so the strontium-to-calcium ratio of their tissues is lower than that of the plants they eat. Carnivorous animals that feed on herbivores discriminate against the reduced amounts of strontium present in their prey and so the strontium-to-calcium ratio of a carnivore is lower still. Ideally the ratio of an omnivore should fall between that of a herbivore and that of a carnivore.

What is needed to test this hypothesis is an archaeological site where the bones of human beings have been preserved along with those of known carnivores and herbivores. The site should also be representative of some important period in prehistory, such as the interval between the disappearance of Old Stone Age hunters and the rise of New Stone Age farmers. Several such sites exist in the eastern Mediterranean. From one of them, Hayonim Cave in Galilee, Andrew Sillen of the Smithsonian Institution has obtained a collection of animal and human bones. The bones come from a 10,000-year-old level representative of the Natufian culture, which flourished in the Levant in the interval between the Old and the New Stone Age. The material was provided to Sillen by colleagues at the Hebrew University of Jerusalem and the University of Tel Aviv Medical School.

Sillen reports the results of his study in *American Journal of Physical Anthropology.* The strontium-to-calcium ratios are expressed as a proportion of the ratio prevailing in the earth's crust as a whole. Sillen writes that the average ratio for a representative herbivore (gazelle) was .98 and that for a representative carnivore (fox) was .63. The human bone samples, all from adults, averaged .77, an intermediate value suggestive of an omnivore whose intake of plant foods was moderate. Sillen's findings support earlier conclusions about the Natufian diet. The presence at the sites of flint-edged sickles and stone artifacts suitable for processing plant foods has long been accepted as evidence that the Natufian hunters exploited local plant resources on a considerable scale.

Galois's Last Hours

The story of Évariste Galois, the boy mathematician who invented group theory in his adolescence, is one of the finest romances in the history of science. As the tale is told, Galois was a brilliant but rebellious student, misunderstood by his teachers and ignored by the mathematical establishment. Committed to radical politics, he was goaded into a duel by a female agent provocateur at the behest of his political enemies. The night before the duel, in a few hours of desperate clarity, he wrote out his mathematical ideas. The next morning, May 30, 1832, Galois was shot in the stom-



How we're putting gold into connectors you can afford.

A lot of companies that depend on electrical connectors are trying to reduce their usage of gold.

They can with AMP.

The reason? We've designed connectors with less gold that perform as well as ever.

One beauty of gold is its unique reliability in electrical contact areas. But the predictability of its price is something else. That's why we developed our own method of selective plating to put the gold only where it's needed. We call it ACCU-PLATE. It's precise. Its advantages work with other metals, as well. And it's saving our customers money.

Providing reliable platings that are cost-effective isn't new to AMP. Over the past decade, we've reduced the gold in our connectors by 33%.

What can we do for you in the '80s?

Whether you call us early with a problem, or simply call us for connectors, you'll get what the world has come to expect from AMP. Connectors and interconnection systems designed for performance. And as the industry leader in research and development, we'll continue to bring you products that reduce the cost of making yours.

AMP has a better way.

For product innovations, for new applications, for ways to save you money, AMP has better ways to help make your company more productive.

Write for our brochure, "AMP Has A Better Way." AMP Incorporated, Harrisburg, PA 17105.

Barcelona • Brussels • Buenos Aires • Dublin • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • London • Luzern • Mexico City
Montreal • Paris • San Juan • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo • Valley Forge • Vienna

AMP means productivity™

ach; he died the following day at the age of 20. Fourteen years later the manuscripts written on his last night were published and group theory was born.

It is perhaps remarkable that much of this dramatic tale can withstand the scrutiny of historical research. Nevertheless, the true romance of Évariste Galois is a somewhat different story. It has recently been investigated by Tony Rothman, a physicist and historian of science at the University of Texas at Austin, who reports his findings in *The American Mathematical Monthly*. Rothman notes that a number of writers, notably Eric Temple Bell, Leopold Infeld and Fred Hoyle, have sought to portray Galois as a genius frustrated by the mediocrity of those around him, who died before his work was completed at the hands of political assassins. In Rothman's account Galois remains a tragic figure but a more credible one.

The greatest impetus to the Galois legend, Rothman writes, was given by Bell's 1937 book *Men of Mathematics*. There Bell describes Galois's final hours: "All night long he had spent the fleeting hours feverishly dashing off his scientific last will and testament, writing against time to glean a few of the great things in his teeming mind before the death which he saw could overtake him. Time after time he broke off to scribble in the margin 'I have not time; I have not time,' and passed on to the next frantically scrawled outline."

The passage, Rothman notes, is almost entirely a fabrication. Galois had already written several articles on group theory, and some of them had been published. The papers he left on his last night were memoirs to which he was making annotations and corrections. At one point, in the margin of a manuscript, he wrote: "There are a few things left to be completed in this proof. I have not the time. (Author's note.)" It is the only instance in all the manuscripts where the phrase appears that Bell attributes to Galois "time after time."

It is undoubtedly true, Rothman asserts, that Galois's teachers in secondary school found him "singular," "bizarre," "original" and "closed." He showed little interest in the humanities and devoted almost all his energy to mathematics. His mathematics teacher, however, recognized his ability and recommended that he be admitted to the École Polytechnique without the usual entrance examination. The recommendation was not followed; Galois failed the examination twice and instead attended the École Normale.

It is also true, Rothman concedes, that there were repeated troubles over the publication of Galois's ideas, but there is certainly no evidence of a conspiracy to suppress his work. According to one legend, the eminent mathematician Augustin Louis Cauchy lost or destroyed several important memoirs sent to him

by Galois. Actually, according to a letter by Cauchy found in 1971 in the archives of the French Academy of Sciences by René Taton, Cauchy had read the memoirs and, apparently convinced of their importance, had planned to present them to the Academy in January, 1830.

As it turned out, Cauchy did not present the papers. The reason remains obscure, but Taton argues that Cauchy may have persuaded Galois to submit them as a single memoir eligible for the Academy's Grand Prize in Mathematics. In February, 1830, the 18-year-old Galois did submit a manuscript to Jean Baptiste Joseph Fourier, the Academy's secretary for mathematics and physics. Fourier died in April, however, and Galois's manuscript could not be found among Fourier's papers.

By all accounts Galois was a political firebrand in a time of great political tension. The restoration of the monarchy in the person of Louis-Phillipe, after the abdication of the Bourbons in the revolution of 1830, was a great disappointment to Galois. He became an ardent republican and was once arrested for toasting the king with a glass in one hand and a knife in the other.

In spite of such provocative acts, the available evidence suggests, according to Rothman, that Galois's death had nothing to do with politics. Galois's opponent in the duel, Pescheux d'Herbinville, was at least ostensibly a man who shared Galois's republican sentiments. Bell does not mention d'Herbinville and instead invokes an anonymous marksman in the pay of the Royalists; it is thereby suggested that Galois was assassinated and the affair of honor that led to the duel was a mere pretext. The name of the opponent is mentioned, however, in a contemporary account by Alexandre Dumas (an account, incidentally, that Bell lists among his sources). D'Herbinville, Rothman contends, is an unlikely candidate for the role of secret agent of the king: he had once been tried with a group of 18 other republicans for conspiring to overthrow the monarchy.

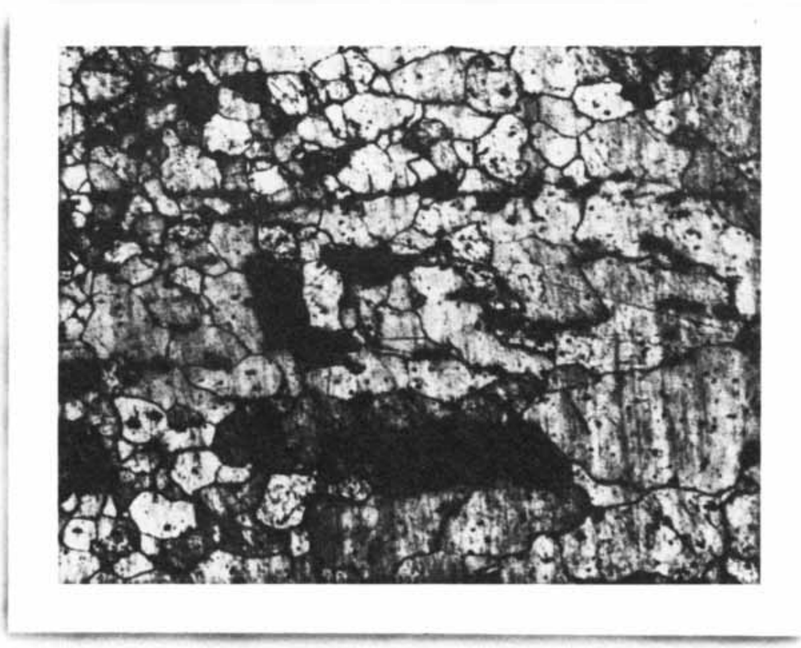
Rothman points out that the events leading up to the duel are consistent with a much simpler interpretation. Galois himself stated just before the duel that he had been "the victim of an infamous coquette and her two dupes." He went on to write, "Forgive those who kill me for they are of good faith," suggesting they were not his political enemies. The identity of the "infamous coquette" was determined in 1968 by C. A. Infantozzi from the partially obliterated signature on a letter. She was Stéphanie-Félicité Poterin du Motel, no agent provocateur but the daughter of a doctor with whom Galois had been staying. In short, the duel may well have been fought over a woman and was probably not the outcome of a political plot.

Geologists say more than 50% of all the oil ever discovered in the United States is still trapped underground in the pores of rocks. Now Texaco's using enhanced recovery techniques to force more of that oil out to meet your future energy needs.

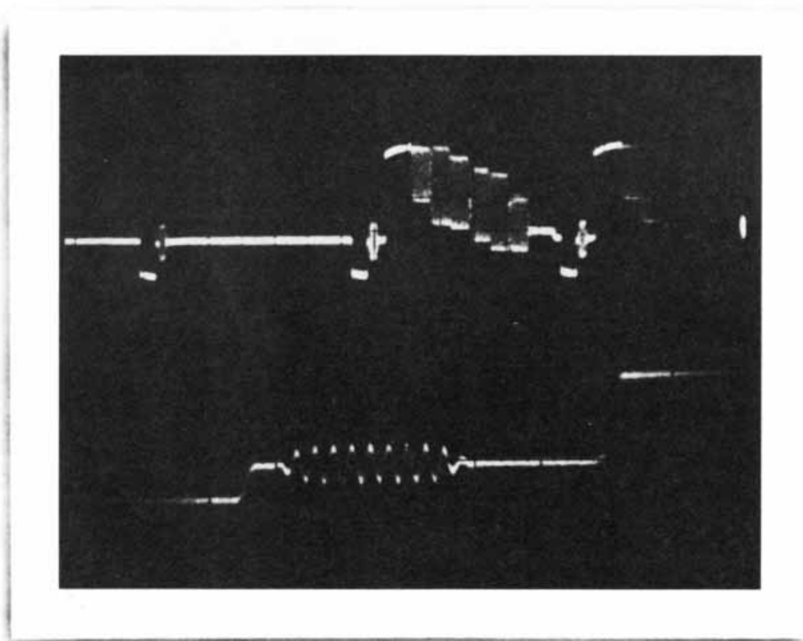
The usual pumping methods just can't tap all of the oil here in the U.S. That's because so much is locked away in the ground like water in a sponge, except the "sponge" is often solid rock. With enhanced recovery techniques like steam-flooding and carbon dioxide flooding, Texaco's now recovering some of that oil. In the years ahead, we'll be trying even more advanced techniques to recover more of the energy you need, and to recover more of it here at home.

TEXACO

You can trust the Star at home and in your car.

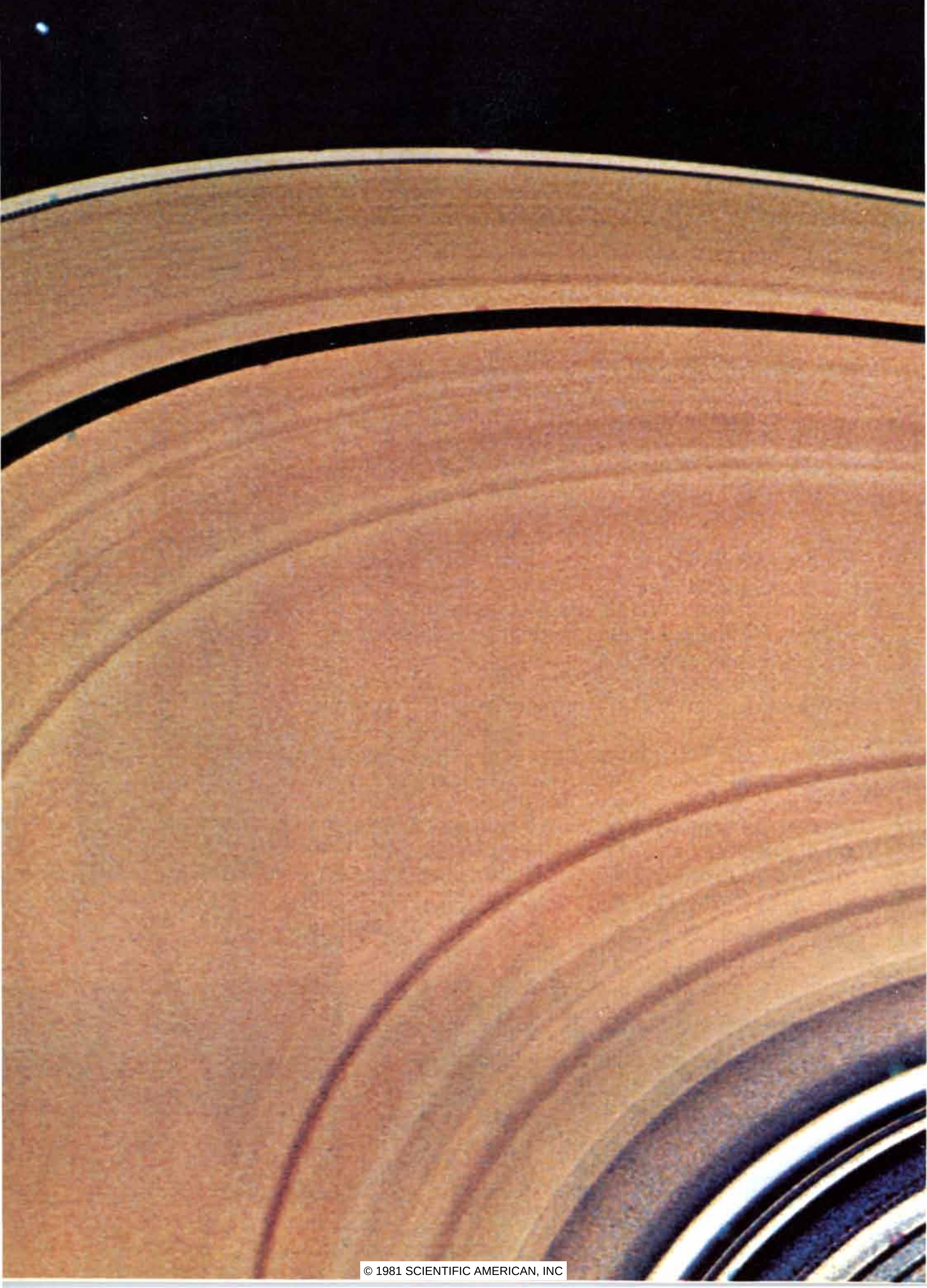


*Steel solenoid frame.
80x magnification. Cross polarized. Bright-field reflected illumination with
xenon source. Green filter.
20,000 ASA–
one-eighth-second exposure*



*Single sweep trace of
wave forms from a television
test signal generator on
new Polaroid Type 612
20,000 ASA film.
Upper trace made at 20μs/cm,
lower trace made at 1μs/cm.*

20,00



Rings in the Solar System

Three of the giant planets are now known to have them, and the rings around Saturn are now known to consist of myriad ringlets. The form of the rings is maintained by a complex interplay of sculpturing forces

by James B. Pollack and Jeffrey N. Cuzzi

The approach of the spacecraft *Voyager 2* to within 100,000 kilometers of Saturn on August 25 marks the climax of a period of planetary exploration in which the rings of Saturn have surprised us perhaps as much as they surprised the first men who saw them almost four centuries ago. Today the rings of Saturn stand revealed in a wealth of detail, including bands, spokes and braids. Some of the detail is unexplained. On the other hand, it emerges that Saturn is not unique in having rings. Jupiter has a ring, and Uranus has at least nine discrete rings. It remains to be determined whether Neptune has rings, and whether rings are therefore ubiquitous among the giant gaseous planets of the outer solar system. Here we shall discuss the structure and composition of the rings of Jupiter, Saturn and Uranus, with special attention to the rings of Saturn. Then we shall take up what is known about the processes that shape the rings. Finally, we shall consider several alternative explanations of how the rings arose.

Strange Appendages

The rings of Saturn were first observed in July, 1610. The observer was Galileo Galilei. Partly because the images produced by his invention, the astronomical telescope, were poor and partly because he had discovered the four largest moons of Jupiter only a few months earlier, he initially thought the blurry, earlike structures he saw were two moons close to Saturn. Soon his opinion changed. For one thing the "strange appendages" did not vary in their position with respect to Saturn from one night to the next. Moreover, in 1612 the appendages disappeared. Today it is plain that the rings had come to lie in an edge-on orientation as viewed from the earth in 1612 and thus had become quite faint.

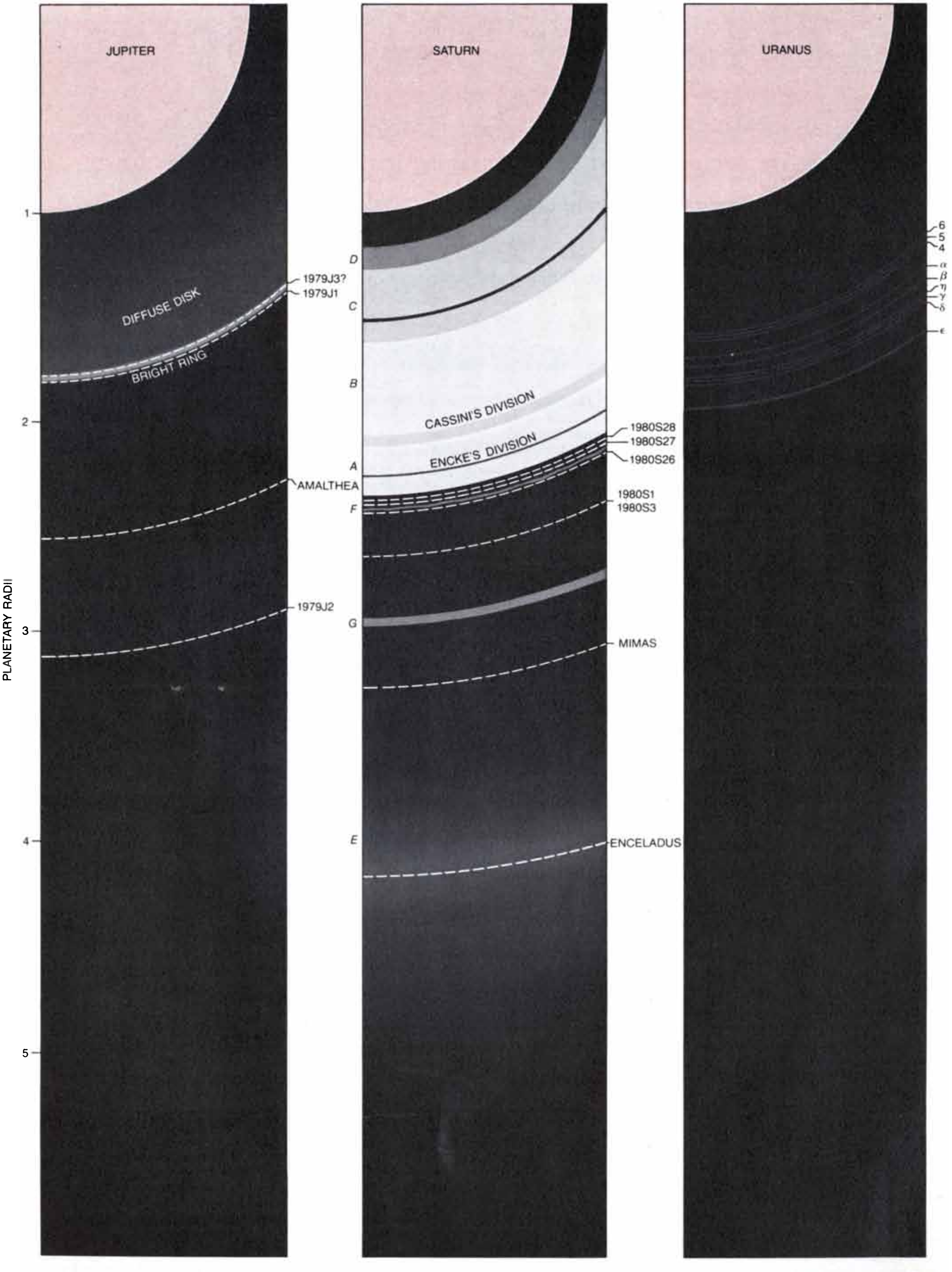
The geometry of the appendages puzzled astronomers. Indeed, it was proposed that the appendages were handles attached to Saturn or that they consisted of several moons in orbit only around the back of Saturn so that they never cast a shadow on the planet. Finally in 1655 Christiaan Huygens proposed that the appendages were the visible sign of a thin, flat, detached disk of matter arrayed in the equatorial plane of the planet. Depending on the position of Saturn and that of the earth in their orbits about the sun, the disk would vary in its tilt toward the earth; hence it would vary in appearance from a narrow line to a broad ellipse.

For the next two centuries it was assumed that the disk was a continuous sheet of matter. A difficulty with this hypothesis emerged, however, as early as 1675, when Jean Dominique Cassini found that the dark band now known as Cassini's division separates the disk into two concentric rings. Moreover, late in the 18th century Pierre Simon de Laplace showed that the combined forces of the gravity of Saturn and the rotation of the disk would rip apart a single broad sheet of matter. Basically any given parcel of the disk maintains its radial distance from Saturn because two forces are in balance. Gravity pulls the parcel inward; centrifugal force pushes it outward. The centrifugal force arises from rotational velocity; thus the disk is rotating. The problem is that for a rigidly rotating disk the forces balance at only one radial distance. Laplace therefore proposed that the rings of Saturn consist of many narrow ringlets, each one thin enough to sustain the slight imbalance of forces across its radial width.

The final step toward the modern view of the rings was taken in 1857, when James Clerk Maxwell won the Adams prize of the University of Cambridge for his mathematical demonstration that the ringlets actually consist of numerous tiny masses in independent orbits. Experimental proof for his hypothesis came from several quarters. In 1895, for example, the American astronomers James E. Keeler and William W. Campbell inferred the velocity of the particles in the rings from Doppler shifts: the altered wavelength of spectral lines in the sunlight the particles reflect toward the earth. They found that the rings rotate about Saturn at a rate different from that of the atmosphere of the planet. Moreover, the inner parts of the rings rotate at a greater velocity than the outer parts, just as the laws of physics would lead one to expect for particles in independent orbits.

The rings of Uranus were discovered by accident. A number of groups of astronomers had set out to observe the passage of the star SAO 158687 behind Uranus on March 10, 1977. The intent was to study the structure of the atmosphere of Uranus. Among the more successful observations were those made by James L. Elliot and his associates on the Kuiper Airborne Observatory, an aircraft equipped with a 91-centimeter telescope. His group (and several others) found that the brightness of the star decreased not only when it passed behind Uranus but also at a number of places close to the planet yet far above its atmosphere. The short-duration dimmings defined a series of distances on one side of Uranus that was roughly symmetrical with the series on the other side of the planet. The symmetry was taken to arise from the presence of fairly opaque, quite narrow and nearly circular rings. Later observations have revealed so far the presence of nine rings,

"A" RING OF SATURN was photographed on August 23 by the spacecraft *Voyager 2.* In the resulting image, displayed in false color, the *A* ring is yellowish; the broadest gap in the ring is Encke's division. The part of the ring outside Encke's division shows a faint set of bands. The bands are more closely spaced near the orbit of 1980S27, a moon discovered in images made by *Voyager 1* and visible as the white dot at the upper left. It is thought the bands are caused by resonances in the ring due to the gravitational effects of the moon. In addition 1980S27 and a second moon, 1980S26, are thought to shape the *F* ring of Saturn, too faint to show up here. The whitish and bluish ringlets at the lower right occupy Cassini's division. Their different colorations in this image indicate that their particles have somewhat different compositions.


JUPITER
SATURN
URANUS
PLANETARY RADII
1
2
3
4
5
DIFFUSE DISK
BRIGHT RING
1979J3?
1979J1
AMALTHEA
1979J2
D
C
B
A
F
G
E
CASSINI'S DIVISION
ENCKE'S DIVISION
1980S28
1980S27
1980S26
1980S1
1980S3
MIMAS
ENCELADUS
6
5
4
α
β
η
γ
δ
ε

all lying within one planetary radius of the top of the Uranian atmosphere.

The rings of Jupiter were discovered by spacecraft. The first hint came when *Pioneer 10* passed near Jupiter in 1974. Jupiter has a magnetic field that traps charged particles in regions of space surrounding the planet. The regions amount to the Jovian equivalent of the earth's Van Allen belt. What *Pioneer 10* detected was a decrease in the count of the high-energy particles in the Jovian belts 50,000 to 55,000 kilometers above the atmosphere of the planet. Mario H. Acuña and Norman F. Ness of the Goddard Space Flight Center of the National Aeronautics and Space Administration suggested that the decrease might be due to the partial absorption of the particles by a close-in satellite or a system of rings. The latter proved to be the case. A faint Jovian ring was detected in 1979 when it was photographed by imaging systems aboard *Voyager 1* and *Voyager 2*. The spacecraft were deflected toward Saturn by the gravitational field of Jupiter. They arrived near Saturn in November, 1980, and August, 1981.

Characteristics of Planetary Rings

The rings of Saturn, Uranus and Jupiter share a number of properties. First, they consist of myriad particles in independent orbits. Second, they lie far closer to their parent planet than the major moons of the planet do; in fact, the bulk of each ring system lies less than one planetary radius away from the surface of the planet. Third, the rings are centered on the equatorial plane of the planet; indeed, almost all the ring material is confined to a thin region in that plane. Fourth, the ring systems of Jupiter and Saturn have a number of tiny moonlets near or within the rings. It is suspected that similar moonlets are associated with the rings of Uranus.

Still, each ring system has its own peculiarities. The ring system of Saturn has seven major sections. Some of them are separated from neighboring sections by more or less empty annular gaps; the borders of the others are marked by changes in the packing density of ring particles. Each section is designated by a letter that reflects not its distance from Saturn but the order in which the sections were discovered or hypothesized.

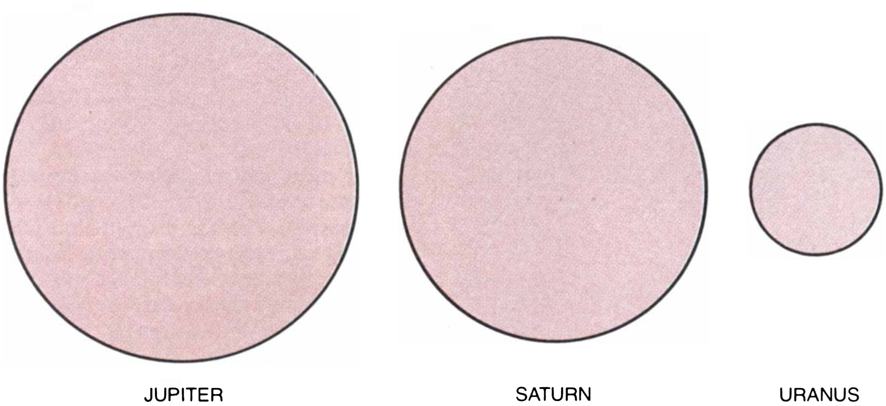
The main body of the ring system of Saturn includes, then, the bright and fairly opaque rings *A* and *B*. They are separated from each other by the 5,000-kilometer width of Cassini's division, a rather transparent region but definitely not an empty one. The main body of the Saturnian system also includes the fainter, less opaque *C* ring, which lies inside the inner edge of the *B* ring. It has a degree of opacity comparable to that of Cassini's division. The still fainter *D* ring lies inside the *C* ring. Finally, three very faint rings, *E, F* and *G,* lie outside the *A* ring. Taken together, the main rings of Saturn (the *A, B* and *C* rings) measure about 275,000 kilometers in annular width, or about three-fourths of the distance from the earth to the moon. In comparison, the thickness of the rings of Saturn is negligible. An upper limit of about a kilometer has been placed on their vertical extent. Compared with their width, the rings are thousands of times thinner than razor-thin.

Information on the composition of the particles in the rings of Saturn can be derived from the rings' ability to reflect or absorb light of different wavelengths. For example, the *A, B* and *C* rings are poor reflectors of sunlight at certain near-infrared wavelengths. This property is characteristic of water ice, which suggests that water ice is a major constituent of the surface of the particles that make up those rings. Water ice, however, is white in color; that is, water ice is more or less equally reflective at all visible wavelengths. In contrast, the particles of the *A, B* and *C* rings are less reflective in blue light than they are in red light. Perhaps some additional substance is present in small amounts. Dust containing iron oxide has been suggested as the source of the reddish color. It has also been hypothesized that compounds generated by the sun's ultraviolet radiation are responsible for the redness. Certain colorless sulfur-containing compounds become polysulfides under ultraviolet radiation, and polysulfides selectively absorb in the blue. One surprise in the Voyager data is that the particles in the *A* and *B* rings have a similar color but are brighter and more reddish than the particles in the *C* ring and in Cassini's division.

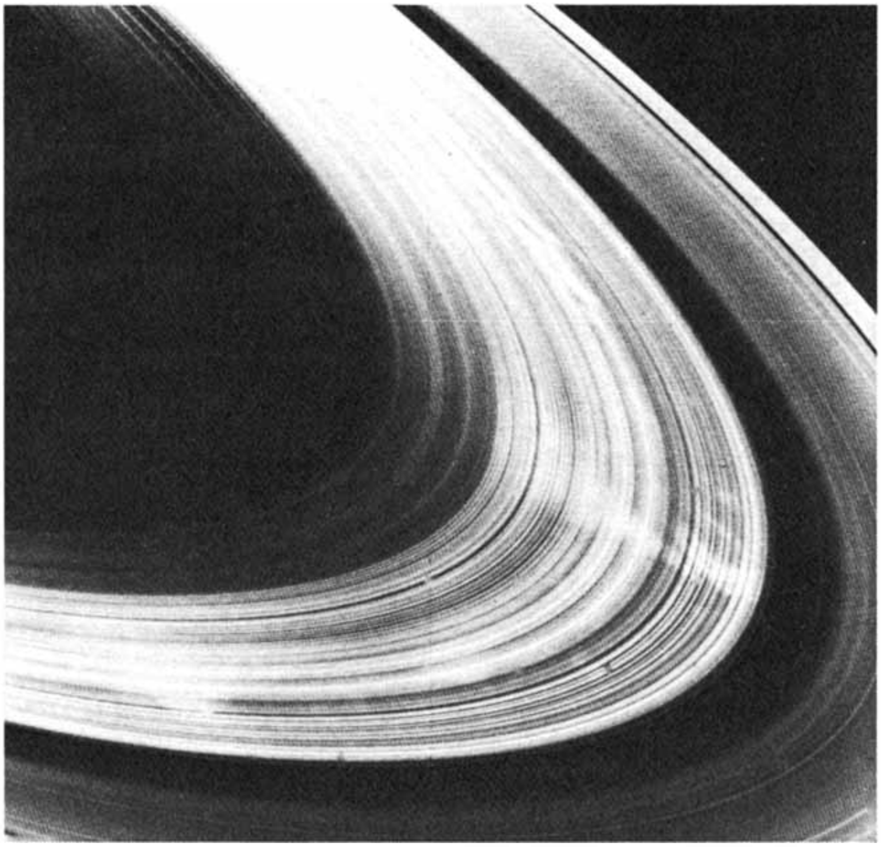
Observations with radar yield further deductions. In 1973 Richard M. Goldstein and Gregory Morris of the Jet Propulsion Laboratory of the California Institute of Technology probed the rings of Saturn with radar waves whose reflection they detected with the 210-foot antenna of the deep-space network at Goldstone, Calif. The high reflectivity of the *A* and *B* rings implied that most of the particles in those two rings are at least comparable in size to the radar wavelengths of several centimeters that the investigators employed. If they had been smaller than the radar wavelengths, they would have been transparent to the radar waves. If they had been much larger than the wavelengths, their emission of thermal radiation at those wavelengths would have been significant. The low level of such radiation limits their size to no more than a few meters.

Data from the Voyager spacecraft have supported and extended the earlier findings. In one type of experiment radio waves from the spacecraft were sent through the rings to the earth and measurements were made of the amount of power scattered by the ring particles at various angles of deflection from the initial path of the waves. As the size of the particles increases with respect to the wavelength, the scattering pattern becomes more narrowly concentrated within small angles of forward deflection. An analysis of the Voyager data by G. Leonard Tyler and Ahmed Essam A. Marouf of Stanford University indicates that the largest abundant particles in the *A, B* and *C* rings are about 10 meters in size. More abundant particles are as small as 10 centimeters, and regional variations in the distribution of sizes are found among the rings.

Just as the scattering of radar waves by the particles in the rings makes it pos-



RING SYSTEMS around Jupiter, Saturn and Uranus are diagrammed at scales on which the radius of each planet is set equal to 1. Jupiter (*left*) has a "bright" ring that is actually quite faint and almost transparent. An even fainter disk of particles extends inward from the ring, perhaps into the atmosphere of the planet. A halo of particles (not shown) gives the system a vertical thickness of some 20,000 kilometers. The system was discovered by spacecraft. Saturn (*center*) has seven rings. The *A* and *B* rings are separated by Cassini's division; the *A* ring includes Encke's division. Letters were assigned to the rings in the order of their discovery. Only the main rings (the *A, B* and *C* rings) are readily visible in earth-based telescopes; the rest (except the *E* ring) were discovered by spacecraft. Uranus (*right*) has no fewer than nine rings. Here their width is exaggerated. They were detected from the earth, and they are designated by numbers or Greek letters. The moons in the ring systems are labeled at the right of each panel; their orbits are in broken lines. The relative sizes of Jupiter, Saturn and Uranus are shown above.



SPOKES AND BANDS in the rings of Saturn were first photographed by *Voyager 1*. The spokes arise sporadically in the *B* ring; they are more or less radial wedges. Each spoke is bright in images of the sunlit face of the rings made when the spacecraft and the sun are on opposite sides of Saturn, so that sunlight reflects forward from the ring particles to the spacecraft (*top photograph*), and each spoke is dark in images made when the spacecraft and the sun are on the same side of Saturn, so that sunlight reflects backward (*bottom photograph*). In both images the banding of the rings into ringlets is evident. The wide dark band is Cassini's division.

sible to detect particles on the order of the size of a radar wavelength, so the scattering of sunlight makes it possible to detect particles the size of a wavelength of visible light. In particular, the strong brightening of a segment of ring when it is viewed at small forward scattering angles implies that particles on the order of a micrometer in size are abundant in the segment. Such an observation can be made only when Saturn comes between the sun and the observer. For observations from the earth this condition cannot be met, but for observations from spacecraft it can be. Thus studies of the Voyager data indicate that particles on the order of a micrometer in size constitute a large fraction of the particles in the *F* ring, a significant fraction in many parts of the *B* ring and a less significant fraction in the outer part of the *A* ring. On the other hand, the *C* ring and Cassini's division show no sign of such small particles.

Structure in Saturn's Rings

Before the passage of the Voyager spacecraft near Saturn a limited amount of structure was recognized in the planet's rings. Cassini's division was of course known, and so was Encke's division, a narrower band in the outer part of the *A* ring. The high-resolution photographs of the rings made by the Voyager spacecraft revealed a number of surprises. Narrow annular regions of differing brightness and opacity appeared, seemingly as numerous as the grooves on a phonograph record. In addition deviations from circularity were found. These include radially oriented, wedge-shaped spokes in the *B* ring and knots, braids and twists in the *F* ring.

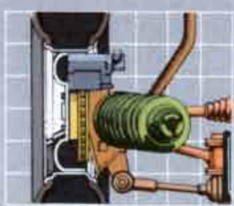
The greatest amount of detailed structure is exhibited by the *B* ring, which also has the greatest density of particles found anywhere in the rings. Variations in the opacity of the *B* ring occur over radial distances as small as 10 to 50 kilometers. In contrast, the *B* ring has an overall width of 25,000 kilometers. The central, most opaque part of the *B* ring is where the spokes appear. Typically each spoke can be seen for a significant part of the 10 hours it takes a parcel of the *B* ring to complete one orbital revolution. Meanwhile new spokes are arising sporadically in new locations in the ring. Compared with their surroundings the spokes appear bright in forward-scattered light and dark in backscattered light. Hence particles on the order of a micrometer in size are particularly abundant in the spokes.

Each part of a spoke orbits Saturn at the same velocity as that of the ring particles at its radial distance. The inner sections move faster; thus a spoke becomes more tilted with time. Eventually it disappears. The narrow end of each wedge-shaped spoke seems to coincide approximately with the distance from

FORD ESCORT WAGON

The best-selling wagon in America.

Escort Wagon came on the scene a year ago, and promptly became America's favorite wagon.* Maybe that's because it's the only American-built



wagon with the traction of front-wheel drive and smooth four-wheel independent suspension.† Or maybe it's because of the great job it does in carrying, commuting and loading.

Capability car.

With the rear seat up, Escort Wagon offers 28 cubic feet of usable cargo room, and a color-keyed cargo cover to conceal your valuables. With the rear seat down, Escort's already considerable volume nearly doubles, to 57.8 cubic feet. Based on EPA Cargo Volume Index. Impressive. But don't think numbers, think of the errands run, the cargo carried. And Escort Wagon was designed to help keep driving costs down with low scheduled maintenance.

Engineering means comfort, too.

Escort Wagon works just as hard at keeping you comfortable. You'll find handsome interiors with optional low-back bucket seats. And four-wheel independent suspension helps smooth your ride because each wheel reacts to bumps independently, with minimal effect on the other wheels or you.

At the heart of the World Car Wagon, the CVH engine.

Escort Wagon's powerplant is just as much an engineering triumph as Escort itself. It's the CVH engine. This Compound Valve Hemispherical head engine and standard 4-speed transmission with fourth gear overdrive helps give you power and excellent economy.

44 EST HWY **28** EPA EST MPG

For comparison. Your mileage may differ depending on speed, distance and weather. Actual highway mileage lower.

And Escort Wagon offers the convenience of Ford's patented split-torque automatic transmission option. Whether you buy or lease, see Escort Wagon at your Ford Dealer now.

FORD ESCORT

FORD DIVISION

*Based on R. L. Polk & Co. registrations for 1981.

†Excludes other Ford Motor Company products.

LOOK OUT WORLD HERE COMES FORD

The $149[95] personal computer.

Introducing the Sinclair ZX81

If you're ever going to buy a personal computer, now is the time to do it.

The new Sinclair ZX81 is the most powerful, yet easy-to-use computer ever offered for anywhere near the price: only $149.95* completely assembled.

Don't let the price fool you. The ZX81 has just about everything you could ask for in a personal computer.

A breakthrough in personal computers

The ZX81 is a major advance over the original Sinclair ZX80—the world's largest selling personal computer and the first for under $200.

In fact, the ZX81's new 8K Extended BASIC offers features found only on computers costing two or three times as much.

Just look at what you get:

- Continuous display, including moving graphics
- Multi-dimensional string and numerical arrays
- Mathematical and scientific functions accurate to 8 decimal places
- Unique one-touch entry of key words like PRINT, RUN and LIST
- Automatic syntax error detection and easy editing
- Randomize function useful for both games and serious applications
- Built-in interface for ZX Printer
- 1K of memory expandable to 16K

The ZX81 is also very convenient to use. It hooks up to any television set to produce a clear 32-column by 24-line display. And you can use a regular cassette recorder to store and recall programs by name.

*Plus shipping and handling. Price includes connectors for TV and cassette, AC adaptor, and FREE manual.

If you already own a ZX80

The 8K Extended BASIC chip used in the ZX81 is available as a plug-in replacement for your ZX80 for only $39.95, plus shipping and handling—complete with new keyboard overlay and the ZX81 manual.

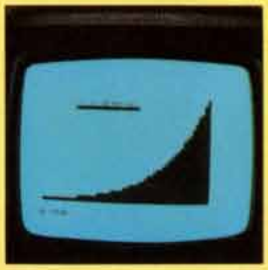
So in just a few minutes, with no special skills or tools required, you can upgrade your ZX80 to have all the powerful features of the ZX81. (You'll have everything except continuous display, but you can still use the PAUSE and SCROLL commands to get moving graphics.)

With the 8K BASIC chip, your ZX80 will also be equipped to use the ZX Printer and Sinclair software.

Warranty and Service Program**

The Sinclair ZX81 is covered by a 10-day money-back guarantee and a limited 90-day warranty that includes free parts and labor through our national service-by-mail facilities.

**Does not apply to ZX81 kits.



NEW SOFTWARE: Sinclair has published pre-recorded programs on cassettes for your ZX81, or ZX80 with 8K BASIC. We're constantly coming out with new programs, so we'll send you our latest software catalog with your computer.

ZX PRINTER: The Sinclair ZX Printer will work with your ZX81, or ZX80 with 8K BASIC. It will be available in the near future and will cost less than $100.

16K MEMORY MODULE: Like any powerful, full fledged computer, the ZX81 is expandable. Sinclair's 16K memory module plugs right onto the back of your ZX81 (or ZX80, with or without 8K BASIC). Cost is $99.95, plus shipping and handling.

ZX81 MANUAL: The ZX81 comes with a comprehensive 164-page programming guide and operating manual designed for both beginners and experienced computer users. A $10.95 value, it's yours free with the ZX81.

The $99^{95} personal computer.

Introducing the ZX81 kit

If you really want to save money, and you enjoy building electronic kits, you can order the ZX81 in kit form for the incredible price of just $99.95.* It's the same, full-featured computer, only you put it together yourself. We'll send complete, easy-to-follow instructions on how you can assemble your ZX81 in just a few hours. All you have to supply is the soldering iron.

How to order

Sinclair Research is the world's largest manufacturer of personal computers.

The ZX81 represents the latest technology in microelectronics, and it picks up right where the ZX80 left off. Thousands are selling every week.

We urge you to place your order for the new ZX81 today. The sooner you order, the sooner you can start enjoying your own computer.

To order, simply call our toll free number, and use your MasterCard or VISA.

To order by mail, please use the coupon. And send your check or money order. We regret that we cannot accept purchase orders or C.O.D.'s.

CALL 800-543-3000. Ask for operator #509. In Ohio call 800-582-1364. In Canada call 513-729-4300. Ask for operator #509. Phones open 24 hours a day, 7 days a week. Have your MasterCard or VISA ready.

These numbers are for orders only. For information, you must write to Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

sinclair

AD CODE SA11	PRICE†	QTY.	AMOUNT
ZX81	$149.95		
ZX81 Kit	99.95		
8K BASIC chip (for ZX80)	39.95		
16K Memory Module (for ZX81 or ZX80)	99.95		
Shipping and Handling	4.95		$4.95
To ship outside USA add $10.00			
		TOTAL	

MAIL TO: Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

NAME ______________________

ADDRESS ______________________

CITY/STATE/ZIP ______________________

† U.S. Dollars

Saturn at which the period of an orbiting particle matches the period of Saturn's rotation. The magnetic field of Saturn is locked into the planet and therefore rotates with it. Hence electromagnetic forces may be partly responsible for the spokes. In this regard it may be notable that bursts of broad-band static were detected by the Planetary Radio Astronomy experiment aboard *Voyager 1*. The bursts appear to have originated from sources in the *B* ring near the regions of intense spoke activity.

In comparison with the *B* ring the *A* ring is featureless and uniformly bright. It does, however, exhibit a variety of narrow, discrete features in its outermost regions. For one thing the dark and relatively empty band of Encke's division, which has a width of about 350 kilometers, contains two or three narrow, irregular ringlets composed of tiny particles. Images made by *Voyager 2* show that the ringlets have kinks reminiscent of those detected in the *F* ring by *Voyager 1*. Four unusual bands with irregular, wavelike structure bracket Encke's division. At a greater distance from Saturn the outer edge of the *A* ring shows many narrow bands about 20 kilometers wide in which small particles are plentiful. The spacing between the bands decreases outward. The outer edge of the *A* ring is quite abrupt. Here too small particles are plentiful.

The *C* ring and Cassini's division have many structural similarities in addition to their relative lack of redness, their similar degree of transparency and their dearth of small particles. Both of them exhibit discrete, regularly spaced bands of uniform brightness. Both of them also contain a variety of narrow, sharp-edged, completely empty gaps with a radial width of from 50 to 350 kilometers. Furthermore, some of the gaps contain even narrower, equally sharp-edged ringlets that are quite opaque. Several such ringlets are eccentric (that is, noncircular) and nonuniform in width. Images from *Voyager 2* show that the opaque ringlets often differ in color from the surrounding ring material. They more closely resemble the stuff of the *A* and *B* rings.

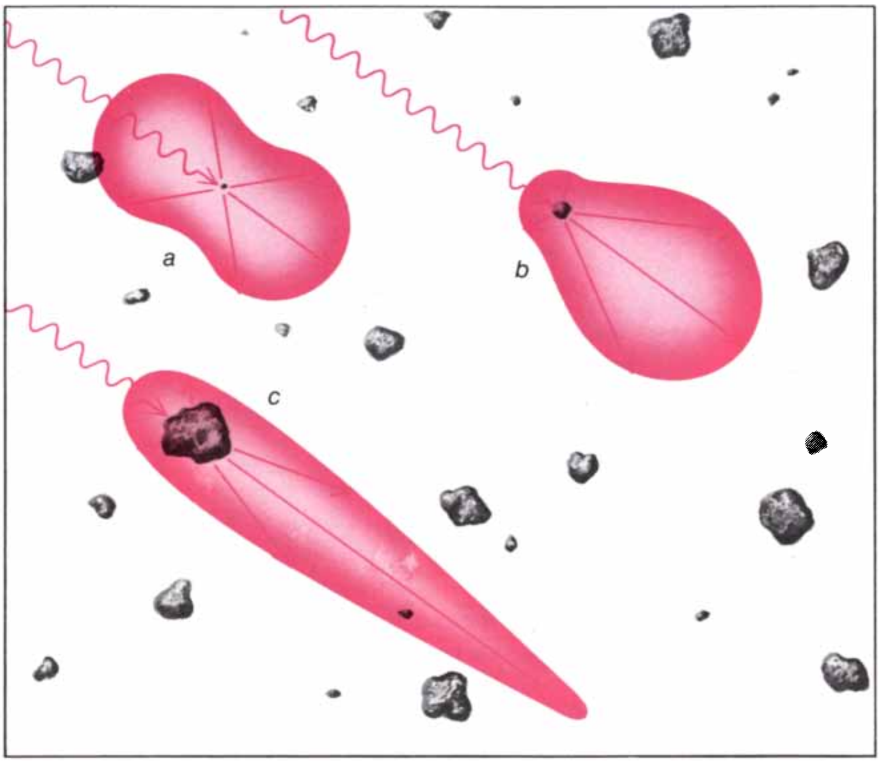
Rings of Uranus and Jupiter

The rings of Uranus are in many ways the complement of the rings of Saturn. The main rings of Saturn (the *A, B* and *C* rings) are broad, and punctuated by narrow gaps. The rings of Uranus are narrow (they each measure from several kilometers to 100 kilometers in radial extent) and are separated by broad, empty regions. Nine rings around Uranus have been confirmed. Each has been assigned a number or a Greek letter. The technique of stellar occultation exploited to study the rings capitalizes on the minute apparent size of the occulted star to yield a high spatial resolution. Nevertheless, several of the rings of Uranus remain unresolved. Such rings are less than five kilometers in radial extent. (The technique of stellar occultation, as applied to the rings of Saturn by *Voyager 2,* reveals structure on scales as fine as hundreds of meters.) The narrowest rings of Uranus (the ones designated γ, δ and η) are nearly circular and lie in a common plane. In contrast, the somewhat broader rings α, β and 4 tend to be slightly elliptical and inclined to the common plane. The ϵ ring is the broadest and most elliptical of the nine. Its radial width varies linearly with distance from Uranus, beginning at a width of 20 kilometers where the ring is closest to the planet and ending at a width of 100 kilometers where the ring is farthest away. A few of the narrow eccentric ringlets in Saturn's rings are now known to exhibit a similar variation.

The dimming of the brightness of stars as they pass behind the positions of the rings of Uranus indicates that the rings have an opacity comparable to that of the more opaque (and therefore brighter) parts of Saturn's rings. The ϵ ring is particularly opaque. Curiously, it is most opaque at its edges. The gravitational forces arising from the oblate shape of Uranus cause the orientation of the elliptical rings, including the ϵ ring, to change continuously. A cycle in which the elliptical shape of the ring rotates once about the planet takes some 6,300 hours. In that time a particle in the ring orbits Uranus some 750 times. Hence two things affect the ϵ ring. First, its orientation changes. Second, the particles in the ring at different radial distances from Uranus orbit the planet at somewhat different velocities, in accord with the physics of orbital motion. Still, the ϵ ring maintains its integrity. That is, its eccentricity, its variable width and the concentration of particles toward its edges all persist.

What composes the rings of Uranus? Because of their narrow width the rings are hard to observe from the earth. Nevertheless, recent observations imply that the ring particles are quite dark. Thus water ice is not their dominant component. Quite likely the ring particles orbiting Uranus are made of silicates rich in compounds that absorb sunlight. Certain iron oxides and complex carbon compounds have the required characteristics. Unfortunately nothing is known about the size of the particles.

The rings of Jupiter consist of three main parts: a bright ring, a diffuse disk and a halo. The bright ring has a width



SCATTERING OF SUNLIGHT or some other form of electromagnetic radiation by the particles in a ring allows the deduction of the size of the particles abundant in the ring. In particular, if a particle is less than about a tenth the size of a wavelength of the incoming radiation, it scatters the radiation almost equally in all directions (*a*). If a particle is somewhat smaller than the wavelength, it tends to scatter the radiation forward (*b*). If a particle is larger than the wavelength, it scatters nearly all the radiation forward (*c*). The lengths of the arrows represent the relative amounts of energy scattered at various angles. The observation that particles in the spokes in the *B* ring of Saturn scatter sunlight mostly forward allows the inference that the spokes are local, transient concentrations of ring particles approximately a micrometer in size.

RUN AWAY

. . . with us to the most intriguing islands on earth . . . through the pages of a fabulous new magazine for the escapist in us all!

Enroll now as a Charter Subscriber at a special low rate . . . and get your copy of the magnificent Premiere Issue, just off the presses!

There are more than half a million islands on earth—and now, at last, there's a magazine devoted to exploring and learning about the most interesting of them. A magazine with the spirit of an adventurer, the soul of a poet, the eye of a scientist and the curiosity of a social anthropologist. A magazine of surpassing beauty, illustrated throughout in full color, bound with a hard spine for durability and the longevity it deserves.

It's called ISLANDS, and if you've been looking for the island of your dreams—to think about or even to visit—it's the logical place to start. Because part of our job will be to help you keep your island fantasies alive and well, by introducing you to islands fit for the visions of the most incurable romantic.

But our enthusiasm for islands goes well beyond the stuff of dreams. We're interested in how islands are formed, altered and destroyed by awesome natural forces. In the remarkable ecosystems that develop in isolation islands afford. In island life forms, so often found nowhere else.

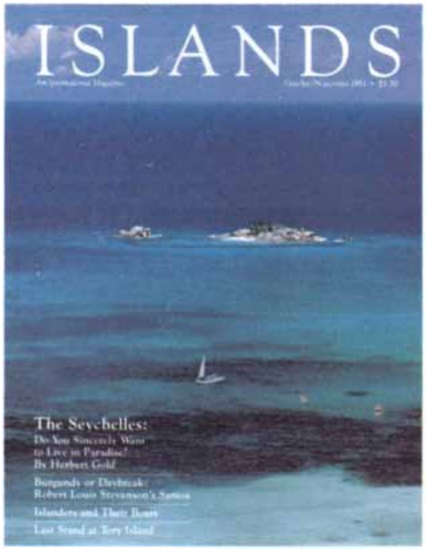
Many island civilizations and cultures have formed and grown in isolation too, with fascinating—and occasionally bizarre—results. We'll tell you about them, trace their histories and evolution, bring you up to date on the life and lifestyles of island societies today.

And we'll range far and wide across the earth to search out and report to you the most interesting island stories, past and present. Stories like these . . .

Do You Sincerely Want To Live in Paradise? (The Extraordinary Seychelles)

Islanders and Their Boats: The Boat Culture of the South Pacific.

The Resettlement of Tristan da Cunha, World's Loneliest Island.



The Wild Horses of Chincoteague—and the Wilder Horses (and Deadly Sand Bars) of Sable.

Burgundy or Daybreak: Robert Louis Stevenson's Tragicomic Last Days in Samoa.

Peril in the Galapagos: The Life Forms that Inspired Darwin Are in Danger.

The Curious Accident that Created Madeira Wine—and Put Madeira on the Map.

The Best-Attended Island Birth: 1963, When Volcanic Surtsey Rose Out of the Sea South of Iceland.

The Goat Men of Aguijan.

The Explosion of Krakatoa—and the Birth of Anak Krakatoa (Child of Krakatoa).

Coconuts: Everything You Wanted To Know (and More).

Tom Sawyer's Island.

Heyerdahl's Search for the Origins of South Pacific Islanders.

Seamounts and Abyssal Hills—Islands of the Future?

Unraveling the Archaeological Mysteries of Easter Island.

Plus articles and features on island travel and tour opportunities. We'll tell you how to lay claim to an island; about scientific expeditions to islands, and what they're discovering; islands to buy or lease; the best books, films, TV and art about islands.

Charter Subscription Offer

If you're as fascinated by islands as the rest of the human race has always been . . . if one corner of your imagination has been set aside for the island of your dreams . . . if you're curious about islands and like the idea of exploring them and experiencing them through the pages of a magnificent new magazine . . . ISLANDS is for you.

To become a Charter Subscriber—and get your copy of the Premiere Issue, just off the presses—fill in this coupon, clip it out and return it to ISLANDS Magazine, 123 West Padre, Santa Barbara, CA 93105. **Send no money now.** We'll enroll you as a Charter Subscriber, send your Premiere Issue and bill you at the Charter rate of only $15 for the year (five additional bimonthly issues)—a discount of $3 off the regular subscription price. You reserve the right to cancel at any time and receive an immediate, unquestioned refund covering the balance of your subscription. To avoid disappointment, act promptly. Quantities of the premiere edition are limited.

YES . . . please enroll me as a Charter Subscriber to ISLANDS, send me the Premiere Issue and bill me at the low Charter rate of just $15 for the year (five additional bimonthly issues). I understand that I may cancel at any time and receive an immediate, unquestioned refund covering the balance of my subscription.

Name (please print) ____________

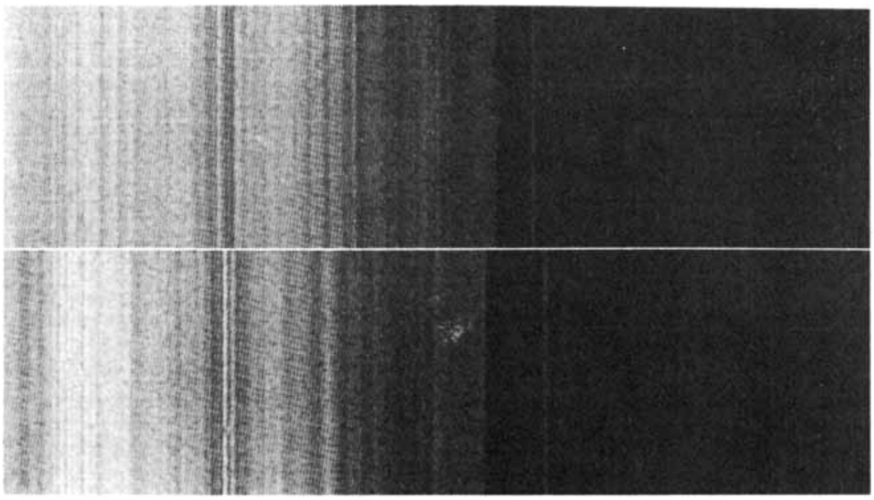
Address ____________

City ________ State ______ Zip ______

ISLANDS
123 West Padre
Santa Barbara, CA 93105
2SA1

of about 6,000 kilometers. Its fairly sharp outer boundary lies about 58,000 kilometers, or .8 Jovian radius, above the surface of Jupiter. In the outer part of the ring a narrow band some 600 kilometers wide is about 10 percent brighter than the rest. Still, the opacity of the "bright" ring is so low that only .001 percent of the sunlight passing through it is intercepted by its particles. The diffuse disk is several times fainter. It extends inward from the inner margin of the bright ring. Indeed, it may intersect the atmosphere of Jupiter. Viewed edge on, the bright ring and the disk appear to be confined primarily to a thickness of no more than 30 kilometers. Remarkably, however, the halo has a vertical extent of some 20,000 kilometers. The halo is highest over the diffuse disk. Its outer boundary extends a little beyond the outer edge of the bright ring.

On the basis of the way the bright ring brightens at small scattering angles it has been found to contain particles whose characteristic size is several micrometers. Such particles are quite ineffective at absorbing high-energy protons and electrons. Hence these particles cannot be responsible for the decrease in the flux of high-energy particles detected in the bright ring by the *Pioneer 10* spacecraft. There must also be particles at least a centimeter in diameter. The particles in the ring are reddish, which makes them the color of many asteroids and moons of the outer solar system.



DISTORTIONS IN THE RINGS of Saturn are evident in a comparison of images made on opposite sides of the rings by *Voyager 2.* In each image the *B* ring is the bright region at the left; Cassini's division is the dark region at the right. The outer edge of the *B* ring is shown to be noncircular, and the presence of a bright eccentric ringlet outside the edge of the *B* ring is confirmed. (The ringlet was discovered by *Voyager 1.*) Particles near the outer edge of the *B* ring orbit Saturn twice for each single orbit of the moon Mimas. This resonance greatly amplifies the gravitational effect of Mimas and thereby distorts the *B* ring; overall the resonance raises bulges on opposite sides of the ring. In contrast, the eccentric ringlet has a single maximum that may result from the gravitational field of a local moonlet that has not yet been detected.

Collisions in the Rings

The architecture of a ring system results from the interplay of a number of forces. These include gravitational forces due to moons outside the rings and the moonlets embedded in them, electromagnetic forces due to the planet's rotating magnetic field and even the gentle forces exerted by the dilute gaseous medium in which the rings rotate.

All the particles in a ring system share a common orbital motion around a planet: they travel in the direction of the planet's rotation. The vertical and radial motions superimposed on the orbital motion of each such particle have no similar constraint. Hence neighboring particles move randomly in these directions with respect to one another, and collisions are inevitable. When the random relative velocities are large, as they might have been if the rings were ever a thick cloud of particles, the collisions are violent, and even if the collisions are rare, a great amount of the energy of relative motion goes into heating the particles and deforming their structure. The consequent loss of energy means that the random velocities rapidly decrease. The decrease in the vertical component of the velocities leads to a flattening of the ring system. Meanwhile the decrease in the radial component leads to more nearly circular orbits. In brief, a fat ring becomes a thin and roughly circular disk quite early in its history.

Even when the ring particles have lost almost all their random motion, the collisions continue. The reason is as follows. The force of gravity exerted by the planet on the particles in a ring weakens with increasing distance from the planet, so that ring particles at greater distances take longer to circle the planet. Thus a ring particle whose orbit is slightly inside that of a second ring particle eventually catches up with it, and the two of them collide if their radial separation is less than the diameter of a particle.

The collision is likely to happen at a relative velocity of less than a centimeter per second. Nevertheless, it can convert a little of the particles' circular orbital motion into random vertical motion. Subsequent collisions will prevent the particles' vertical velocities from becoming very great. A steady state will therefore be reached that determines the thickness of the ring. If the particles have a wide range of sizes, the smaller particles will gain vertical velocity mostly by being gravitationally deflected in near-miss encounters with the bigger particles. They will lose their vertical velocity mostly by colliding with other small particles. Under these circumstances the small particles will attain a vertical extent several times the size of the biggest abundant particles. In the case of the *A* and *B* rings of Saturn the small particles would be expected to occupy a vertical thickness of 10 to 100 meters. Measurements made by experiments on the Voyager spacecraft indicate that the main rings of Saturn are certainly no more than a few hundred meters thick. The range of the measurements makes the observational thickness compatible with the predicted thickness.

The collisions of neighboring particles will also convert some of their circular orbital motion into radial motion. Hence the rings will spread radially. An isolated, unconstrained ring will spread until its particles are far enough apart for their collisions to virtually cease. The bright ring of Jupiter may have attained this end state. The width and the very low opacity of the bright ring may reflect the collisionally induced spreading of its larger particles over the age of the solar system.

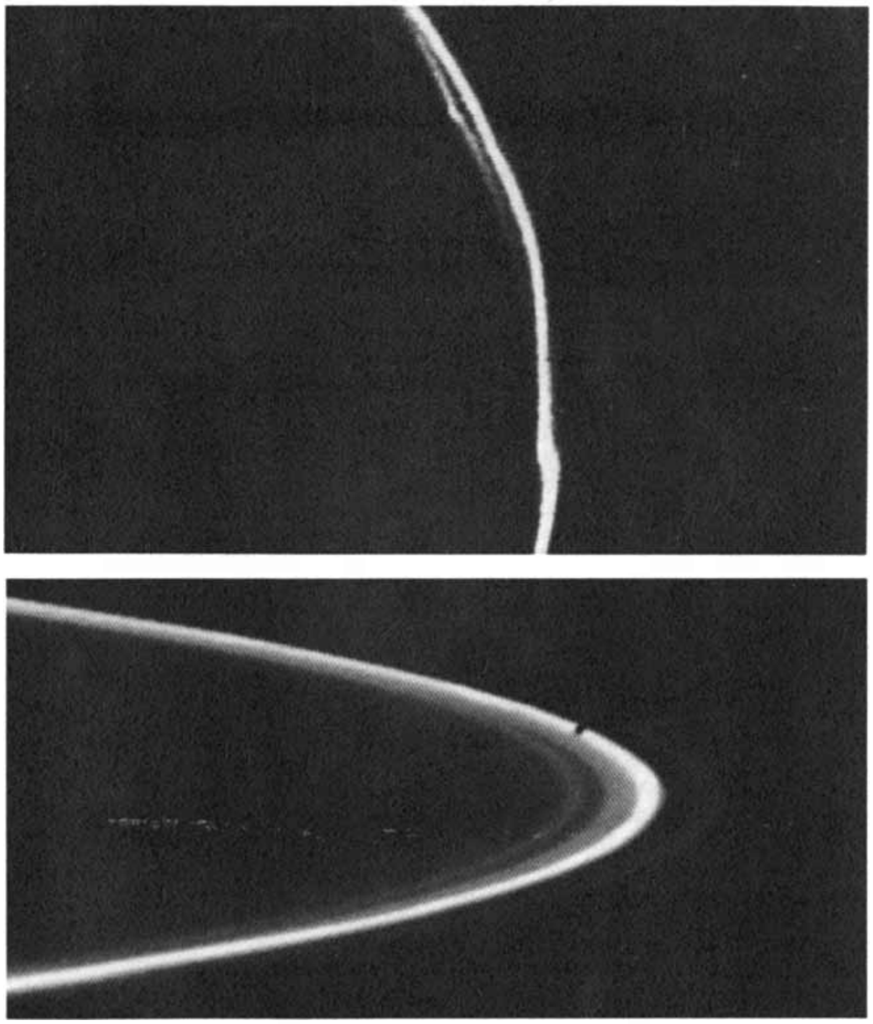
The rings of Saturn and Uranus, however, display abrupt, well-defined boundaries that limit regions densely filled with particles. Other processes must therefore be counteracting the rapid spreading induced by frequent collisions. An important role in these processes may be played by moonlets embedded in the rings or adjacent to them. The gravitational fields of larger, more distant moons may serve to lock some of the local moonlets into fixed orbits and in that way prevent the moonlets from being dislocated by their gravitational interactions with the ring particles around them.

Fundamentally the motion of bodies in orbit around a much more massive planet is dominated by the planet's gravitational field. In certain instances, however, the gravitational attraction between two relatively tiny orbiting bodies may be amplified and thus may come to affect their motion significantly. Such amplification is known as resonant forcing or simply as a resonance. Consider that if the orbital period of a moon is an exact multiple or fraction of that of another moon, the net gravitational effect of each moon on the other is in essence a push or a pull applied repeatedly at the same point in the cyclical motion. Thus the effect is amplified. In some instances resonant forcing locks pairs of moons into orbits whose periods maintain a fixed ratio of two small integers. Such commensurabilities are well known in the Jovian and Saturnian satellite systems, where they involve at least four pairs of the major moons and probably several of the newly discovered moons as well.

Resonance and the Rings

A rather different situation applies to a resonance in a disk of particles. Close to the radial distance from the planet at which the particles in the disk would have an orbital period commensurate with that of one of the planet's moons, the amplification of the gravitational effect of the moon over long periods of time causes the orbits of the particles to become noncircular. Thus the particles are made likely to collide with their less perturbed neighbors. As a result the particles are lost from a band at the radial distance of a resonance. Typically the band has a natural width of some tens of kilometers. Examples of such bands in the rings of Saturn probably include the dozens of narrow gaps in the outer part of the *A* ring, which seem to have resulted from resonances caused by the newly discovered moons designated 1980S1, 1980S3, 1980S26 and 1980S27. The predominance of small particles near the gaps probably attests to the violent collisions induced locally by each resonance.

One additional effect of the resonances that are distant from a moonlet should also be mentioned here. In 1978 Peter M. Goldreich of Cal Tech and Scott D. Tremaine, now at the Institute for Advanced Study in Princeton, proposed that spiral waves of fluctuation in the density of the ring material are present in the rings of Saturn. It had been suggested earlier that similar waves are responsible for the spiral-arm pattern of galaxies such as the Milky Way. In images made by *Voyager 1* spiral density waves seem to be faintly visible in Cassini's division; their tightly wound pattern is reminiscent of a watch spring. It is thought the waves are raised there by resonances. The waves may convey the effects of the resonances over great distances; hence the waves may turn out to play a central role in transporting material within the disk formed by Saturn's rings.



BRAIDS IN THE "F" RING of Saturn were revealed in photographs made by the *Voyager 1* spacecraft when it passed near the planet in November, 1980, but when *Voyager 2* arrived nine months later, they were gone. The ring itself is some 80,000 kilometers (1.3 Saturnian radii) from the surface of Saturn and 4,000 kilometers from the outer edge of the *A* ring. In November, 1980 (*top*), it consisted of three strands, each some 30 kilometers wide. The outer two strands exhibited kinks, warps and knots, and they seemed to be braided, or at least to intersect. The gravitational fields of the moons designated 1980S26 and 1980S27 may account for such structural distortion. By August, 1981 (*bottom*), the *F* ring had changed its structure substantially. An unbraided strand was dominant; three fainter strands appeared as companions.

The resonances created in a disk by a moon are spaced closer together as the orbit of the moon is approached. At some critical distance the radial spacing between successive resonances becomes equal to the natural width of each resonance. Within this critical distance the resonances overlap. The result is a continuous zone of transfer of ring material that clears the ring material away from the orbit of the moon. The width of the zone and the degree of clearing in it will depend on the mass of the moon and the density of the surrounding ring material. The more massive the moon, the wider the zone, but the denser the packing of ring particles, the more often it will happen that collisions between particles will propel some of the particles back into the zone. Jack Lissauer and Frank H. Shu of the University of California at Berkeley and M. Hénon of the Nice Observatory have shown that if moonlets from several kilometers to tens of kilometers in size were embedded in the rings of Saturn, the moonlets could cause much of the fine structure in the rings. As of this writing, however, no moonlets have been detected in even the most likely locations in the images made by *Voyager 1* and *Voyager 2*.

If the moon is adjacent to a ring, still another effect is possible: the overlapping resonances that surround the orbit of the moon can prevent the ring from spreading and give the ring a sharp boundary. The outer edge of Saturn's *A* ring is probably maintained in this way by the moons 1980S26 and 1980S27. The ring in turn repels the moons. As it happens 1980S26 is quite close to a radial distance from Saturn that would trap it in a resonance with the much more massive moon Mimas or Tethys. Such a resonance would "anchor" its radial distance, so that it could continue to sculpture the outer edge of the *A* ring. The only trouble with this hypothesis is that precise measurements seem to place

How to make better mistakes.

Everybody makes mistakes. Especially when they're typing.

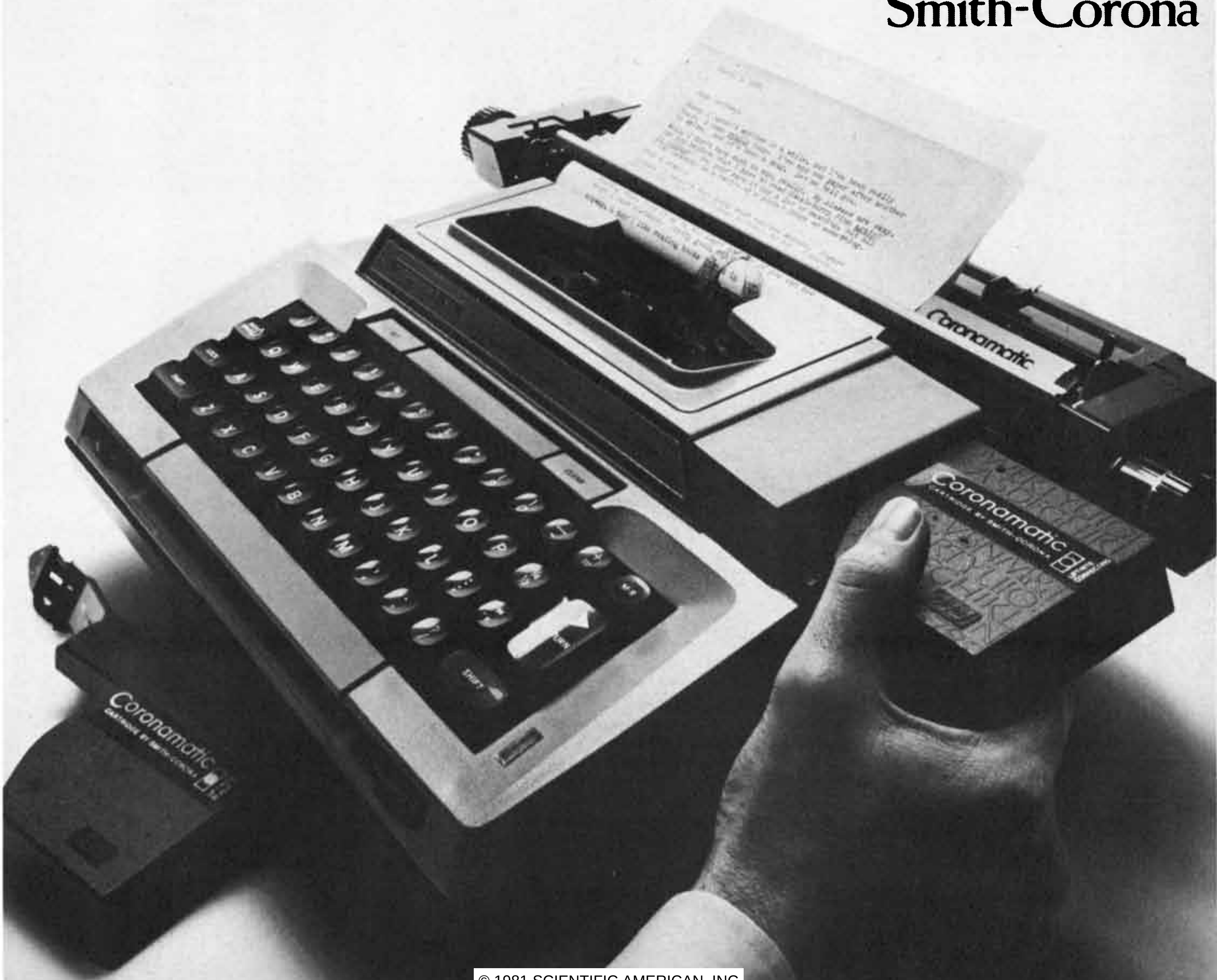
But the best typing mistakes are the ones that no one ever sees. And that's what Lift-Rite™ from Smith-Corona® is all about.

Lift-Rite is the correction cartridge that actually lifts mistakes right off the page. Without a trace of white paint or powder, because there isn't any. Mistakes disappear neatly, easily, smoothly, permanently, because they lift clean off the paper!

Compare a Smith-Corona with any other portable. Any one. Listen to it. Type on it. Examine it closely. Make no mistake.

It's Smith-Corona with Lift-Rite that should be your next portable typewriter.

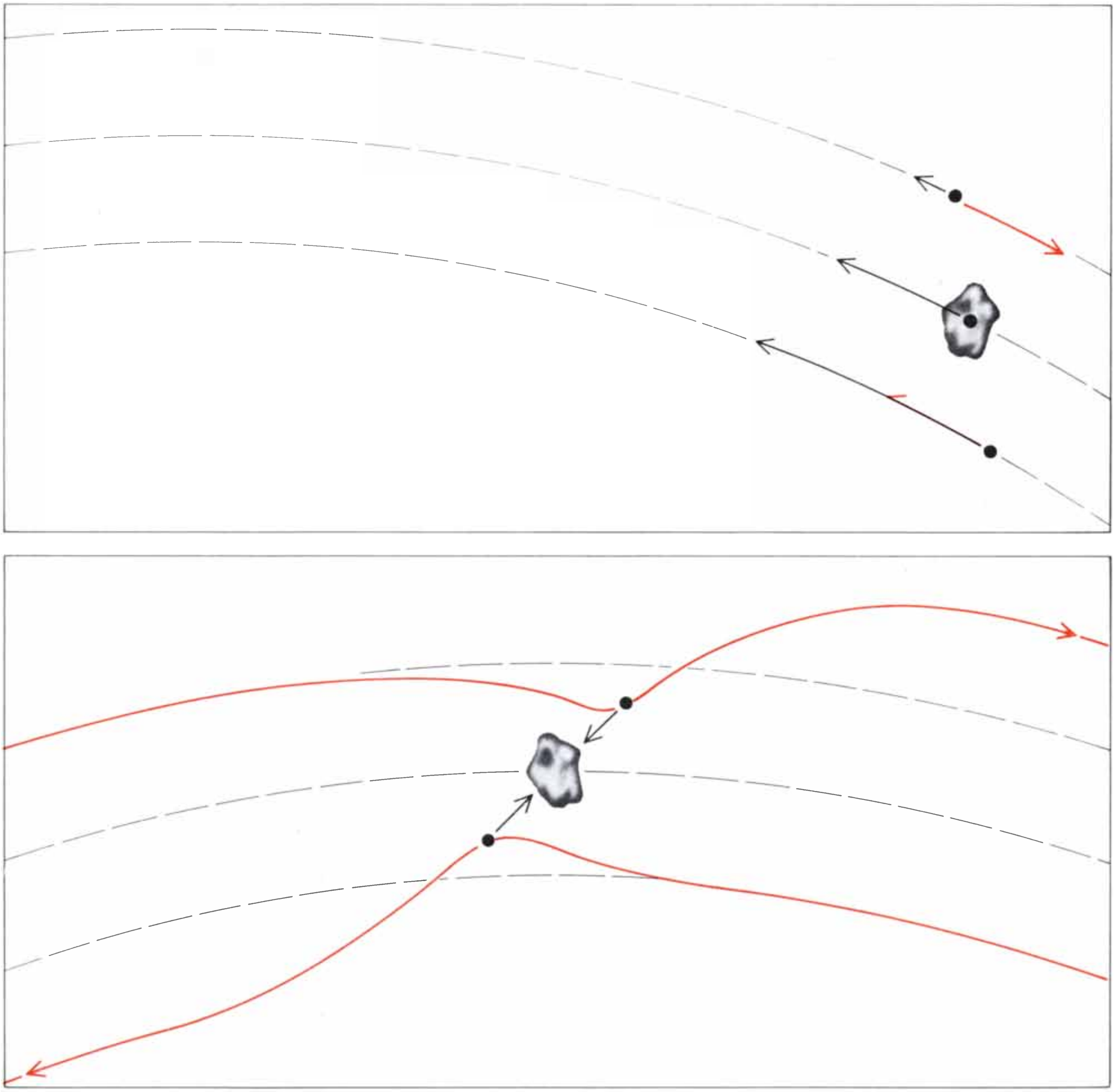
Smith-Corona



1980S26 slightly away from the anchoring distance.

Finally, two moons whose orbits are close together can prevent a narrow ring between their orbits from spreading radially. For example, 1980S26 and 1980S27, whose orbits are separated radially by only some 2,000 kilometers, may act as shepherds to confine the narrow, multistrand *F* ring between them. When one or both of the moons pass close to a given strand of the ring, their gravitational tug may generate distortions such as the ones seen in the images of the ring made by *Voyager 1*. Still, it is not clear why the *F* ring should consist of multiple narrow strands rather than a single somewhat broader strand, and why one of the strands recorded by *Voyager 1* was untwisted. In images made by *Voyager 2* the *F* ring exhibits no braids or other distortions, and the number of strands seems to have changed.

Whether shepherding moonlets produce smooth ringlets (such as the narrow, eccentric ringlets in Saturn's *C* ring and Cassini's division and probably the ε ring of Uranus) or kinky ringlets (such as those of Saturn's *F* ring and the ones in Encke's division that were found to be kinky in images made by *Voyager 2*) may depend on the ringlets' degree of opacity, or equivalently their particle density. In a ringlet that is rather transparent collisions between particles are relatively infrequent; hence the "echoes" of past



GRAVITATIONAL SHEPHERDING of particles by moonlets in the midst of a ring system may account for some of the banding in the rings of Saturn. The top drawing shows a moonlet and two representative particles in orbit around a planet. In accord with physical law the inner particle moves faster than the moonlet, which in turn moves faster than the outer particle (*black arrows*). Thus the moonlet is being overtaken by the inner particle at the same time as it in turn overtakes the outer particle (*colored arrows*). Each particle is drawn toward the moonlet by gravitational attraction; hence each particle is closer to the moonlet just after they are neck and neck in their orbits around the planet than it was just before. The bottom drawing shows the result. The net gravitational tug the moonlet exerts on the outer particle is in the direction of the outer particle's orbital motion, and so the outer particle is raised to a higher orbit. Conversely, the tug the moonlet exerts on the inner particle is opposite to the direction of orbital motion; the inner particle falls into a lower orbit. Ultimately the moonlet clears out a band surrounding its trajectory. The more massive the moonlet is, the wider the band will be.

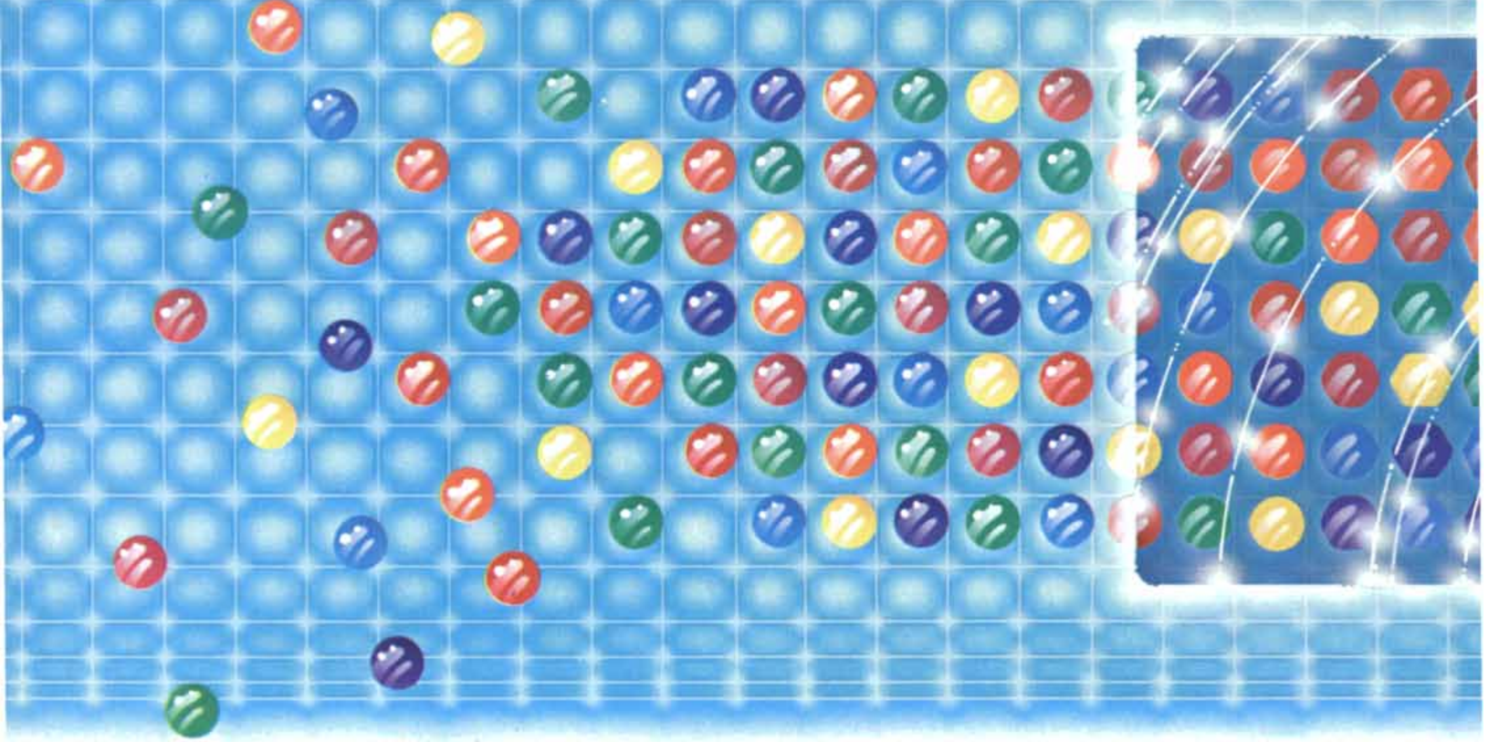
The power to tune in the world is placed instantly at your fingertips with Sony's ICF-2001 Worldband Radio.

Push a button and listen to opera from Germany, the news from Russia, the weather in England, or your favorite station in your own hometown.

You can even pre-set the memory buttons for six different stations on FM, AM, SW, or SSB/CW (Morse Code) broadcasts. And you can do it all without the skills and training of an astronaut.

The 2001 operates on AC, DC or car-battery power. So you can take the whole world with you, no matter where on Earth you go. **THE WORLDBAND RADIO**

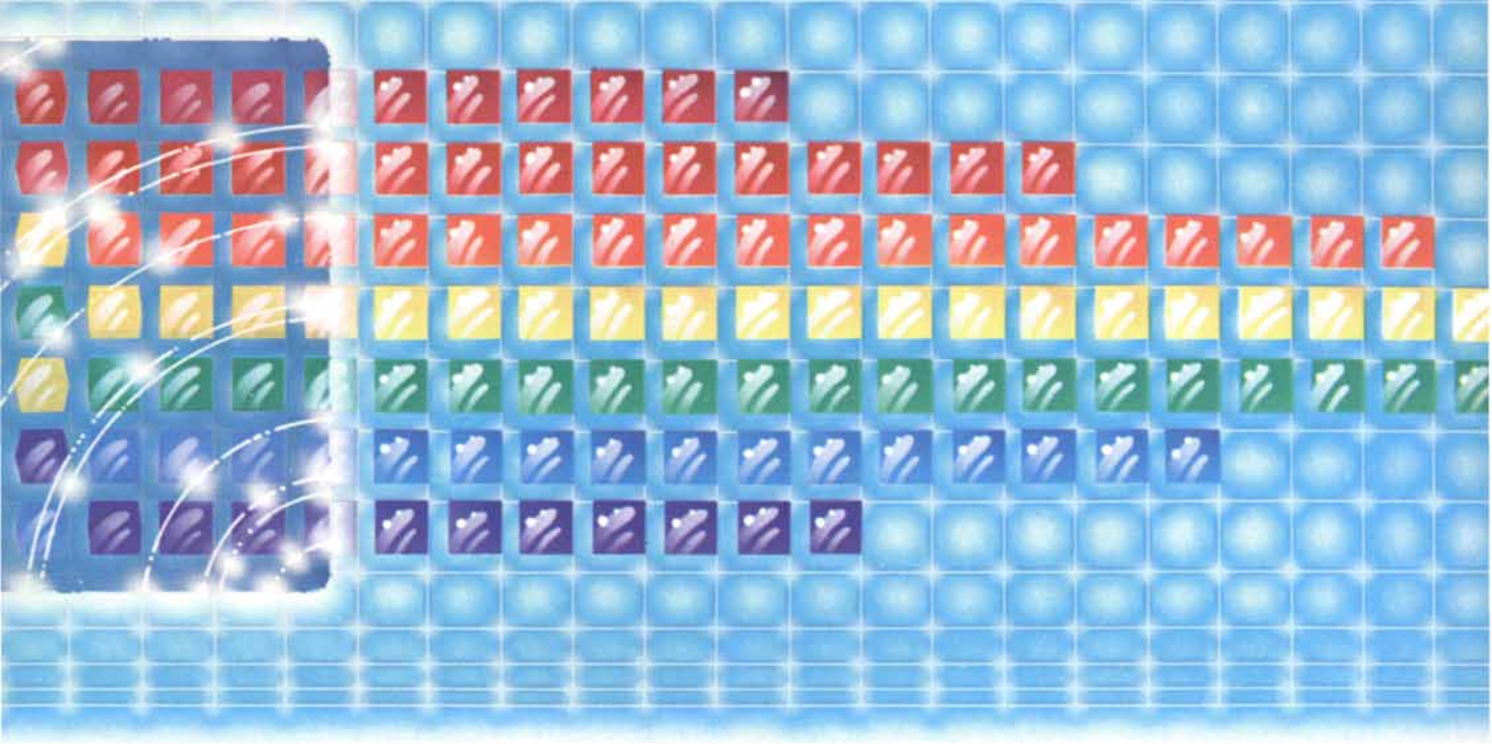
SONY®
THE ONE AND ONLY



Understanding information management.

We are no longer an industrial society.

More Americans are now in information jobs than in manufacturing and agriculture combined. Today's most pervasive technological trends have to do with information and its management. Yet to begin to understand the boundless promise of our new information society, we need to define its terms.



The management of information occurs in three distinct stages:

Organization

Information (words or data) is gathered and structured so that it can be put into a system. The business or institution is also organized to take advantage of the efficiencies of information management.

Processing

The information may then be stored, examined, organized, translated or directed by people or machines.

Transmission

Information is transported and made available to users as they select. Often information may be managed simply by organizing it and transmitting it to the proper user.

Present day systems to manage information are but one of the products of the knowledge business.

The Bell System constantly produces new knowledge, and turns this knowledge into practical advances. Electronic switching (organization, processing) and lightwave communications (transmission) were laboratory phenomena just yesterday. So were the integrated communications systems that ease the way for so many American businesses.

All are realities within the Bell network, which organizes, processes and transmits analog and digital information worldwide.

The ideas, applications and systems of the knowledge business are available to you now. The Bell System can help you use them now, to increase productivity and profits.

The knowledge business

Who moves more people than VW for less money? It's Ram Tough Dodge's new Mini Ram Wagon.

Compare Dodge Mini Ram 250 to VW Vanagon—passenger seating, gasoline mileage, fuel capacity and range, horsepower and rear loading area. Mini Ram gives you more.

Then compare standard features like power steering and factory-installed options like air conditioning and your choice of 6 sound systems: all available in Mini Ram. Not in Vanagon.

Now compare price. Mini Ram costs $1,325 less than Vanagon. For panel-side van buyers, there's a Dodge Van with the same performance, efficiency and interior dimensions as Mini Ram Wagon at a price VW hasn't seen in years—just $6,928.*

Underneath it all, Mini Ram is a Ram Tough Dodge Wagon. With the best rust protection of all wagons. There's 100% electrocoating, plus 370 sq. ft. of galvanized steel in critical areas . . . compared to VW's 4 sq. ft.

New Dodge Mini Ram Wagon. It does more than VW Vanagon and it does it for less. As a pure people mover, it just might be the best buy on the road today.

Here's how Mini Ram beats Vanagon

	Dodge Mini Ram Wagon 250	VW Vanagon
Sticker Price (MSRP)	**$9,365.00***	**$10,690.00***
EPA EST. MPG	19 MPG**	17 MPG**
Horsepower	95	67
Passenger Seating	8	7
Overall Length	178.9"	179.9"
Wheelbase	109.6"	95.8"
Fuel Capacity	36 gal.	15.9 gal.
EPA Est. Range	684 miles**	270 miles**
Side Door Width	49"	42"
Removable Rear Seat	Yes	No
Power Steering	Standard	Not Available
Color-Keyed Carpeting	Standard	$410 with int. pkg.
Factory-Installed Air Conditioning	Option	Not Available

Dodge Trucks

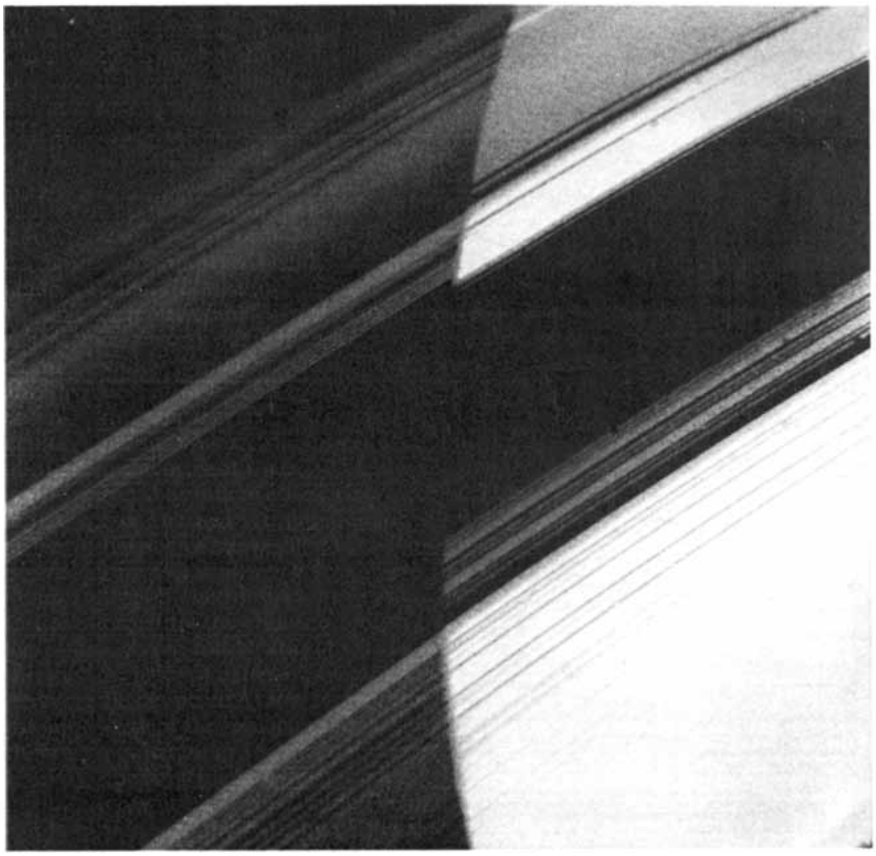
DODGE TRUCKS ARE RAM TOUGH

*Sticker price comparison, excluding title, taxes and destination charges, unadjusted for different levels of standard equipment. Whitewall tires $85 extra. Prices and specifications as of October 1, 1981.

**Use these numbers for comparison. Your mileage and range may differ depending on speed, distance and weather. Calif. est. lower.

gravitational perturbations may persist. In a ringlet that is more nearly opaque the greater incidence of collisions may damp the perturbations.

A full understanding of the physics of the interactions of rings with moonlets is not yet at hand. For example, Stanley F. Dermott and Thomas Gold of Cornell University have offered a hypothesis in which a narrow, eccentric ringlet with sharp edges is maintained by a single small moonlet hidden inside it rather than by a pair of surrounding moons or moonlets. Moreover, the lack of success (so far) at finding moonlets embedded in the main rings of Saturn means that other explanations of the fine radial structure in places such as the *B* ring must be explored. One hypothesis proposes that ring particles may respond to transient increases or decreases in their packing density by a further increase or decrease. Suppose the density of particles increases locally. The collisions between the particles become more numerous, and since some of the energy of the collisions is lost in the heating and deforming of the particles, their energy of random motion decreases. Unless the further collisions at lower velocities lose a smaller proportion of their energy the process of collision and coalescence continues. It is proposed that the overall result might be radial variations in the opacity of a ring.



VARIOUS DEGREES OF TRANSPARENCY and thus of the packing density of particles in the rings of Saturn are evident in an image made by *Voyager 1* that shows the unlit face of the rings as they cross the bright disk of the planet. The disk shines through most of the *C* ring (*bottom*), and it shines through Cassini's division to a similar extent, although both include ringlets whose opacity is greater. In contrast, much of the *B* ring (*middle*) is almost opaque.

Other Forces in the Rings

Small particles in the rings can be affected by forces other than gravity. Electromagnetic forces are an important example. The rings of Jupiter, of Saturn and possibly of Uranus lie in the midst of a plasma of low density, that is, they lie in a tenuous gas consisting of negatively charged electrons and positively charged ions. The electrons are less massive than the ions; therefore they move faster, and initially they collide with particles in the ring more often than the ions do. Eventually the particles become negatively charged by their absorption of electrons. At that point their charge repels the arrival of further negatively charged particles. More important, the ring particles themselves are now accelerated by an electromagnetic force as they cross the planet's magnetic field. If the particles are smaller than about .1 micrometer, the electromagnetic force is greater than the planet's gravitational attraction, and so it dominates their motion.

Several aspects of the structure of ring systems might thereby be explained. With Jupiter, for example, the axis of the magnetic field is tilted by some 10 degrees with respect to the axis of the planet's rotation. In such circumstances electromagnetic forces can give small particles a vertical distribution much greater than that of larger particles. The vertical extent of the halo of Jupiter's ring system is comparable to the extent that would be expected for particles whose size is .1 micrometer or smaller. With Saturn electromagnetic forces may turn out to be important for the peculiar structure of the *F* ring. Moreover, a number of ingenious theories have been advanced to explain how the spokes in Saturn's *B* ring arise without a simultaneous disturbance of the finely banded radial structure of the ring. Some of these theories invoke a rain of charged particles from either the planet or the rings themselves as a way of transferring electric charge to micrometer-size particles and lifting them from the surface of larger particles. The lifted particles are recaptured when they later collide with other large particles.

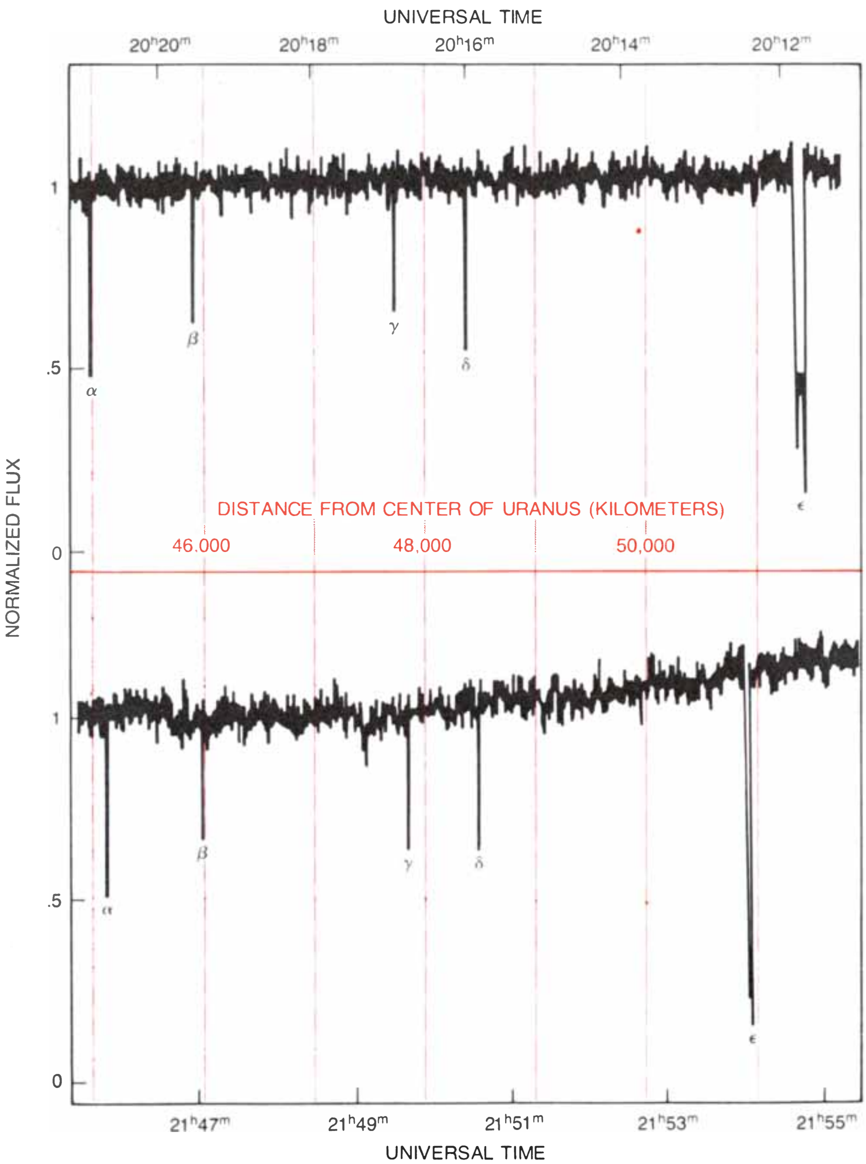
A second force that may dominate the motion of small particles in the rings is gaseous drag. Here the friction due to the presence of the plasma causes ring particles to spiral toward the planet. The smaller the particle, the faster the decay of its orbit. In only some 20 years, for example, gaseous drag can cause a micrometer-size particle to move from the outer edge of the bright ring of Jupiter to the inner edge. Another 200 years would take it through the diffuse disk and into the planet's atmosphere. The relatively structureless appearance of Jupiter's bright ring may arise in part from this rapid radial motion of the small particles in it. Indeed, the diffuse disk may be populated by particles that move into it from the bright ring.

In addition to gravity, electromagnetic forces and gaseous-drag forces, the small particles in a ring system are subjected to collisions that destroy them. For one thing interplanetary space contains solid bodies whose size ranges from less than a micrometer to more than a kilometer. The smaller bodies are by far the most numerous. The interplanetary bodies are in orbit around the sun, whereas the ring material is in orbit around a planet. Hence the two can collide at velocities of several tens of kilometers per second. If the interplanetary body is larger than about a hundredth the size of the ring particle, the collision destroys the particle. On that basis it can be estimated that a micrometer-size particle in any of the ring systems lasts for only about 10,000 years. Another type of collision may be even more destructive. In Jupiter's ring and in the outer rings of Saturn it is likely that a particle smaller than 10 micrometers is eroded by its collisions with high-energy ions in the planet's Van Allen belt well before a micrometeoroid hits it.

How did the rings arise? The sun, the

planets, the moons, the asteroids and the comets of the solar system are all thought to have formed about 4.6 billion years ago inside the solar nebula, a diffuse cloud of gas and dust. The sun and the giant planets Jupiter, Saturn, Uranus and Neptune formed in large part (or even entirely) from the gases of the nebula; Mercury, Venus, the earth, Mars, Pluto, the moons, the asteroids and the comets formed from solidified matter. The composition of the solidified matter would have depended on the temperature of the nebula from place to place. In general, however, the solidified matter consisted of varying mixtures of "rock" (which included silicates and iron) and "ices" (which included water).

At first Jupiter and Saturn, and perhaps Uranus and Neptune, were several hundred times larger than their current dimensions. Under the influence of their own gravitation they gradually shrank. As they did so they rotated faster. Eventually the increasing centrifugal force associated with the rotation caused the outermost part of each incipient giant planet—in essence a gaseous envelope—to become distinct from the more nearly spherical concentration of gas inside it, which continued to contract. The envelope served as a source of the solid grains from which the large moons of the planet accreted. Some of the grains would have formed from the condensation of gases such as water vapor as the envelope cooled. Perhaps the same process of accretion of grains into larger objects gave rise to a ring



URANUS' RINGS WERE INFERRED from the series of dimmings recorded as a star passed behind Uranus in 1977. The pattern of dimmings minutes before the star disappeared behind the planet (*top*) matched the pattern minutes after it reemerged (*bottom*). For four of the dimmings the match was almost exact; thus four rings are nearly circular. A fifth ring, designated epsilon, is notably eccentric. Further observations have shown four more rings. The data were collected by James L. Elliot and his colleagues on the Kuiper Airborne Observatory.

system somewhat closer to the planet.

After an interval on the order of a million years the circumplanetary envelopes disappeared. Perhaps they were blown away by a wind of ionized gas from the youthful sun, or perhaps friction within the envelopes caused them to collapse onto their planets. In any case the formation of moons, moonlets and ring particles ended.

Circumplanetary Conditions

Today Jupiter and Saturn give off about twice as much energy to space in the form of infrared radiation as the amount of energy they absorb in the form of sunlight. The excess represents the conversion of gravitational energy into heat by each planet's past and present contraction. At the time their moons were forming the planets were contracting much more rapidly. Hence Jupiter and Saturn and perhaps the other giant planets may have radiated many thousands of times more heat then than they do now. This heat may well have controlled the temperature and thus the composition of the solid matter in the circumplanetary envelopes. At a given time the temperature would have increased toward the planet and at a given place the envelope would have steadily cooled with time.

Several hypotheses about the origin of the ring systems around three of the giant gaseous planets posit that ring material formed in the circumplanetary envelopes at positions close to its current radial distances. Variations in temperature and other conditions near each planet may thus have been the cause of striking differences in the composition of their close-in moons and their ring material. The mean densities of the four largest moons of Jupiter imply that the inner two consist almost exclusively of rock and the outer two of comparable amounts of ice and rock. Since Jupiter's ring is well inside the orbit of the innermost large moon, any particles that formed near the ring could have consisted only of rock. Indeed, they could have consisted only of the relatively rare minerals that condense at high temperatures.

In contrast, the innermost large moons of Saturn all consist substantially of ice. Thus the temperatures even at the small distance from Saturn where the planet's rings now lie may have been cool enough to allow the formation of particles made up in large measure of water ice. The differences between the conditions around Jupiter and those around Saturn are readily explained. First of all, the amount of heat generated in early times by the contraction of a gaseous sphere depends approximately on the square of its mass. Jupiter, with more than three times the mass of Saturn, emitted 10 times more heat. It may also be that not much rock was available near Saturn for incorporation into particles. When Saturn's circumplanetary envelope formed, the planet's atmosphere probably extended into the region now occupied by the rings and the innermost moons. Hence most of the rocky grains there may have been kept in Saturn's atmosphere. They may also have been claimed by the moons farther out.

The hypotheses that the temperature near Saturn was relatively low and that little rock was available there are consistent with the finding that the inner moons of Saturn have a lower mean density and consequently a larger fractional content of ice than most of the outer moons. The hypotheses may also help to explain why Saturn's rings seem to be almost pure water ice, whereas Jupiter's are rock. Yet it seems the rings of Uranus are rock. One would think they would have been ice.

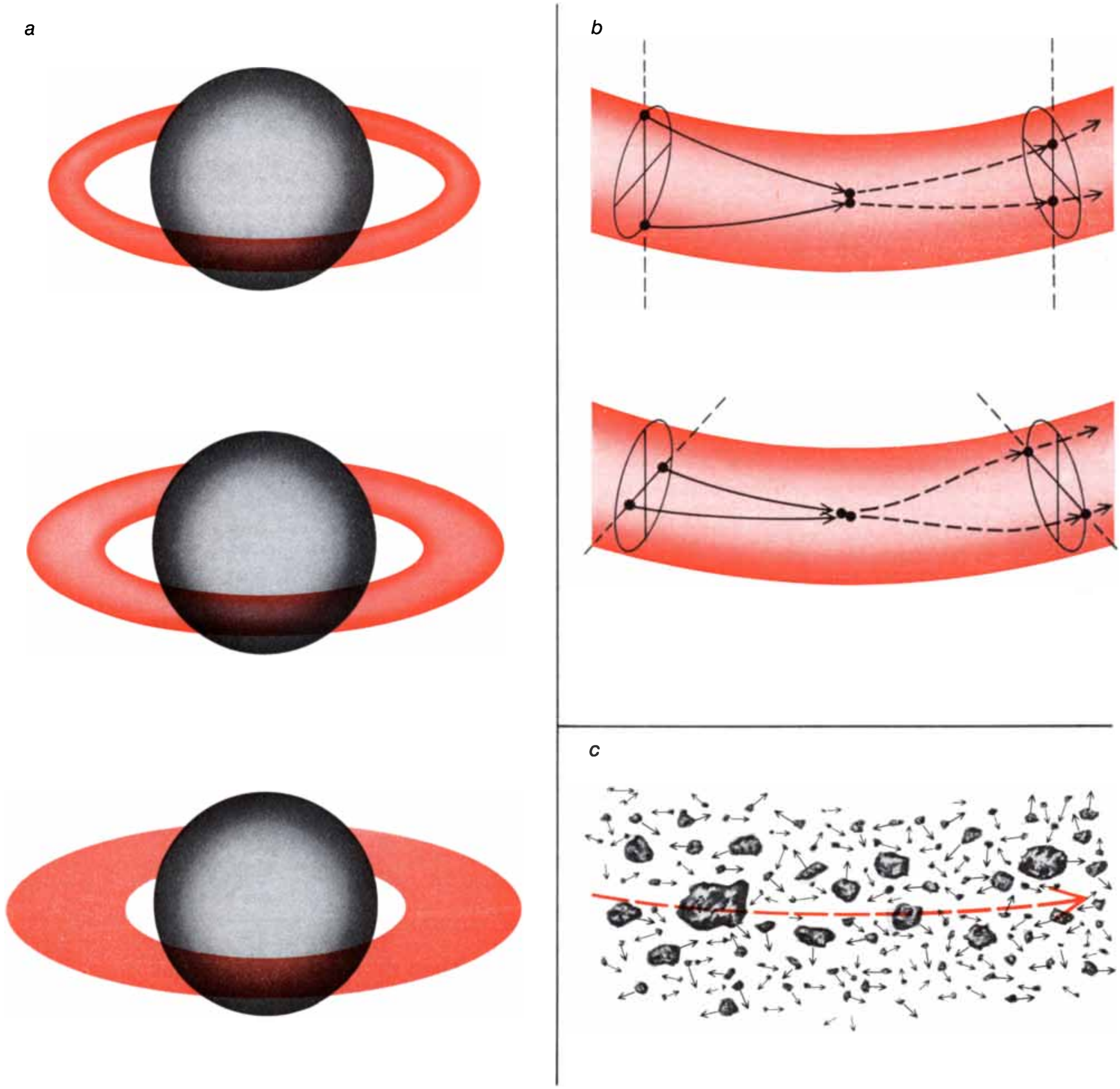
Rival Hypotheses

The various hypotheses that explain rings around the giant gaseous planets differ primarily in their account of the relation between the ring particles observed today and the primordial ring material. According to one hypothesis, a single large body was fragmented into myriad pieces when it came close to a planet and the fragments then formed rings. The body may have been a large meteoroid that suffered a chance gravitational encounter with the planet, or it may have been a moonlet that formed in the planet's envelope. In either case the agent of fragmentation would have been tidal disruption: the shearing force that arises because the gravitational attraction exerted by the planet on the body is greater for the parts of the body closer to the planet than it is for the parts farther away. The creation of Saturn's rings by tidal disruption was first proposed by the French mathematician Edouard Albert Roche in 1848. Roche had calculated that tidal forces exceed the cohesive self-gravitation of a liquid moon if the moon comes closer than about 1.5 Saturnian radii to the surface of the planet. This disruption threshold—the Roche limit—lies close to the outer edge of the main rings of Saturn.

It is quite unlikely, however, that a moonlet near Saturn would have been liquid. It would have been solid, and a solid moonlet is held together not only by self-gravitation but also by the forces that order the atoms in crystalline matter. According to Hans R. Aggarwal and Verne Oberbeck of the Ames Research Center of NASA, a solid moonlet smaller than about 100 kilometers in diameter cannot be tidally disrupted at any distance from the surface of a planet. Moreover, a larger solid moon cannot be disrupted at a distance greater than .4 planetary radius from the surface. That distance places the disruption threshold inside the inner edge of the main rings of

Saturn. It also is unlikely that the tidal disruption of a stray meteoroid could yield ring particles near Saturn. The particles, like their parent body, would have velocities sufficient for them to escape from the planet's vicinity.

There is nonetheless a subtle way in which tidal forces could have been important. As Roman Smoluchowski of the University of Texas at Austin has shown, the gravitational attraction of particles of equal size is insufficient to hold them together against disruption by tidal forces inside the classical Roche limit (the limit calculated by Roche for a liquid moon). In contrast, two particles that differ greatly in size can resist tidal disruption at distances well inside the classical limit. As it happens, the limit set by the tidal disruption of two particles of equal size and the limit set by the tidal disruption of two particles of unequal size lie close to the outer and inner edges of the ring systems around Jupiter, Saturn and Uranus. Within the outer disruption limit particles may have accreted, but so slowly that they were kept from aggregating into large moons. Within the inner disruption limit growth may have been almost impossible.



EVOLUTION OF A RING SYSTEM probably began in the early solar system, when the contraction of a rotating gaseous cloud into a giant planet such as Saturn left behind a gaseous envelope strung out by the centrifugal force associated with its rapid rotation (*a*). The particles accreting in the envelope must have orbited the planet, but they also had random motions that caused them to collide (*b*). The collisions were inelastic: they deformed the particles and also heated them. The corresponding loss of kinetic energy decreased their vertical motion. The collisions between two particles at nearby radial distances had an additional effect. In each such collision the inner particle would have been orbiting the planet faster; it would hit the outer particle from behind. The resulting transfer of momentum would raise the orbit of the outer particle and lower the orbit of the inner one. Thus the envelope spread horizontally. Ultimately a flat disk emerged in which the random velocities of the particles with respect to one another had only about a ten-millionth the magnitude of their mean orbital velocity (*c*). It is thought the smallest particles in the rings today result from the collisional erosion of the larger particles.

The second major hypothesis regarding the history of ring particles was suggested by Eugene Shoemaker of the U.S. Geological Survey. It postulates that a single large moon in the ring region (or perhaps a number of moons) collided catastrophically with a stray meteoroid. The photographs of the moons of Jupiter and Saturn made by the Voyager spacecraft do in fact show that the moons are scarred by a large number of craters produced by high-velocity collisions. Mimas, one of the smaller moons of Saturn, has a crater spanning a third of one hemisphere. The collision that made this crater must have come close to destroying Mimas. Nearer Saturn two objects with a diameter of about 100 kilometers occupy almost identical orbits. They may be the largest fragments of a catastrophic collision between a moon and a giant meteoroid.

There are several reasons why catastrophic collisions may have occurred preferentially in a region to be occupied later by rings. First, the major moons of Jupiter and Saturn tend to be smaller at distances closer to their planet. At a given energy of collision small moons are more likely to fragment than large ones. Second, the gravitational field of a planet focuses the trajectories of meteoroids, so that the flux of meteoroids passing close to the planet is significantly greater than the flux at increasing distances.

A final hypothesis regarding the history of ring particles postulates that the larger bodies in the rings are simply the result of the limited extent of the accretion of matter in the circumplanetary envelope at distances close to the planet. The accretion began with the cooling of the envelope and the resulting condensation of gaseous matter into tiny solid grains. Gravitational forces and gaseous drag caused the grains to settle into the equatorial plane of the envelope. There the grains continued to grow by the condensation of vapor onto their surface. Such growth could bring them to sizes as large as a few meters. The particles that make up most of the visible rings of Saturn range in size from a few centimeters to a few meters. They may be the product of such a process. Any moonlets in the rings would then be the local products of a stage of growth in which the meter-size bodies accreted further as the result of gentle collisions.

The ring particles in all three ring systems are small and numerous. According to the accretional hypothesis, there are several reasons why this is the case. First, the undisturbed formation of tiny grains could not begin at a given distance from an incipient giant planet until the planet had shrunk to a size inside that distance and the circumplanetary envelope had cooled sufficiently. Thus less time was available for the formation of grains near the planets than was available at greater distances. Second, the tiny grains near Jupiter could form only from the relatively rare, high-temperature condensates that were available. Finally, the arrival of a moonlet at a certain size would mean that its overlapping resonances spanned a width comparable to its dimension. Fresh material could no longer reach it, and so it would stop growing. In Saturn's rings the calculated limit to growth is a few kilometers to some tens of kilometers. Rather different conditions must have prevailed at greater distances from Saturn and the other giant planets, since bigger moons formed there.

Continuing Erosion

What about the small particles in the ring systems? We have already noted that gaseous drag causes micrometer-size particles to spiral from the outer edge of the bright ring into the atmosphere of Jupiter in a mere few hundred years. Plainly such particles could not have survived from the time the planet had a gaseous envelope. They must be forming today. According to Joseph A. Burns of Cornell, they result from the erosion of larger bodies in or near Jupiter's bright ring.

Ring particles larger than a centimeter or so are not readily destroyed by their collisions with interplanetary micrometeoroids. Instead each collision excavates a tiny crater around the point of impact, and an amount of matter 1,000 to 10,000 times greater than the mass of the impacting body is ejected. Many of the micrometer-size particles in the ring systems may originate, then, as ejecta. It can be estimated that if a moonlet is smaller than about 10 kilometers in diameter, most of the ejecta resulting from a collision with an interplanetary body can escape from the moonlet's gravitational field. The ejecta that escape from the moonlet would lack the energy to escape from the planet around which the moonlet orbits; thus they would take up orbits in the rings. (The continuing erosion of a parent population might also be a source of bodies in Saturn's rings on the order of a centimeter or a meter in size.)

In sum, the moonlets in the rings and also the largest ring particles probably date back to the early history of the solar system: they are contemporaries of the moons of the giant planets. The smallest ring particles are forming even now. It is suspected that the large moons of the outer solar system and also several planets (including the earth) formed by the accumulation of many bodies of smaller size. Surely among the multitudes of particles in the rings similar processes are being enacted on a small scale today. The ring systems thus offer a double challenge. One seeks to deduce the processes that made them and then one seeks to use this knowledge to gain insight into how the solid moons and planets formed.

The Fluid Phases of Matter

After a century-long effort to understand the structure of liquids and gases a successful theory has been based on one of the earliest and simplest models, in which the molecules are hard, inert spheres

by J. A. Barker and Douglas Henderson

The difference between a liquid and a gas is obvious under the conditions of temperature and pressure commonly found at the surface of the earth. A liquid can be kept in an open container and fills it to the level of a free surface. A gas forms no free surface but tends to diffuse throughout the space available; it must therefore be kept in a closed container or held by a gravitational field, as in the case of a planet's atmosphere. The distinction was a prominent feature of early theories describing the phases of matter. In the 19th century, for example, one theory maintained that a liquid could be "dissolved" in a vapor without losing its identity, and another theory held that the two phases are made up of different kinds of molecules: liquidons and gasons. The theories now prevailing take a quite different approach by emphasizing what liquids and gases have in common: They are both forms of matter that have no permanent structure, and they both flow readily. They are fluids.

The fundamental similarity of liquids and gases becomes clearly apparent when the temperature and pressure are raised somewhat. Suppose a closed container partially filled with a liquid is heated. The liquid expands, or in other words becomes less dense; some of it evaporates. In contrast, the vapor above the liquid surface becomes denser as the evaporated molecules are added to it. The expansion of the liquid and the higher density of the vapor increase the pressure in the container. With further heating the density of the liquid continues to diminish and that of the vapor continues to increase until ultimately they are equal. The combination of temperature and pressure at which the densities become equal is called the critical point. Above the critical point the liquid and the gas can no longer be distinguished; there is a single, undifferentiated fluid phase of uniform density.

The significance of the structural and dynamical differences between liquids and gases was first questioned in the 1870's by the Dutch physicist J. D. van der Waals. At the time there was already a well-established theory of an ideal gas, that is, a gas in which the relations of temperature, pressure and volume follow a particularly simple law. Van der Waals suggested that a liquid be considered a dense gas, which could be described by a modification of the equation of state for an ideal gas.

The more recent theories of fluid structure that we shall describe here can be traced back to van der Waals's work. The basic conceptual model that underlies the theories is a fluid composed of hard spheres. The properties of the model can be explored by means of simulations done with a computer. The model can also serve as a basis for the description of real fluids by perturbative methods, in which the basic model is refined by adding to it a sequence of progressively subtler modifications.

Virtually all elements and chemical compounds have a solid, a liquid and a vapor phase. A transition from one phase to another is accompanied by (or can be provoked by) a change in temperature, pressure, density or volume. The nature of such transitions can be illustrated by again considering a substance in a closed container, where the various conditions that determine its phase can be controlled.

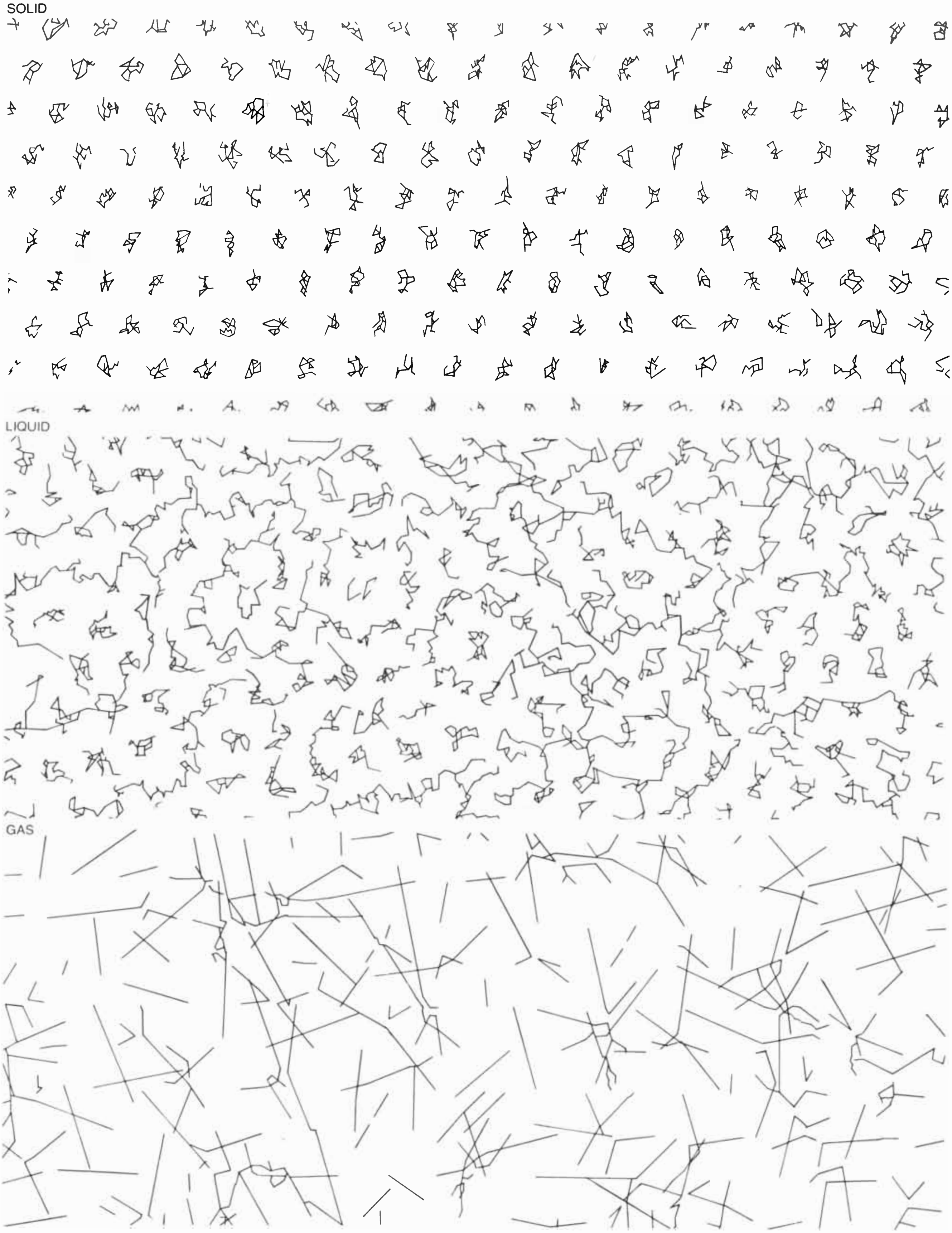
Suppose the substance is initially in the solid state and the temperature is held fixed as the volume of the vessel is increased. At first the volume of the solid increases only slightly, and so the pressure declines. When the solid begins to melt, however, the pressure suddenly becomes constant because the total volume occupied by the coexisting solid and the liquid keeps pace with the increasing volume of the container. When all the substance has been converted into a liquid, the pressure again begins to fall, although not as quickly as it did for the solid. After a further increase in the volume of the chamber the boiling point is reached; it is marked by another plateau in pressure as the formation of vapor compensates for the increase in the container's volume. Finally, when all the liquid has evaporated, the pressure once more resumes its decline.

If this experiment were carried out at a temperature exceeding the critical point of the substance, there would be no discontinuous transition from liquid to vapor. Instead the pressure and density of the fluid would fall off smoothly. It should be emphasized, however, that the solid-fluid transition persists up to the highest temperatures at which the molecules retain their identity. There is no critical temperature for melting.

The task of a theory that would describe the phases of matter is to account for observations such as these in terms of the motions of the individual atoms or molecules that make up the substance. In other words, a connection must be established between the thermodynamic properties of common substances and the mechanics of the individual molecules.

Under most circumstances the motions of molecules can be described with acceptable accuracy by the methods of classical, or Newtonian, mechanics. In classical mechanics the acceleration of a molecule (that is, the change in its velocity) is determined entirely by its mass and by the forces acting on it. If the positions and velocities of all the molecules in a specimen of matter are known at a given instant, and if the nature of the forces between the molecules is understood, then in classical mechanics the positions and velocities can be calculated for all other times, both in the past and in the future. By making such a calculation the events that constitute a change of phase might be traced in ultimate detail.

One realm where classical mechanics cannot provide an adequate description is in determining the forces between molecules. The only significant forces are electromagnetic ones; they arise from a distortion that develops in the cloud of electrons surrounding a molecule when another molecule is nearby. Because an electron is much lighter than a molecule and moves correspondingly

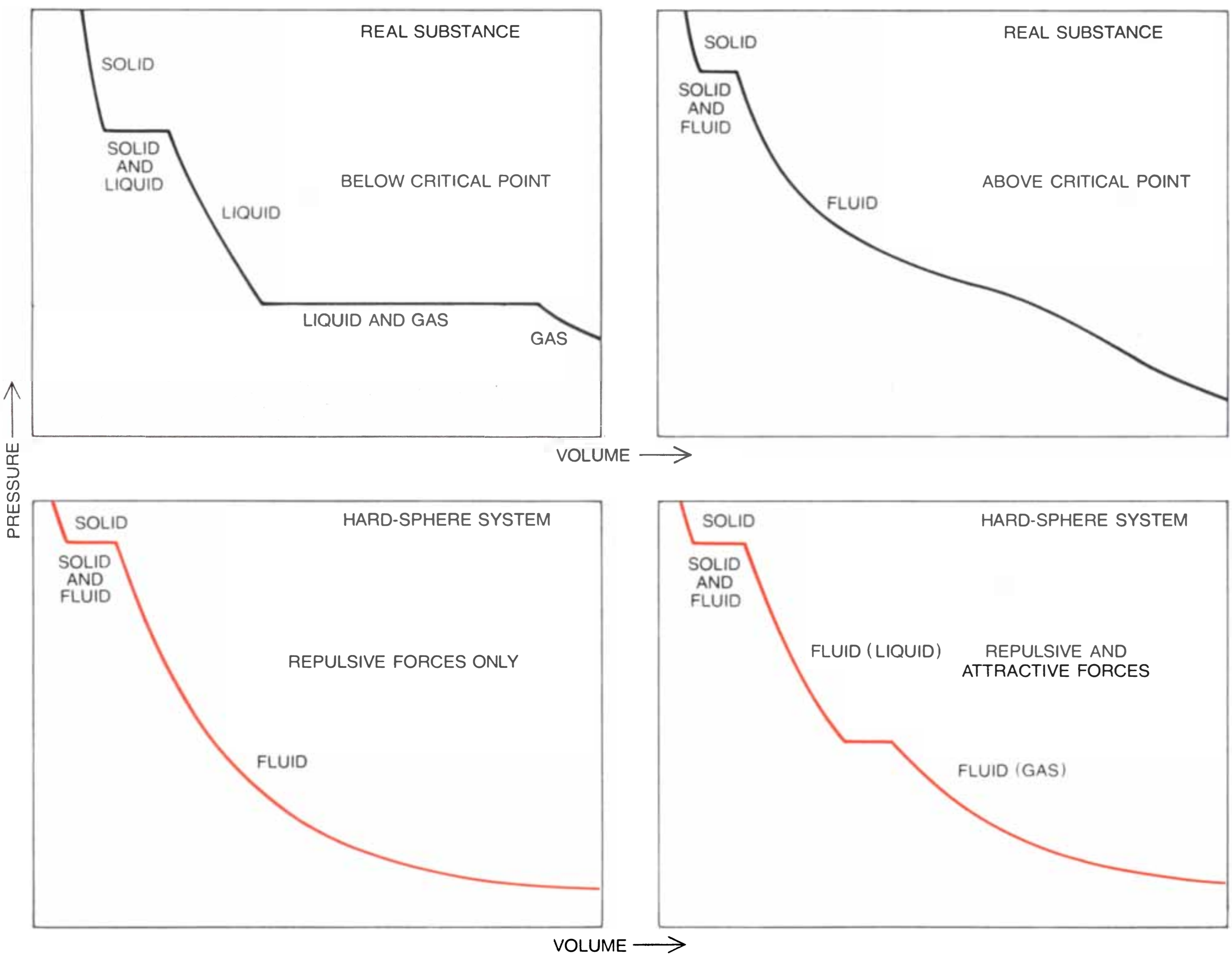


TRAJECTORIES OF MOLECULES in a solid, a liquid and a gas were simulated with the aid of a computer. The two-dimensional system of molecules has the same phases and phase transitions as a real substance, but the molecular positions and motions are more easily displayed. In the solid the molecules are constrained to vibrate about fixed lattice sites, whereas molecules in the liquid and the gaseous phases are free to wander. The only substantial differences between the two fluid states are those of density and of frequency of collision. The computer program calculates the trajectories by solving the equations of motion for some 500 two-dimensional molecules. The simulation was done by Farid Abraham of the International Business Machines Corporation Research Laboratory in San Jose, Calif.

faster the configuration of the electron cloud can be described only by the methods of quantum mechanics. Calculations based on quantum mechanics are far more difficult than those based on classical mechanics, but two important simplifications can be adopted in defining the intermolecular forces. First, precisely because the electron moves much faster than the atomic nuclei of a molecule do, one can assume that the nuclei are stationary. Second, it turns out that the force between two molecules is almost independent of the presence of other molecules. Hence a good approximation of all the forces acting within a system made up of many molecules (even molecules in rapid motion) can be deduced from the force between just two stationary molecules.

Even the much reduced task of defining the nature of the force between two stationary molecules has proved to be a considerable challenge. Indeed, so far it has been done precisely only for the inert gases, whose molecules are actually individual atoms that have a particularly simple electronic structure and a spherical symmetry. The importance of the latter property is that it means the force between molecules is the same in all directions and depends only on the distance between them. A force that also depends on the orientation of the molecules can give rise to interesting properties in the condensed phases of matter; it does so, for example, in water. The study of such anisotropic forces is currently an active area of investigation, but we shall confine our attention mainly to the isotropic forces such as those of the inert gases.

The force between two inert-gas atoms can be determined in several ways. For helium, where the system of two atoms includes only four electrons, the quantum-mechanical equations that define the configuration of the electron cloud can be solved numerically with a large computer. Such a calculation has been done by our colleagues Bowen Liu and A. Douglas McLean at the International Business Machines Corporation Research Laboratory in San Jose, Calif. For the heavier inert gases and for more complex molecules complete calculations are generally not feasible, but it is possible to measure the intermolecular forces experimentally. Some of the measurements are quite indirect: the force is deduced from certain properties of the gases that depend on two-molecule collisions, such as the viscosity and the thermal conductivity. Recently more direct methods have become available, in which beams of atoms are made to col-



PRESSURE AND VOLUME of a real substance and of a model system composed of hard spheres vary sharply as the phases change. The upper diagrams show the relation of the pressure to the volume in the real substance at two temperatures. Below the temperature called the critical point two phase transitions are observed, each of which is represented by a plateau in the pressure-volume diagram. In these regions the substance separates into two distinct phases, and an abrupt phase transition can result from a smooth change in pressure or temperature. Above the critical temperature the solid-fluid transition persists but liquid and gas cannot be distinguished; there is only a supercritical fluid whose volume can be varied continuously without a change in phase. The lower diagrams show the pressure-volume relation for the hard-sphere system, in one case where only repulsive forces act between the spheres and in another case where attractive forces have been added. In the hard-sphere system the addition of attractive forces creates a plateau much like the one that separates liquid from gas at subcritical temperatures in the real fluid. Apparently attractive forces are needed to form the liquid state.

lide or the radiation absorbed and emitted by weakly bound diatomic molecules of the inert gases is analyzed.

The results of such calculations or measurements are usually expressed in a potential-energy function, from which the force between two molecules can easily be determined. The function specifies the energy associated with the interaction of the molecules in relation to the distance between them. The molecules tend to approach whatever separation minimizes their total energy, and this tendency is observed as a force that impels the molecules toward the optimum separation. On a graph of the potential-energy function the intermolecular force is given by the negative of the slope of the curve; at the point of minimum energy the slope is zero and so there is no force. At the optimum separation the energy of the interaction is negative, that is, the energy of the combined system of two molecules is less than that of two isolated molecules.

In rough outline the potential-energy functions for almost all molecules have the same basic form. When the molecules are infinitely far apart, they do not interact at all, and so both the potential energy and the force between the molecules are set equal to zero. As the molecules are brought closer their electron clouds are distorted in such a way that the total energy is slightly reduced; as a result there is an attractive force that tends to bring the molecules closer still. This process continues until the molecules are approximately one molecular diameter apart. In order to reduce the separation further, energy must be expended because the electron clouds begin to overlap; the force then becomes strongly repulsive, reflecting the repulsive interactions of the electrons. The equilibrium point of minimum energy and zero force is attained where the attractive and the repulsive interactions are exactly in balance.

The simplicity of the potential-energy function for a pair of molecules immediately raises a question: How can such a simple interaction between molecules give rise to the three phases of matter with their distinctive properties? The function has only one point of minimum energy. In a solid the molecules are arranged so that their spacing is as close as possible to this minimum-energy distance. What then accounts for the existence of liquids and gases? Since their energy must evidently be greater than the minimum, why do they not reduce their energy by condensing into solids? One might expect the potential-energy function to have a second and a third minimum to accommodate the fluid phases, but the additional minimums do not exist.

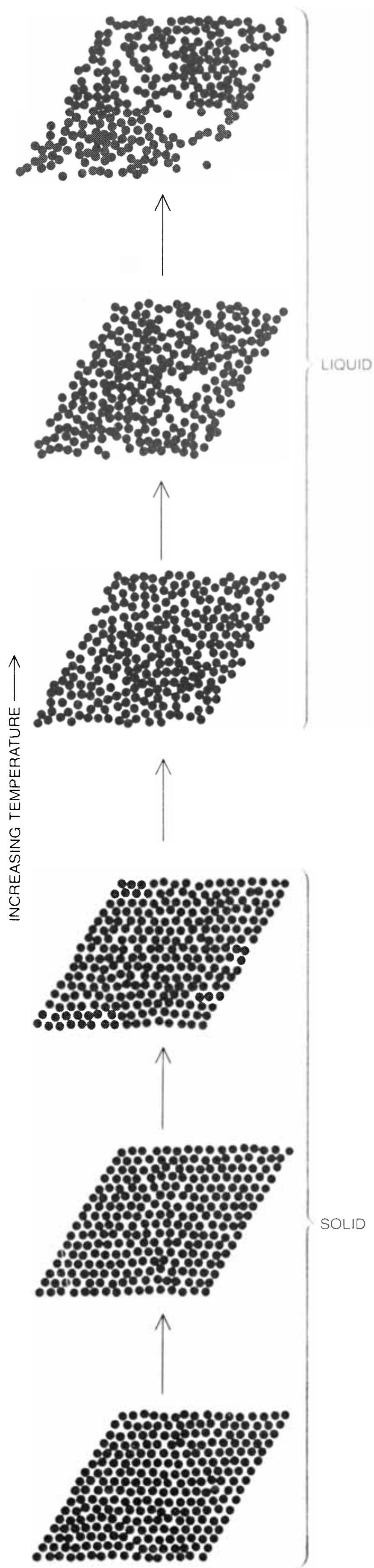
The resolution to this paradox lies in the fact that the system of many molecules in a macroscopic specimen of matter is not a simple mechanical system, and energy is not the only quantity that has an influence on its behavior. The system of molecules is governed by the laws of thermodynamics and statistical mechanics, where an important role is also assigned to the temperature and to the measure of disorder called entropy.

A simple mechanical system, such as a ball rolling down a mountainside, consistently comes to rest in its state of lowest energy, namely at the bottom of the valley. Similarly, two molecules in isolation always tend to assume their minimum-energy separation. A thermodynamic system, on the other hand, approaches not the state of lowest energy but the state of lowest free energy. The free energy (A) is equal to the difference between the energy (E) and the temperature (T) multiplied by the entropy (S); thus a thermodynamic system is governed by the equation $A = E - TS$.

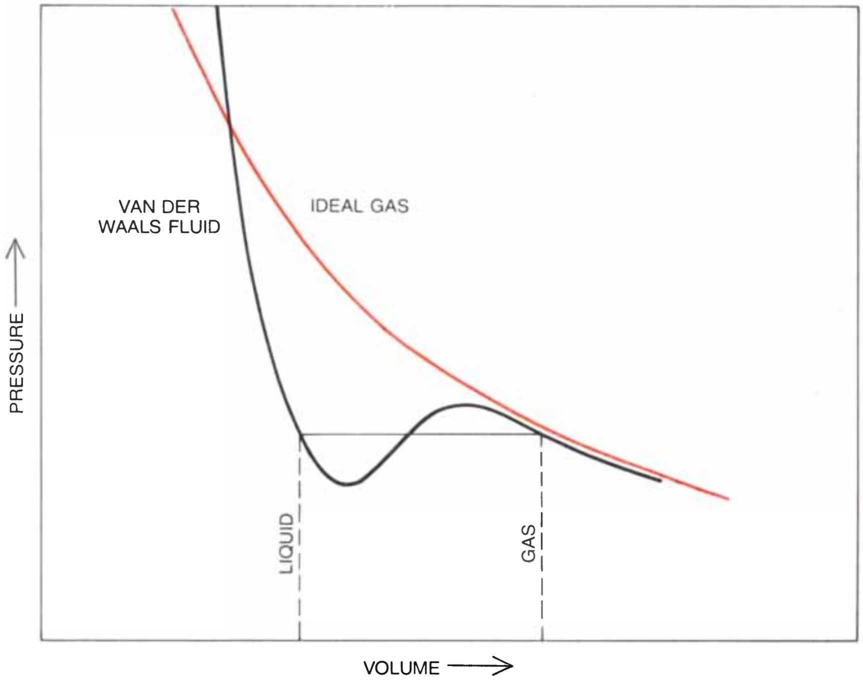
From this relation the salient characteristics of a thermodynamic system can be deduced. At low temperature the energy of the molecules makes the greater contribution to the free energy and so the solid phase, which minimizes the energy, is the favored state. At higher temperature the entropy of the system becomes the predominant influence. As a result the fluid phases are stable at elevated temperatures even though they constitute higher-energy configurations than the solid phase.

The concepts of entropy and free energy can be understood by considering the relation between the macroscopic and the microscopic descriptions of a thermodynamic system. The macroscopic state of the system is specified by quantities such as the temperature and the density, which are collective properties of all the molecules. The microscopic state is defined by the detailed configuration of the molecules. It is an essential point that many microscopic states can correspond to one macroscopic state. In other words, many configurations of the molecules can give rise to the same values of temperature, density and other quantities.

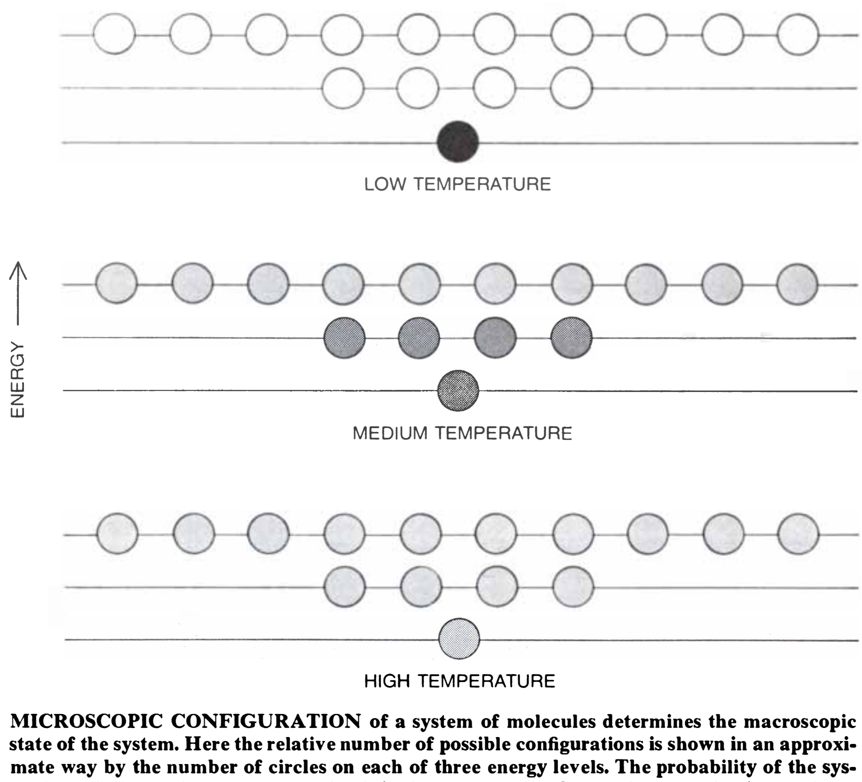
In statistical mechanics entropy is interpreted as a measure of the intrinsic probability of a macroscopic state. The probability is determined by the number of microscopic states that correspond to



CONFIGURATION of the molecules of a two-dimensional substance is seen in "snapshots" made during a computer simulation. In the bottom panel the substance is a solid, which expands and becomes slightly disordered as the temperature increases. At the melting temperature there is a discontinuous transition from solid to fluid. In the fluid the molecules move about freely, creating transitory voids as the substance expands further. The simulation was done by Abraham.



THEORY OF LIQUIDS was formulated by J. D. van der Waals by modifying the equation of state that defines the relation of pressure, volume and temperature in an ideal gas. The dip in the van der Waals potential results from introducing attractive forces between the molecules and from modifying the available volume to exclude the volume taken up by the molecules themselves. The attractive intermolecular forces serve to reduce the total pressure of the fluid.



MICROSCOPIC CONFIGURATION of a system of molecules determines the macroscopic state of the system. Here the relative number of possible configurations is shown in an approximate way by the number of circles on each of three energy levels. The probability of the system's having a particular configuration is indicated by the darkness of the shading in the circle. The probability is determined by both the energy of the configuration and the temperature of the system. At low temperature the lowest-energy configuration is the only one with an appreciable probability; this configuration corresponds to the solid state. At intermediate temperature the low-energy configurations are still favored, but because there are many more high-energy configurations each energy level can have roughly the same total probability; this situation prevails during the transition from the solid to the liquid phase. At high temperature the probabilities of all configurations are nearly equal. One is almost certain to observe a high-energy configuration because those configurations are more numerous. The solid has melted.

the macroscopic state and by the probability assigned to each of the microscopic states. All microscopic states of equal energy have equal probability, but the probability of a state declines steeply (following an exponential curve) as the energy of the state is increased or as the temperature is decreased. The mathematical expression that defines the dependence of the probability on the energy and the temperature is called the Boltzmann factor, after the Austrian physicist Ludwig Boltzmann.

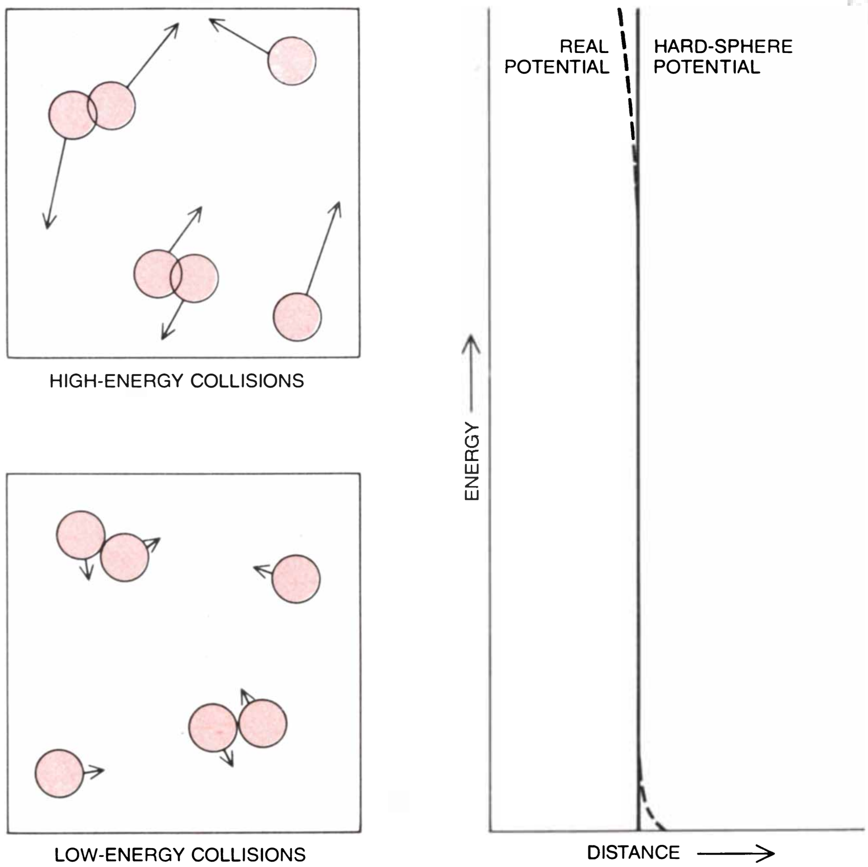
The Boltzmann factor implies that the individual microscopic states having the lowest energy are always the most probable ones, but it does not follow that the observed macroscopic state will necessarily have a low energy. Some macroscopic state of higher energy may be generated by a much larger number of high-energy microscopic states. Although each of the high-energy microscopic states has a low probability, the states can be so numerous that the system is overwhelmingly likely to occupy one or another of them. In this view fluids exist because there are many more ways to organize molecules in a liquid than there are in a solid, and there are still more ways in a gas.

The theory of statistical mechanics was presented in complete form by the American physicist Josiah Willard Gibbs in his book *Elementary Principles in Statistical Mechanics,* published in 1902. Gibbs showed how a system of particles at a given temperature distributes itself among the available energy levels. His elegant results bridge the gap between two great world views: the mechanical and the thermodynamic.

With an understanding of intermolecular forces and of statistical mechanics the basic conceptual tools needed to investigate the phases of matter were in hand; some of the practical tools, however, were still lacking. By the mid-20th century there were satisfactory theories of solids, which are highly ordered, and of gases that are no more than moderately dense, which are highly disordered. There was no fully workable theory, however, for liquids and compressed gases.

The main impediment to the development of such a theory was the large number of molecules that must be included in any description of a liquid or a dense gas. Such many-body problems cannot be solved analytically; indeed, even the three-body problem has no exact solution except in special circumstances. Approximation methods had to be developed to describe dense fluids. The most direct method is the simulation of the fluid with a computer.

The first step in constructing a mathematical model of a fluid is to choose an appropriate potential-energy function to define the force between each



EFFECTIVE DIAMETER OF HARD SPHERES decreases with increasing energy or temperature. At low energy molecular collisions are on the average less forceful and there is negligible overlapping of electron clouds. At higher energy molecular collisions are more violent and the molecular electron clouds overlap more. Because the molecular centers then approach each other more closely the effective diameter of the molecules is smaller. This effect can be incorporated into a hard-sphere fluid by making the diameter a function of temperature.

pair of molecules. A function derived from quantum-mechanical calculations or from experimental measurements can be employed, but it is often advantageous to have a function with a simpler mathematical structure. For example, in about 1930 John E. Lennard-Jones introduced a simple function that approximates quite closely the form of the function that determines the interactions of inert-gas atoms. Much exploratory work has been done with an even simpler function, namely that of the hard-sphere model. In the hard-sphere fluid there is no attractive potential and the spheres are perfectly rigid, so that no interpenetration is possible. Thus the repulsive potential is infinitely steep.

A model consists of a collection of molecules, which must be assigned initial positions. At each step in the evolution of the model new positions are determined that represent a new state of the system. In some cases the velocities of all the molecules are also determined at each stage. From averages of all the configurations observed after many steps, macroscopic properties of the model fluid can be determined and compared with experimental findings for real fluids.

One method of generating successive states of the model system is called the Monte Carlo method because its essential mechanism is a random-number generator whose function is conceptually equivalent to that of a roulette wheel. The algorithm on which the method is based was invented in 1953 by Nicholas C. Metropolis, Arianna W. and Marshall N. Rosenbluth and Augusta H. and Edward Teller; it was employed in its early stages by William W. Wood. The algorithm generates a sequence of configurations of a many-particle system that imitates some of the statistical properties of molecular motion. In particular the probability of a given configuration's appearing in the sequence is proportional to the Boltzmann factor for that configuration, just as it is in a real thermodynamic system.

In the Monte Carlo method the computer program moves one molecule at a time in accordance with the rules of statistical mechanics. Because of the close packing of molecules, however, an arbitrary move may bring a molecule so close to one of its neighbors that they overlap. Repulsive forces then cause the potential energy to become large and positive. Since the Boltzmann factor declines exponentially as the energy increases, the overlapping causes the Boltzmann factor to be effectively zero for most configurations. In practice this difficulty is overcome by calculating for each configuration in a series the change in energy resulting from each move. If a molecule comes too close to its neighbors, the program in effect moves it back to its original position. With this modification the frequency of each configuration is proportional to its Boltzmann factor.

Another computer simulation method was developed by B. J. Alder and T. E. Wainwright of the Lawrence Berkeley Laboratory, Aneesur Rahman of the Argonne National Laboratory and Loup Verlet of the Laboratory of Theoretical and High-Energy Physics at Orsay in France. It is called the method of molecular dynamics, and it relies on a direct solution of the classical equations of motion for a system of molecules. At each stage in the evolution of the model the potential-energy function is employed to calculate the force exerted on each molecule in the sample by the neighboring molecules. From the net force the acceleration is determined, and from the acceleration the new position and velocity are calculated.

Unlike the Monte Carlo method, the molecular-dynamics method provides information about the evolution of the system with the passage of time. It can therefore be employed to calculate time-dependent properties of the fluid system such as viscosity. Properties such as the entropy can also be found by averaging the configurations observed over some period of time.

One description of a fluid that can be derived from a computer simulation is the radial distribution function, which gives the average density of molecules as a function of radial distance from a reference molecule. In a model fluid it is a straightforward procedure to calculate the radial distribution function: one simply tabulates the distances between a selected molecule and all the other molecules. The calculated distribution function for the model can then be compared with the function for a real solid or fluid. The radial distribution function of a real material is determined experimentally by measuring the angles at which X rays or neutrons are scattered by the atoms or molecules.

The radial distribution function readily distinguishes among the phases of matter. At small distances the function is equal to zero in all cases because two molecules cannot occupy the same position. Beyond this forbidden zone the function for a gas at low density and high temperature is almost uniform at all distances because the distribution of the molecules is random. In a crystalline solid the distribution function is a series of sharp peaks.

The most interesting radial distribution functions are those of liquids and dense gases. A first peak at a distance of about one molecular diameter represents the nearest-neighbor molecules. Because there is a high probability of neighboring molecules' being in contact, the nearest neighbors fall within a narrow range of distances and the peak is sharp. After the first maximum there is a minimum in the function, followed by a smaller maximum at about two molecular diameters, corresponding to the formation of a second shell of molecules. At greater separation the peaks become broader, and beyond about four diameters the modulation of the distribution function damps out entirely; there the density of the molecules approaches the uniform density of the liquid as a whole.

The radial distribution function for argon has been determined precisely by neutron scattering. The results of computer simulations of liquid argon are in excellent agreement with experimental findings both for the radial distribution function and for thermodynamic properties such as the relations of pressure, volume and temperature.

Although the Monte Carlo and the molecular-dynamical methods employ different algorithms, the working details of the two models are similar. If the calculations are to be completed in a reasonable amount of time, the number of molecules must be limited. In most applications the number ranges from about 100 to about 1,000. Typically the molecules are assumed to fill a box whose dimensions are determined by the fluid density. With a fairly small number of molecules in the box surface effects take on exaggerated importance.

The computer program partially corrects for the surface effects through the use of periodic boundary conditions. The box is surrounded on all sides by many periodic replicas of itself, like unit cells in a crystal. A molecule in the box can interact with "ghost" molecules in the replica boxes. Moreover, whenever a molecule moves out of the box, its ghost automatically enters through the opposite side. By this means the number of molecules is held constant.

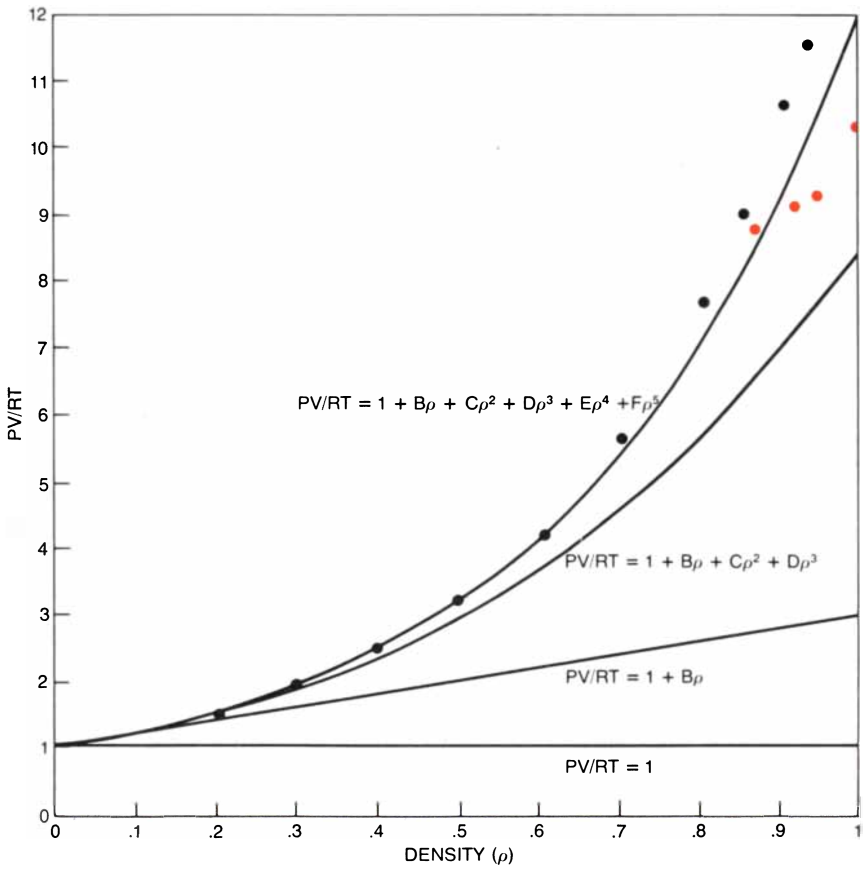
It has turned out that a system with as few as 100 molecules can give remarkably accurate estimates of the properties of condensed phases of simple substances such as argon. The only region of the phase diagram where the models fail is in the vicinity of the critical point. Near the critical point fluctuations in density and other properties develop at many scales of size. Long-range interactions, including the electrostatic interaction between ions and the interaction between polar molecules such as water, also present special difficulties.

A simulation done with a digital computer represents an attempt to calculate the properties of a fluid from first principles. The second major line of development in the theory of fluids takes a quite different approach. First a mathematically tractable model is sought that provides a reasonably good approximate description of the system, then successive refinements are added to it.

Methods of this kind, which are called perturbation methods, were first applied to the calculation of planetary orbits. In that context the initial approximation is to consider the sun and the planet in isolation. Then perturbations to the orbit caused by the presence of other planets are added to the approximation one at a time. The corrections usually take the form of a power series in which some small quantity, x, is raised to progressively higher powers. Each term in the series describes the effect of a perturbation. Evaluating the series becomes increasingly difficult as more terms are added, but if x is small, the importance of the terms diminishes rapidly as x is raised to higher powers. Hence only a few terms may be needed to achieve the desired accuracy. The most difficult part of this method is finding an x that is small enough to yield accurate results with a reasonable number of terms.

In applying perturbation methods to the theory of fluids there is an obvious candidate for the reference system: it is the ideal gas, made up of molecules that do not interact at all. The pressure in such a gas is directly proportional to the temperature and inversely proportional to the volume; the constant of proportionality is R, the gas constant. The relation is expressed mathematically by the equation $P = RT/V$, which is known as the ideal-gas law.

Perturbation methods suggest a way



PERTURBATION SERIES yields a sequence of progressively more accurate descriptions of a hard-sphere fluid. The reference system, with which the series begins, is an ideal gas (one in which molecules do not interact). The equation of state for the ideal gas can be written as $PV/RT = 1$, where P is the pressure, V is the volume, R is a constant and T is the temperature; hence for the ideal gas the value of PV/RT remains constant at 1 no matter what the density is. In a real gas at high temperature the value of PV/RT tends to increase with density. The increase is accounted for in the theory by adding to the equation of state a series of terms in which the density is raised to successively higher powers. Thus the equation takes the form $PV/RT = 1 + B\rho + C\rho^2 + D\rho^3$ and so on. The sum of six such terms provides a good approximation to the results of computer simulations of the hard-sphere model done by the authors (*black dots*) and by B. J. Alder and T. E. Wainwright of the Lawrence Berkeley Laboratory (*colored dots*). For a model of the liquid state, however, an ideal gas is not a satisfactory reference system. At liquid densities and low temperature successive terms become larger, not smaller.

of modifying the ideal-gas law to describe the relations of pressure, temperature and volume in a real gas. A series of terms is added to the law in which the density of the gas is raised to successively higher powers. This perturbation theory of imperfect gases was worked out in detail by Joseph E. Mayer. It provides a satisfactory account as long as the density does not become too great. At high density the successive terms in the series become larger rather than smaller and so an accurate calculation is impossible. For this reason the ideal gas is not a satisfactory reference system for a theory of liquids. The underlying explanation of this failure is that the molecular arrangements that are important in the ideal gas are quite different from those in a liquid.

A more promising reference system for liquids is the hard-sphere model. In this model the repulsive interactions that become important at high density are approximately accounted for and the commonest molecular arrangements are similar to those in liquids.

The first attempt to describe liquids by a hard-sphere model was made by van der Waals. He proposed that the ideal-gas law be modified in two ways. First, he argued that the volume V should be replaced by $V - b$, where b represents the volume from which a molecule is excluded by repulsive interactions with other molecules. From the assumption that molecules are hard spheres a value for b can be calculated. The second proposed modification was to subtract from the pressure an "internal pressure" defined by the expression a/V^2, where a is a constant. It is through the second modification that the effect of attractive forces is introduced. Hence the equation for the pressure of a liquid becomes $P = RT/(V - b) - a/V^2$.

Van der Waals justified subtracting the internal pressure by appealing to his intuitive ideas about the "association" of molecules. As a molecule nears the surface of a container, van der Waals suggested, it has more molecules behind it than in front of it. As a result not all the attractive forces of the other molecules cancel one another. The effect is to create a slight force drawing the molecule back into the volume of the liquid, and therefore the force exerted by the molecule on the walls of the container is reduced. A reduced force on the wall is equivalent to a lowered pressure. The effect becomes noticeable only in liquids and compressed gases, where the molecules are packed close enough for the attractive force between them to become significant. The internal pressure is therefore a function of density.

On the basis of these formulations van der Waals was able to demonstrate that the liquid and the gas states merge continuously into each other. His general ideas about the continuity of states and the packing of molecules were sound; even his specific approximations, such as the expression a/V^2 for the internal pressure, are nearly correct. Nevertheless, van der Waals did not have a satisfactory theory of the hard-sphere fluid. Such a theory did not become available until computer simulations of hard-sphere fluids began about 20 years ago.

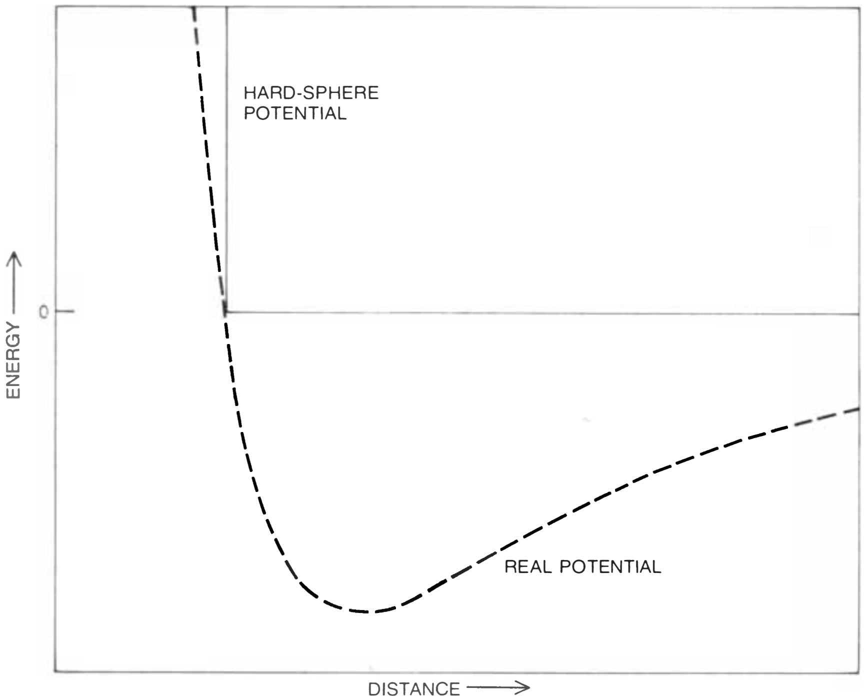
The weak point in the van der Waals theory of fluids was the quantity b, representing the excluded volume. If b is assumed to be a constant, the van der Waals equation yields only a poor approximation of the hard-sphere pressure. At high density the approximation grossly overestimates the pressure because the volumes excluded by different molecules overlap, so that the total excluded volume is less than the sum of the volumes excluded by all the molecules considered in isolation. As a result attempts to fit the van der Waals equation to experimental data often lead to physically implausible values of b. This poor estimate of the hard-sphere pressure obscured the fact that the van der Waals theory is basically correct.

In the 1950's and 1960's a number of workers began to investigate the application of the hard-sphere model to the description of real fluids. Fifteen years ago we contributed to this work by formalizing van der Waals's ideas about the hard-sphere reference model in a rigorous but practical theory. In our work we employed ideas put forward by Alder, E. B. Smith, J. S. Rowlinson, Donald A. McQuarrie and Joseph L. Katz. Later developments were introduced by John D. Weeks, David Chandler and Hans C. Andersen.

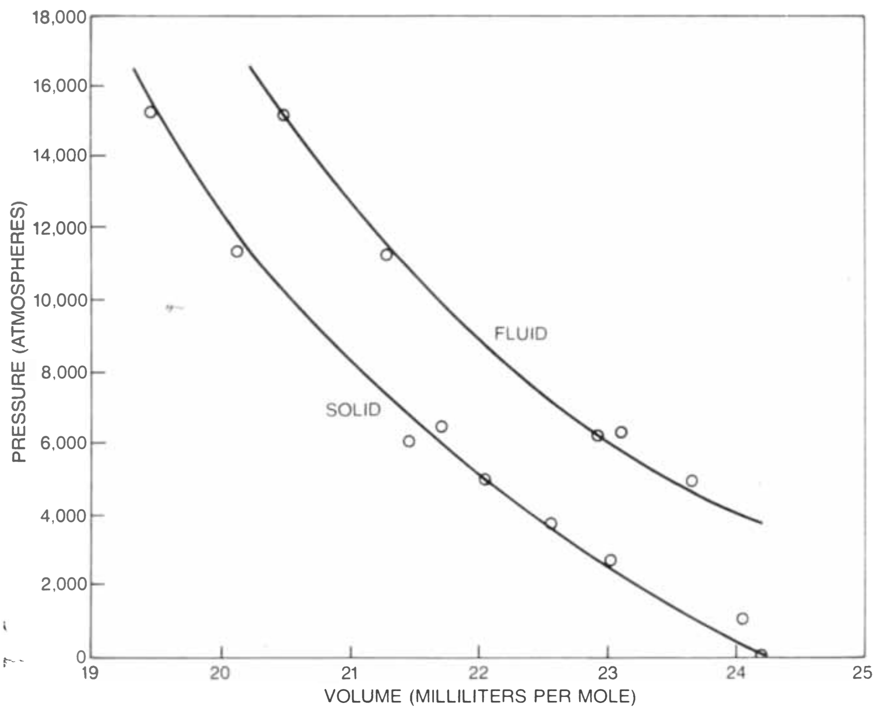
In seeking a rigorous formulation of the ideas of van der Waals our approach is to divide the intermolecular potential into short- and long-range parts. The potential is divided at the point where it changes from attractive to repulsive. The two parts of the potential then serve as the basis for the first two terms in a perturbation series; the variable that is raised to successively higher powers is the depth of the potential divided by the temperature. The first term represents the steep, short-range repulsion, related to the hard-sphere diameter. The second term gives the major effect of the attractive part of the potential, which is weak but of long range.

Because the short-range potential is very steep, it is reasonable to assume that its contribution to the free energy is virtually the same as that of the infinitely steep hard-sphere repulsion. This assumption is part of a crucial step in making our theory work. We developed a mathematical criterion for specifying the diameter of the reference spheres, a criterion that takes into account the repulsive interactions of a real liquid.

The careful choice of a diameter for



POTENTIAL-ENERGY FUNCTION determines the force acting between two molecules. The force is manifested as a tendency of the molecules to adjust their separation to whatever distance minimizes the potential energy. The potential for a system of hard spheres is represented by a vertical line at a separation corresponding to the diameter of the spheres and a horizontal line at all greater separations. The vertical line signifies that the energy becomes infinite when two molecules are pressed closer than their own diameter; in other words, an infinite repulsive force prevents the molecules from overlapping. The horizontal line indicates there is no force between the molecules when they are separated by more than their own diameter. A real potential-energy function differs from the hard-sphere potential in two ways. First, the repulsive part of the potential is not quite vertical, so that some overlapping of molecules is possible. Second, there is a region of negative potential energy, where the force is attractive.



SOLID AND FLUID PHASES of a substance can both be described accurately by a model in which a single potential-energy function defines the interactions of the molecules. Here the relation of pressure to volume is graphed for argon; the dots represent experimental data and the continuous lines are the results of a computer simulation based on the potential function.

the hard sphere is important because real molecules are softer than hard spheres and their potential-energy functions are "fuzzy." In effect the diameter of a real molecule is not a fixed quantity but depends on the energy with which molecules collide. The effective diameter must therefore be described by a weighted average of the repulsive potential-energy distribution, so that the thermodynamic properties of the hard-sphere system correspond to the statistically averaged properties of a real system. This effective hard-sphere diameter is not a constant, nor is it a function of density; instead it depends on the temperature, or the average energy of the molecules. By altering the diameter, this mechanism allows the simple but somewhat unrealistic hard-sphere potential to mimic the softer potential that governs real molecules.

The second term in our equation of state is an expression for the total energy of the intermolecular potential averaged over all molecular arrangements in the hard-sphere reference fluid. This term corresponds to the internal pressure in the van der Waals equation of state and can readily be determined from the radial distribution function of the hard-sphere fluid. The effect of the term is mainly to lower the energy and pressure of the liquid.

Neither of these terms reflects the fact that attractive forces between molecules of the kind hypothesized by van der Waals cause the structure of a liquid to differ from that of an idealized reference fluid of the same density. The attractive forces pull the molecules into regions of lower energy and thereby reduce their free energy. At high densities the effect is small because of the close packing of the molecules. At lower densities, however, the change in the entropy becomes appreciable. A third term in our equation accounts for the low-density effects and appears as a slight entropy difference between the real fluid and the reference fluid. With the addition of these three terms our perturbation sum yields satisfactory results for the free energy of a fluid. The theoretical model is in excellent agreement with experiment.

Van der Waals, in accepting the Nobel prize in physics for 1910, commented on his introduction of the internal pressure into his theory of liquids: "I confess that this is a detour. Perhaps there is a direct way. In the search for this way Gibbs's *Elementary Principles in Statistical Mechanics* will be a necessary guide." The derivation in the 1960's of a theory similar to van der Waals's result, and the continuing clarification and refinement of that theory through perturbation theory and statistical mechanics, may be seen as a completion of the program van der Waals began in the 1870's.

Games for Thinkers

from WFF 'N PROOF publishers

WFF 'N PROOF

2 to 4 players

$16.00*

The Game of Modern Logic, By Layman E. Allen, (Professor of Law and Research Social Scientist, University of Michigan). The original game of symbolic logic. Twenty-one-game kit that starts with speed games that challenge intelligent adults. The kit provides entertainment and practice in abstract thinking relevant for philosophy, mathematics, English, and computer programming for Jr. and Sr. high school.

QUERIES 'N THEORIES

For teen-agers and adults

2 to 4 players

$16.00*

The Game of Science & Language. (A Simulation of Scientific Method and Generative Grammars), By Layman E. Allen and Joan Ross (University of Michigan) and Peter Kugel (Massachusetts Institute of Technology). Through inductive reasoning, players learn scientific method of inquiry and gain skill in organizing, analyzing, and synthesizing data while engaged in an intriguing game of linguistics.

THE EMPHASIS IS NOT ON WHAT TO THINK, BUT HOW TO THINK!

EQUATIONS

2 to 4 players

$13.00*

The Game of Creative Mathematics, By Layman E. Allen. Our most popular game of mathematics. Five-game kit for use in intermediate grades through high school. Includes arithmetic operations such as addition, subtraction, multiplication, division, exponents and radicals in a variety of number bases. Like chess, the game can be as simple or as complex as the players make it.

ON-SETS

2 to 4 players

$13.00*

The Game of Set Theory, By Layman E. Allen, Peter Kugel, (M.I.T.), Martin Owens (Mitre Corp.) The game of set theory. This 30-game package can instruct students in kindergarten through high school in basics of set theory as well as extend a challenge to any adult exploring "new math." Encourages players to enjoy mathematics while learning concepts of union, intersection, logical differences, complement, identity, inclusion, null and universal sets.

THE PROPAGANDA GAME

2 to 4 players

$13.00*

By Lorne Greene (N.B.C., Paramount Studios), and Robert Allen (Director of Academic Games, Nova University). Students learn the fascinating techniques used by professionals to influence public opinion by learning to recognize "bandwagon" appeals, faulty analogy, out-of-context quotes, rationalization, technical jargon, emotional appeals, and many more. Particularly fascinating for social studies classes, English, problems of democracy, debate.

ON-WORDS

2 or More Players

$13.00*

The Game of Word Structures, by Professors Layman E. Allen and Frederick L. Goodman and Doris Humphrey and Joan Ross, Research Associates (University of Michigan). An instructional game that incorporates any aspect of words that players wish to introduce. Players practice and learn spelling, counting, grammar, phonetics, word roots, inflectional endings, prefixes and suffixes.

WFF'N PROOF, 1490-ED South Blvd., Ann Arbor, Mi. 48104

____a.	WFF 'N PROOF (logic)	$16.00*
____b.	QUERIES 'N THEORIES (science & language)	16.00*
____c.	EQUATIONS (mathematics)	13.00*
____d.	ON-SETS (set theory)	13.00*
____e.	PROPAGANDA (social studies)	13.00*
____f.	CONFIGURATIONS (geometry)	7.75*
____g.	ON-WORDS (word structures)	13.00*
____h.	TRI-NIM (problem solving)	6.75*
____i.	REAL NUMBERS (arithmetic)	2.75*
____j.	WFF (beginner's logic)	2.75*
____k.	QWIK-SANE (puzzle)	2.75*
____l.	TAC-TICKLE (pure strategy)	2.25*
____m.	THE MEDITATION GAME	2.25*
____n.	TEACHERS MANUAL	2.00*
____o.	THINKERS BOOKENDS	19.00*
____	5-game Special (a-e)	64.95*
____	6-game Special (a-f)	69.95*
____	7-game Special (a-g)	79.95*
____	13 GAME SPECIAL (a-n) & Teachers Manual	
	With Bookends	119.50*
	Without Bookends	106.50*

Prices include postage and handling.

My check is enclosed for $____________. Send items checked.

Name______________________________

Address____________________________

City______________ State__________ Zip__________

The Color Patterns of Butterflies and Moths

The wing patterns are mosaics of tiny scales. A few simple rules guide the development of more than 100,000 different patterns and provide an approach to the study of how animals develop

by H. Frederik Nijhout

Few things in nature match the beauty and the variety of the patterns on the wings of butterflies and moths. This order of insects—the Lepidoptera—consists of some 100,000 species, and virtually every one of them can be distinguished from the rest solely by the color pattern of its wings. The phenomenon is even more remarkable when one examines how the patterns are formed, as I have done over a period of years, and finds that the answer is essentially rather simple.

Biologists have long puzzled over the physiological basis of these highly diverse color patterns because they exemplify a fundamental problem of animal morphology. What are the processes regulating the development of the stereotyped spatial patterns of tissues, organs and appendages that characterize a given species of animal, and what are the essential differences in these processes that give each species its characteristic appearance? There is reason to believe the development of color patterns can be achieved, at least in principle, by the same kinds of processes that guide the development of morphological features, because all development is ultimately the outcome of progressive changes in the expression of genes. In color patterns this expression results in the localized synthesis of relatively simple pigment molecules.

The scales of the lepidopteran wing (Lepidoptera means scaly wings) are quite small and form a layer over the membrane of the wing. Each scale is a flattened outgrowth of a single cell and is about 100 micrometers long and 50 micrometers wide. (A micrometer is a thousandth of a millimeter.) The scales overlap like roof tiles and completely cover the membranous surface of the wing at a density of from 200 to 600 per square millimeter. Anyone who has ever handled a butterfly by the wings knows that the scales rub off easily and that the color pattern goes with them. It turns out that the membrane of the wing is colorless and translucent.

When one examines a wing at high magnification, it becomes evident that the color pattern is a finely tiled mosaic. Each tile of the mosaic is a scale, and each scale is of a single color. The typical pattern consists of only a few colors: three, four or five is the rule. Fine gradations in color and hue are achieved by variations in the ratios of the different scales, in much the same way as the color photographs accompanying this article are reproduced by an array of fine dots of only four colors. For example, the pattern of the forewing of the buckeye butterfly is made up of scales of only four colors: black, buff, red and brown. The white scales at the center of the large "eyespot" are colorless; their whiteness is due to the reflection of light by tiny air bubbles inside the scales. Although each scale appears to have only one pigment, the amount of pigment can vary. In the buckeye butterfly at least three intensities of brown scales can be detected on microscopic examination. Similarly, the buff and red scales appear to have the same pigment, but it is much more dilute in the buff scales than it is in the red ones.

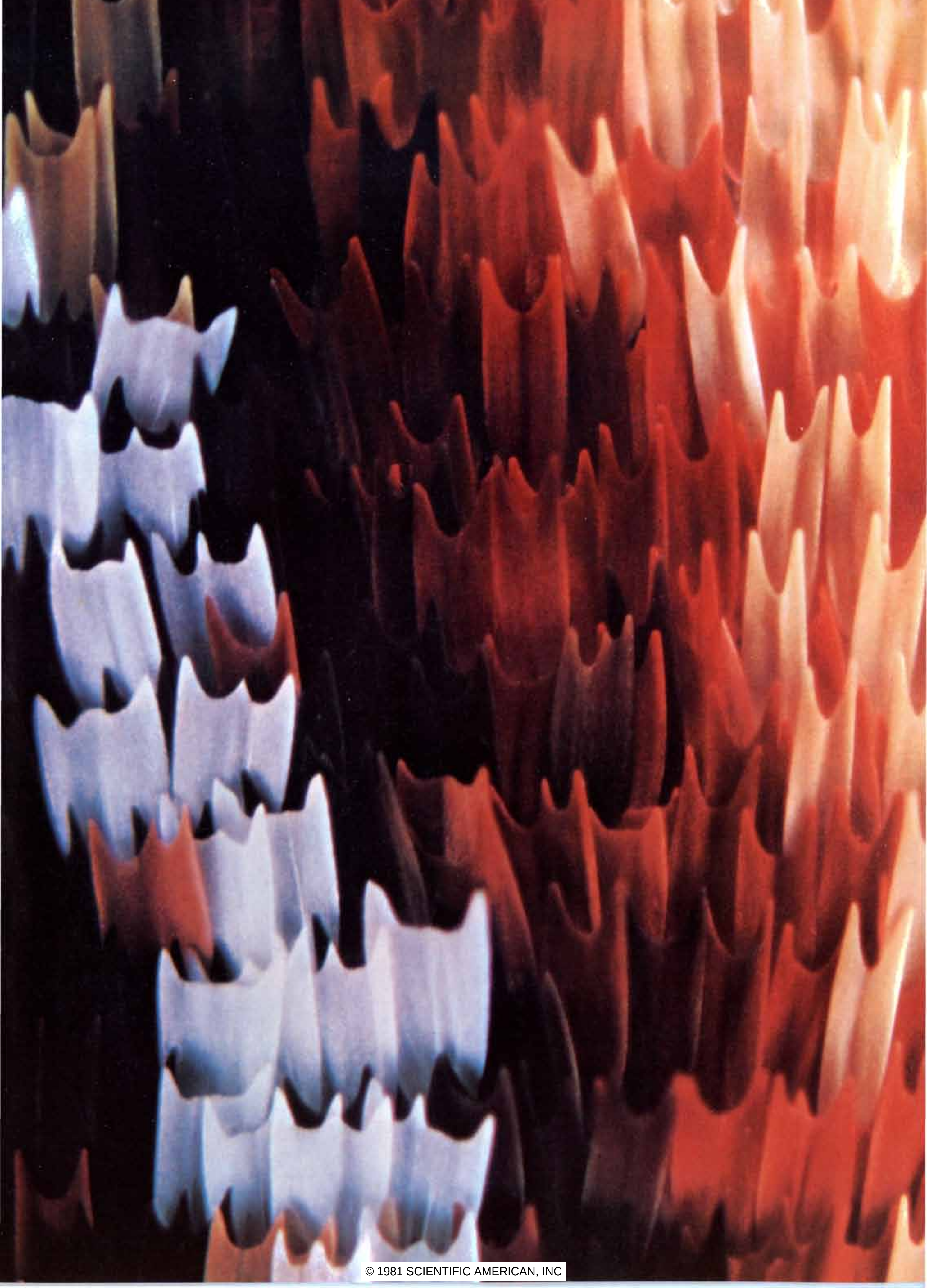
The colors of the scales of butterflies and moths can be of two kinds: pigmentary and structural. The pigmentary colors are created by only a few types of chemical. The vast majority of colors appear to be due to melanins. All blacks and grays and most tans, browns, brownish reds and yellowish colors are forms of melanin. The next-largest category of wing pigments consists of pterins, which account for many of the bright reds, oranges and yellows and a few of the whites. A number of unidentified pigments, some of which have been characterized as flavones, carotenoids and ommochromes, have also been detected in a few species; the colors they account for are yellowish to brownish.

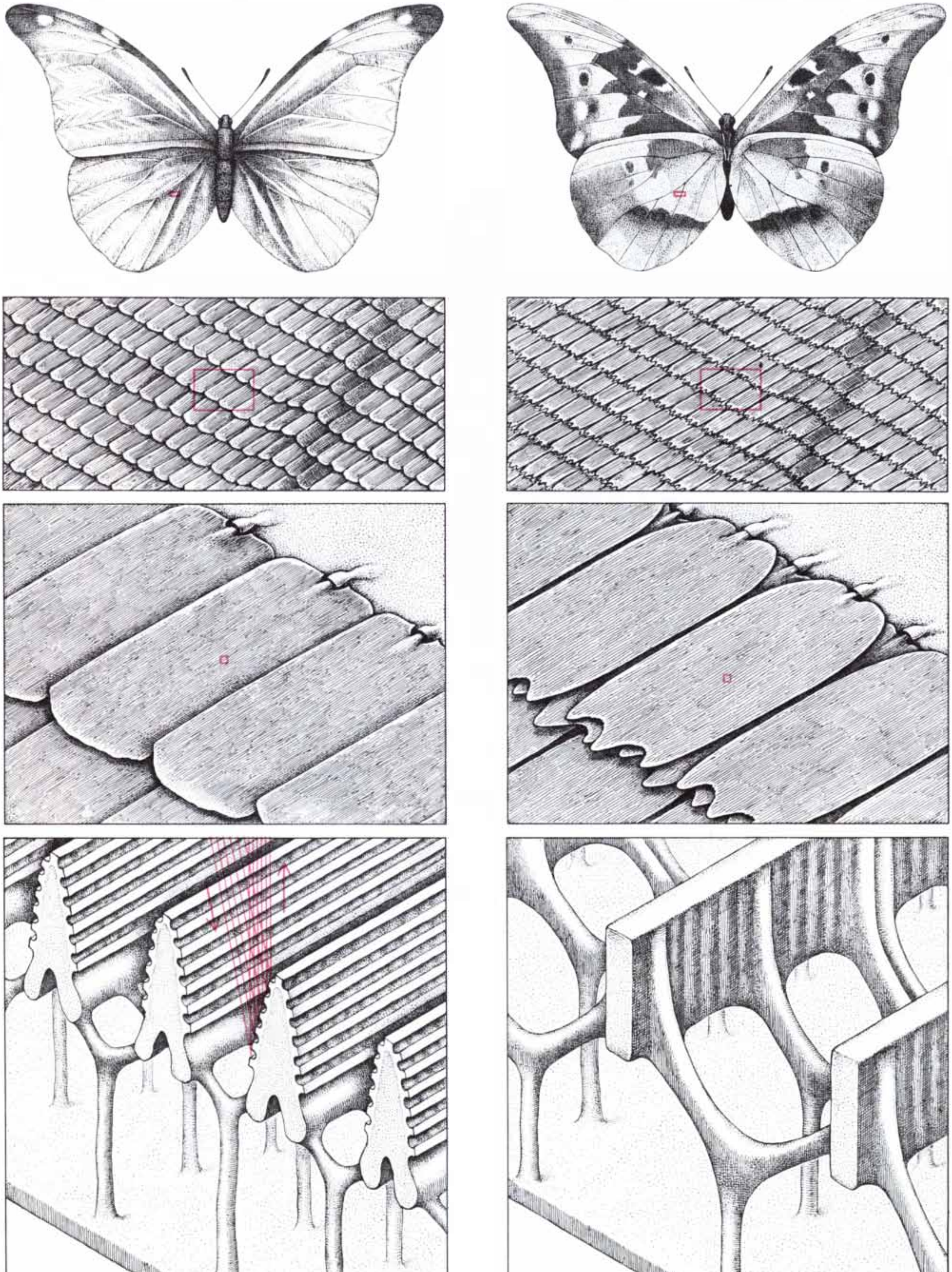
Some of the most striking colors are due not to chemical pigments but to small structural features of the scales. I have mentioned the eyespot whites. In addition all the iridescent or metallic colors, all blues and the greens of butterflies (but not the greens of moths) are structural. These colors come about through the interference, diffraction or scattering of light.

Iridescent scales owe their color primarily to the interference of reflected light. Depending on the species, interference colors arise when light is reflected by horizontal ridges on numerous thin vertical vanes on the surface of the scale or by horizontal stacks of thin lamellae that make up the outer cuticle of the scale. Iridescent scales also bear a pigment, usually a dark melanin, that absorbs much of the transmitted light and causes the reflected component to appear particularly brilliant.

It turns out that the diverse patterns of colors on lepidopteran wings follow a rather straightforward set of rules of design. The rules were first put forward in the 1920's independently and almost simultaneously by B. N. Schwanwitsch of the University of Petrograd (now Leningrad) and Fritz Süffert of the University of Freiburg. They found that a

WING SCALES of a large moth, *Hyalophora cecropia*, are depicted at an enlargement of 150 diameters in the photomicrograph on the opposite page. The area is close to the "eyespot" that appears near the outer end of each of the insect's two upper wings. The white scales are in fact colorless; their whiteness is attributable to the reflection of light by tiny air bubbles inside the scales. The scales ranging in color from reddish brown through red to buff derive their color from the fact that they have different amounts of the same ommochrome pigment.





great many color patterns can be derived from a single archetype, or ground plan. It is usually called the nymphalid ground plan because both Schwanwitsch and Süffert first recognized it in members of Nymphalidae, a large family of butterflies.

Over the years it has become clear that from this ground plan one can derive most of the wing patterns of the major families of butterflies and moths. It is important to note, however, that the ground plan should not be seen as representing an ancestral color pattern. It is a purely hypothetical pattern from which the patterns of real species can be obtained by the proper manipulations.

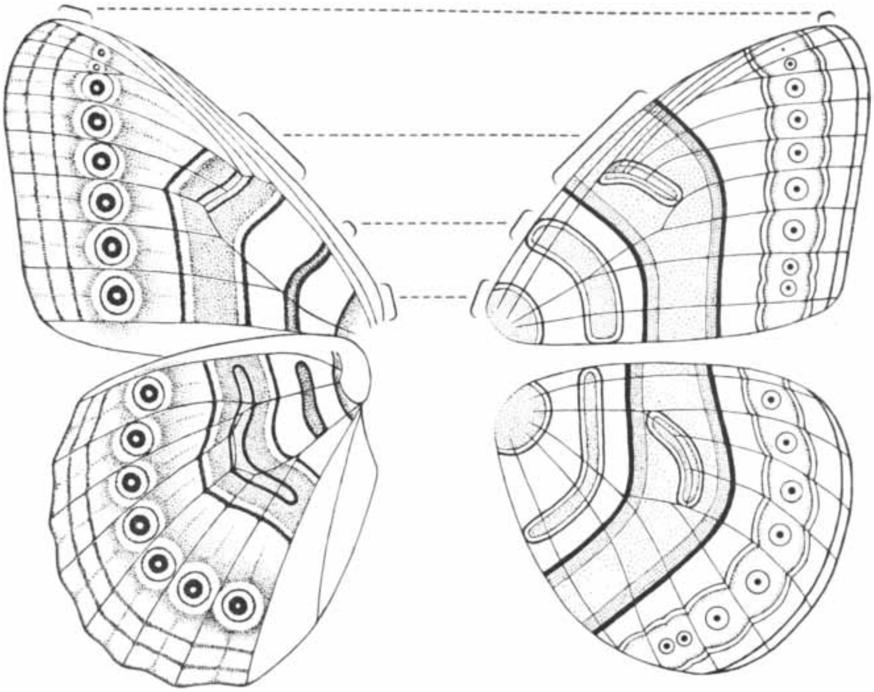
The salient features of the nymphalid ground plan are as follows. Along the margin of the wing is a row of border ocelli, or small eyespots. They are made up of a series of concentric circles with their centers always situated on the midline of a wing cell. (The term wing cell refers to an area of wing membrane that is bordered by veins. It should be distinguished from the scale cells.) The ground plan has one ocellus in each of the peripheral wing cells.

In the middle region of the wing are a number of colored bands that run from the front margin of the wing to the rear margin. This is called the central symmetry system. It consists of two sets of pigment bands that lie to the left and right of an imaginary line running roughly through the middle of the wing. The color sequence on one side of the line is the mirror image of the sequence on the other side. The colors of the bands are therefore symmetrical with respect to the midline of the wing.

A prominent pigmented stripe, the discal spot, is usually visible at the midline of the central symmetry system. Its location normally serves to define the axis of symmetry. In many moth species the discal spot is elaborated into a large round eyespot. Finally, one or more basal symmetry systems may be found near the base of the wing. They consist of matched sets of color bands as the central symmetry system does, although the basal systems are considerably less elaborate.

No species is known to display the entire ground plan. The plan is the maximum pattern from which actual patterns can be derived by the expression or exaggeration of some elements and the suppression of others. Border ocelli are usually present in only a few wing cells, and the proximal band of the central symmetry system (the band closest to the body) is often missing. Moreover, the bands of the central symmetry system can be quite irregular, and there is no discernible relation between the shapes of the distal members (those farthest from the body) and the proximal ones. Each element of the pattern can be expressed in a wide variety of pigments or structural colors to yield either high contrasts or subtle gradations of color.

The most fundamental modification of the ground plan entails lateral shifts of certain segments of the bands of the central symmetry system. Schwanwitsch called this phenomenon dislocation, because the visual effect is strikingly like dislocation in geology, where layers of sedimentary rock have been displaced along a fracture. In the Lepidoptera dislocation describes the interruption of a pigment band where it crosses a wing vein and the lateral shift of a substantial part of the band along the vein. Species-specific patterns of dislocation account for by far the greatest amount of variation in wing patterns. In many instances the effect of the dislocation is to generate a discontinuous array of colored dashes and dots. Dislocation also applies to the ocelli, which are not always perfectly lined up.



NYMPHALID GROUND PLAN, a conceptual system from which many of the color patterns of the wings of butterflies and moths can be derived, was worked out independently in the 1920's by B. N. Schwanwitsch of the University of Petrograd (Leningrad) and Fritz Süffert of the University of Freiburg. The plans are shown here in a somewhat simplified way, Schwanwitsch' on the left and Süffert's on the right, with similarities and differences indicated. About 75 percent of the color patterns of butterflies and moths can be derived from the plans.

A strong argument can be made for the view that many aspects of the design and development of lepidopteran color patterns can be readily understood in terms of current concepts in developmental biology. Over the past decade many developmental biologists have come to accept the suggestion by Lewis Wolpert of the Middlesex Hospital Medical School in London that in order for development to proceed properly cells must come to have information about where they are [see "Pattern Formation in Biological Development," by Lewis Wolpert; SCIENTIFIC AMERICAN, October, 1978]. Therefore a mechanism must exist by which cells become aware of their position in a developing organ or tissue, thereafter utilizing this positional information to turn on or off whatever biochemical processes are specified by their genes for cells in that location. For the development of color patterns this would mean that a scale cell must learn where it is on the wing if it is to manufacture the right pigment for that location.

FINE STRUCTURE of the scales of *Morpho rhetenor*, a South American butterfly, is portrayed on the opposite page in a series of drawings at an increasing scale. The drawings at the left depict the top side of the wing, which in the male of the species is a metallic blue. The color is structural, that is, it arises from the interaction of light with the wing scales. The blue scales owe their color to the fine structure of vanes on their surface. The vanes have horizontal ridges whose top and bottom surfaces are spaced about .22 micrometer apart; that is approximately half of a wavelength of blue light, so that the blue wavelengths interfere constructively as light is reflected from the surfaces. The drawings at the right depict the underside of the butterfly's wing, which is brown. This color arises from pigments in the skeleton of the scale.

In developing his theory of positional information Wolpert suggested that since cells can only know where they are with respect to something else, every developing system must have at least one reference point. He did not specify the biochemical or physiological nature of these points, and a good deal of theoretical effort has been expended to define the kinds of processes or phenomena that could play such a role. Almost all the models that have been developed propose the existence of a gradient of some kind: a physical factor or a chemical substance that changes gradually in dimension or concentration within a developmental field. The easiest models to visualize propose the existence of discrete sources of a diffusible chemical: a morphogen. The concentration of the morphogen would be highest in the cells at and near the source and gradually lower in the cells progressively farther away. Depending on the concentration the morphogen could differentially affect the expression of certain genes in the cells through which it diffuses.

Accepting this idea for the moment, suppose the affected genes code for enzymes in the biosynthetic pathways of various pigments. The simplest patterns that can be generated in the circumstances are rings and circles. All that is required is for cells finding themselves at or within some critical distance from the source of positional information to differentiate in the same way. In this instance all such cells would synthesize the same pigment.

I have already noted that circles and rings of pigments are common features of the color patterns of Lepidoptera, appearing as border ocelli and discal eyespots. Perhaps, then, one might be able to employ the idea of positional information as a model to approach the analysis of eyespot development. In this model the development of a circular eyespot would depend somehow on a specialized activity of the cells at its center. (It is worth noting that almost all eyespots have a clearly marked central spot, usually of a lighter color than the surrounding circles of pigment.)

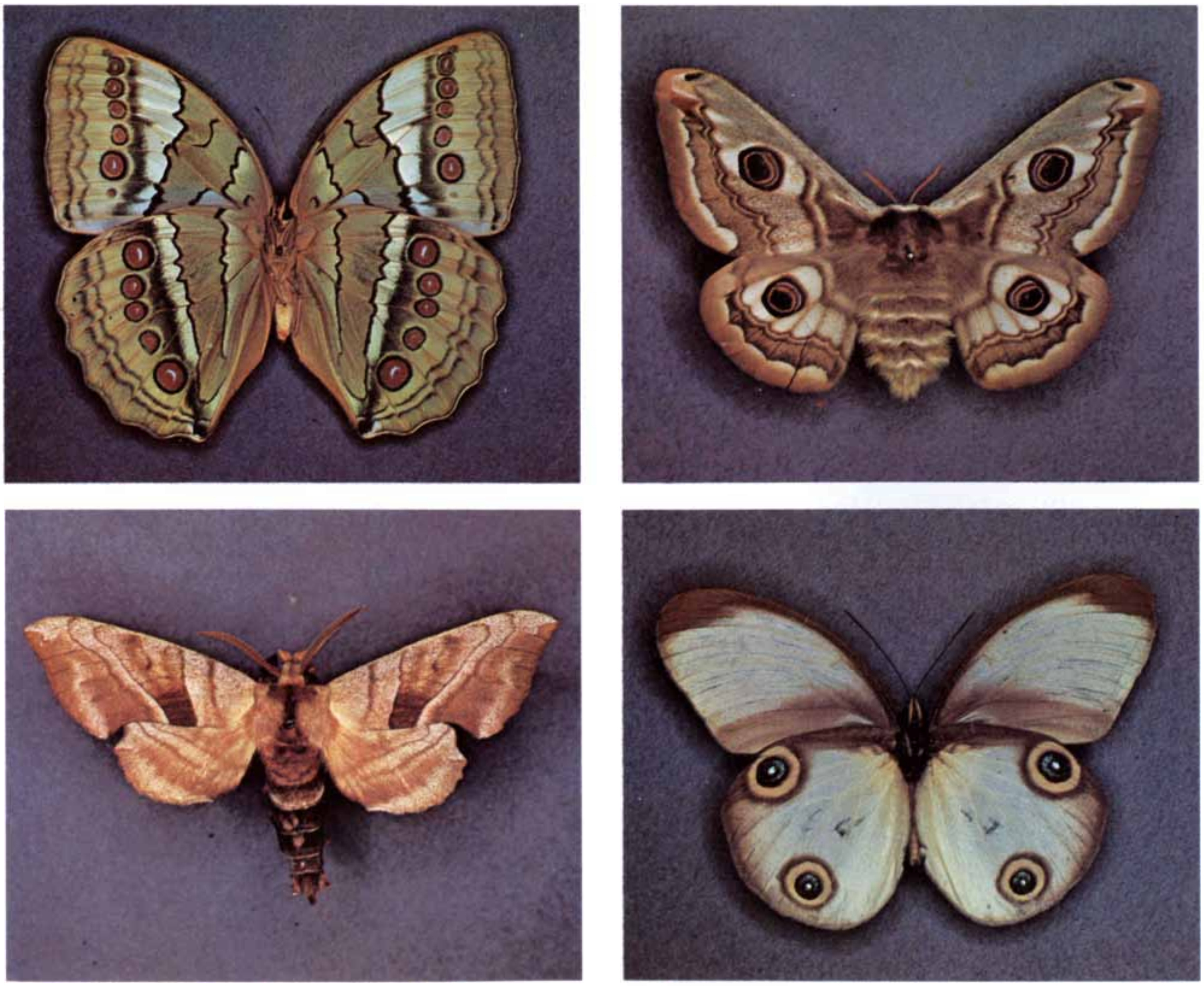
It should be obvious that the physiological processes determining the location and shape of the color pattern must operate while the wing is still unpigmented. Therefore one of the most important practical problems in doing experimental work on the determination of the color patterns is to be able to predict exactly where on the wing a pattern element will develop. For some years I have been working on the development of the color pattern in the buckeye butterfly (*Precis coenia*). This species is particularly useful for such analysis because it has a number of landmarks on the wing cover of its pupa, the developmental stage before the insect emerges as an adult. With the aid of such markings the position of several parts of the presumptive color pattern on the underlying unpigmented wing can be accurately predicted long before the patterns actually develop.

The presumptive position of the central spot of the large eyespot on the forewing is particularly easy to identify. Hence it is possible to penetrate the pupal cuticle with a fine cautery needle and kill the small group of cells at the center of the eyespot. If this is done early enough in development, say a few hours after pupation, when the wing has just formed, the development of the eyespot is totally and specifically inhibited. Similar cautery done 48 hours or more after pupation has no effect. When the cells are killed at various times up to 48 hours, intermediate-size eyespots develop. The wing does not become pigmented until the fifth day of the pupal stage (about 24 hours before the emergence of the adult butterfly); therefore pattern determination precedes pattern differentiation (the synthesis of pigment) by about three days.

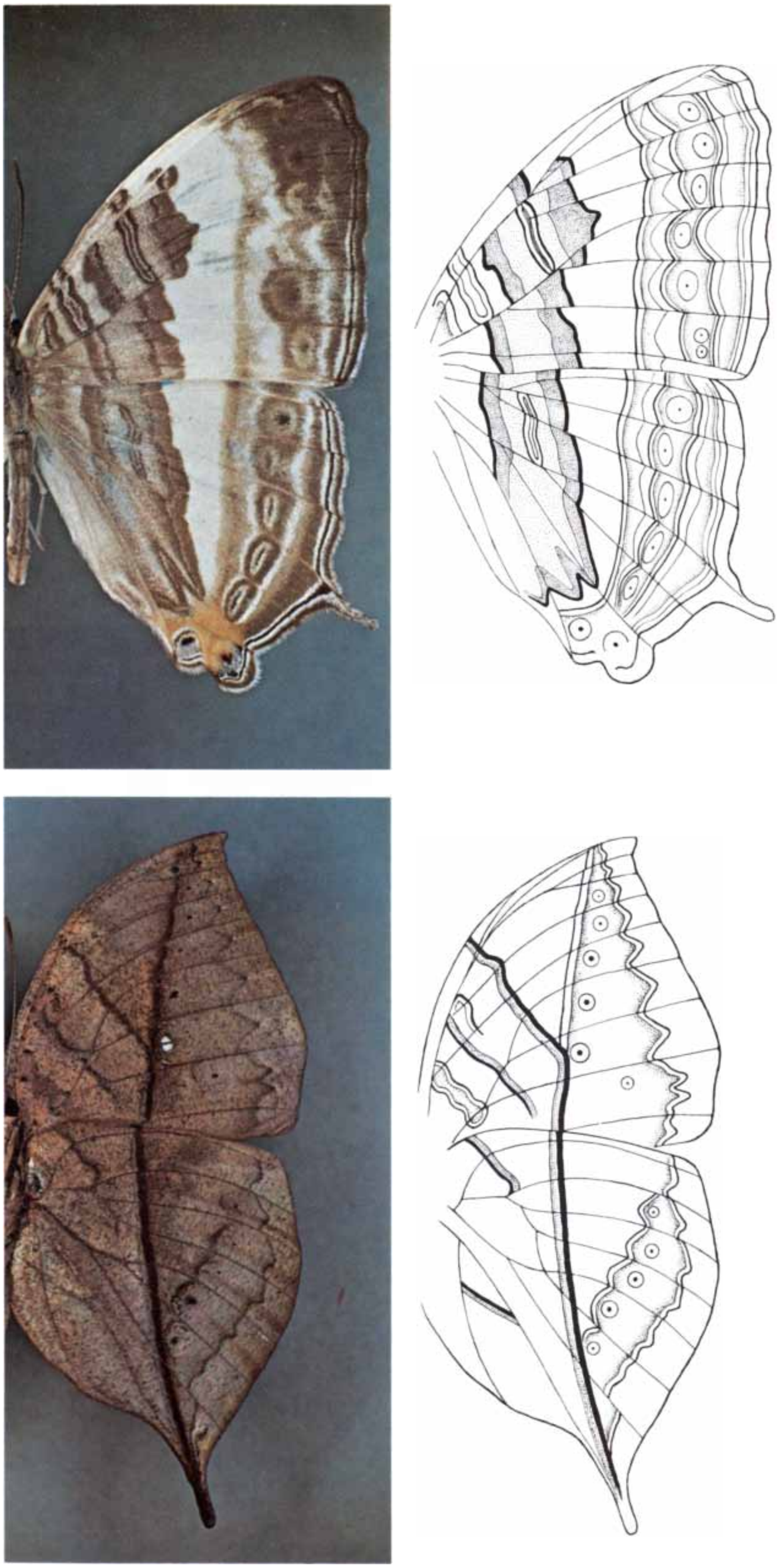
If eyespot development depends on some special property of the cells at the center of the spot, it should be possible to transfer this property to other locations on the wing by transplanting the cells. Such a transfer proves to work. When I transplanted a small piece of epidermis from the area that would have formed the center of an eyespot to a different part of the wing (from which a piece of epidermis of similar size had been removed), I found that a circular pigment pattern developed around it. Epidermis transplanted from other regions of the wing did not give rise to such a pattern. It therefore appears that eyespot determination is the intrinsic property of a small number of cells (perhaps only a single cell) and that they are one of the sources of the positional information postulated by Wolpert, at least for one element of the color pattern. I have named such a group of cells a focus.

I have mentioned that the forewing of the buckeye has only four pigments, all melanins. Although it is not known what the molecular differences are among these differently colored substances, there is reasonable evidence that each one needs for its synthesis from a simple precursor only a single type of enzyme, a tyrosinase. It also appears that tyrosinases with different specificities may serve for each of the four colors of melanin. If they do, the action of the focus would be, at least in principle, a simple one. The focus would somehow cause the scale cells in its domain to switch from the production of, say, a yellow-melanin tyrosinase to a black-melanin tyrosinase.

It is likely that all eyespots are organized around such foci and that the system of border ocelli develops around a



DERIVATIVES OF GROUND PLAN are evident in these photographs. *Stichophthalma camadeva,* a butterfly (*upper left*), comes the closest of any species to possessing all the elements of the ground plan. *Saturnia carpini,* a large silk moth (*upper right*), has a clear central symmetry system and a well-developed discal spot but only a hint of border ocelli ("eyespots"). *Cressonia juglandis,* a hawk moth (*lower left*), has a more subdued pattern, seen in many moths, that clearly reveals a central symmetry system. *Taenaris rothschildia,* a tropical butterfly (*lower right*), has a much reduced pattern made up of border ocelli that have developed into disproportionately large structures.



DEPARTURES FROM GROUND PLAN indicate the flexibility of the system. Most of the variability arises from dislocation: the interruption of a band of pigment where it crosses a wing vein and the lateral shift of the band along the vein. Two examples are depicted here, each by means of a photograph of one wing and a companion drawing. The butterfly *Cyrestis cocles,* shown at the top, exhibits clear but minor dislocations of the bands of the central symmetry system. The dead-leaf butterfly *Kallima inachus,* shown at the bottom, has evolved a pattern resembling a dead leaf on the jungle floor, complete with fungus spots and leaf veins.

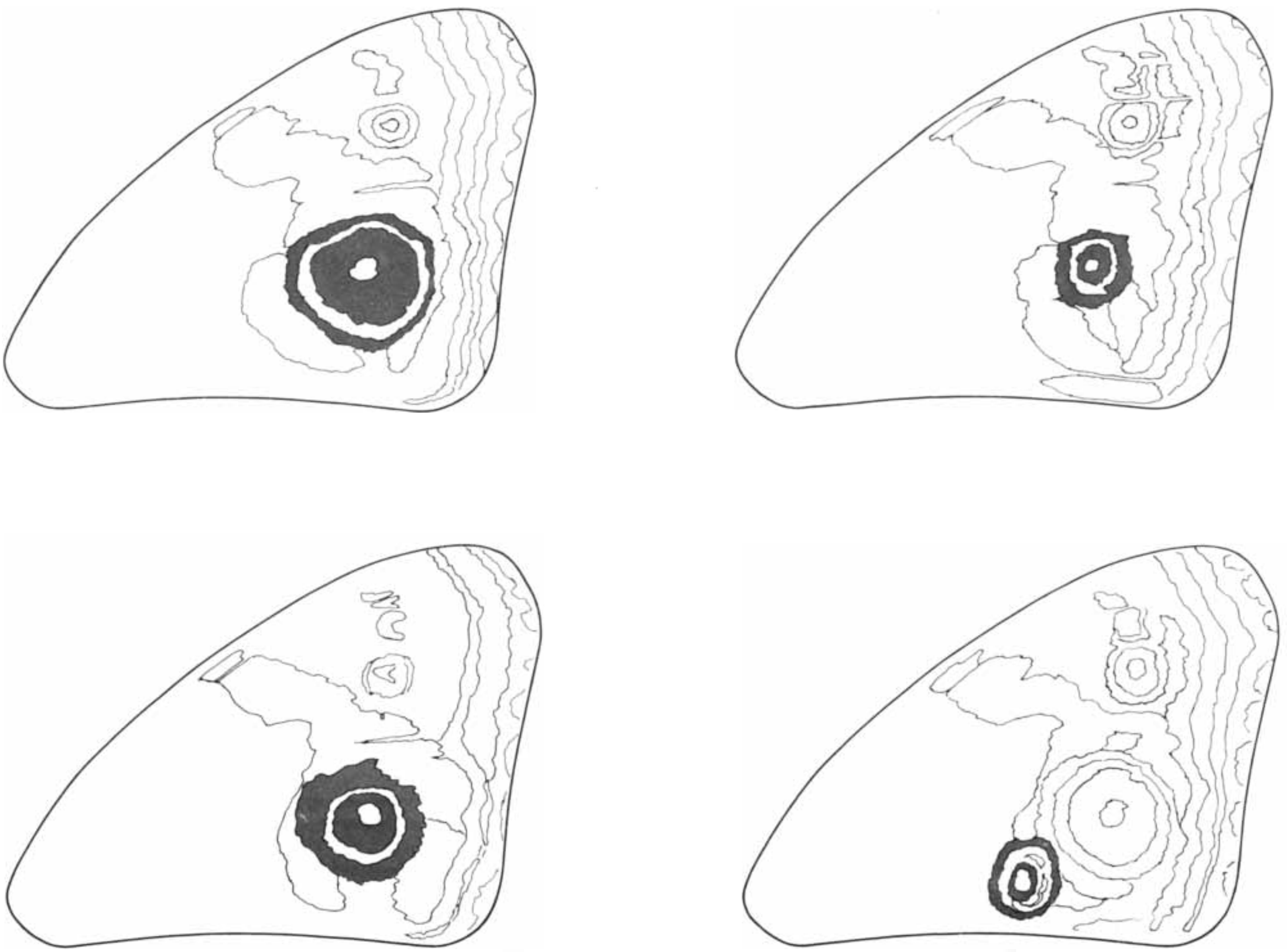
row of foci. In the nymphalid ground plan there would then be one focus per wing cell, with each focus situated on the midline of the cell.

There is good reason to believe the central symmetry system is also generated around a row of foci. In many species this system consists of a series of merged concentric circles. The centers of the circles, like those of the border ocelli, lie on the midline of the wing cells. Moreover, there is good experimental evidence for the existence of functional foci in the central symmetry system. In the 1930's Alfred Kühn and Melitta von Engelhardt of the University of Göttingen did a series of experiments in which they studied the effects of cautery at various places in the central symmetry system of the Mediterranean meal moth (*Ephestia kühniella*). By analyzing the defects in the color pattern that developed after cautery they showed that the central symmetry system of this moth is generated by at least two and probably three "sources of determination," one near the forward margin of the wing, one near the rear margin and perhaps one near the center of the wing. Similar experiments by other investigators with other species of moth have confirmed these findings.

Like the central symmetry system, the basal symmetry systems are often broken up into a short series of merged concentric circles, each with its center situated on the midline of a wing cell. It therefore appears that the nymphalid ground plan is best interpreted as being organized around three rows of foci that run from the forward margin of the wing to the rear margin. A distal row of foci controls the development of the border ocelli, a central row the central symmetry system and a proximal row the basal symmetry system. The central symmetry system and the border ocelli are merely different manifestations of basically similar processes.

The bands of a symmetry system can be thought of as originating from the fusion of the peripheral rings of large ocelli. One trouble with this idea is that the bands of the central symmetry system are more often than not straight lines rather than segments of circles. The border ocelli too are often significantly noncircular.

Before I take up the development of noncircular patterns around a focus it will be useful to summarize what I perceive to be the rules governing pattern development in the lepidopteran wing. As far as is known the patterns on the upper and lower surfaces of the wing develop independently. Pattern elements develop around a finite number of foci. In most cases a focus determines only the pattern of the wing cell in which it is situated. The color pattern of each wing cell is thus determined independently of the patterns of other wing



TWO EXPERIMENTS indicate how the development of an eyespot is controlled in a butterfly or a moth. The experiments are depicted here in the buckeye butterfly (*Precis coenia*). The butterfly's normal pattern (*upper left*) is greatly changed if a small patch of cells at the center of the eyespot is killed by cautery shortly after pupation (the stage of development before the insect emerges as a butterfly). When cautery is done within four hours after pupation, the spot fails to develop (*upper right*). Cautery 24 hours later results in the development of a medium-size eyespot (*lower left*). Cautery more than 48 hours after pupation has no effect. If the central cells are not killed but instead are transplanted to another area of the wing, a pattern of circles develops (*lower right*). It is not a full eyespot because time is required for the tissue in the area of the transplantation to heal, leaving less than the 48-hour induction period required for a full-size spot to arise.

cells. Hence the color pattern is not only a mosaic of scales but also a mosaic of wing cells that are independent units of development. It is possible for one focus to determine the pattern in several adjoining wing cells, but only if they lack active foci.

Species-specific patterns within the nymphalid ground plan arise in part by variations in the number of foci that are actually expressed and their position on the wing. Lateral shifts of foci and the patterns they generate within each wing cell account for the phenomenon of dislocation. The development of species-specific patterns also depends on variations in the rules governing the shape of the pattern that develops around a given focus.

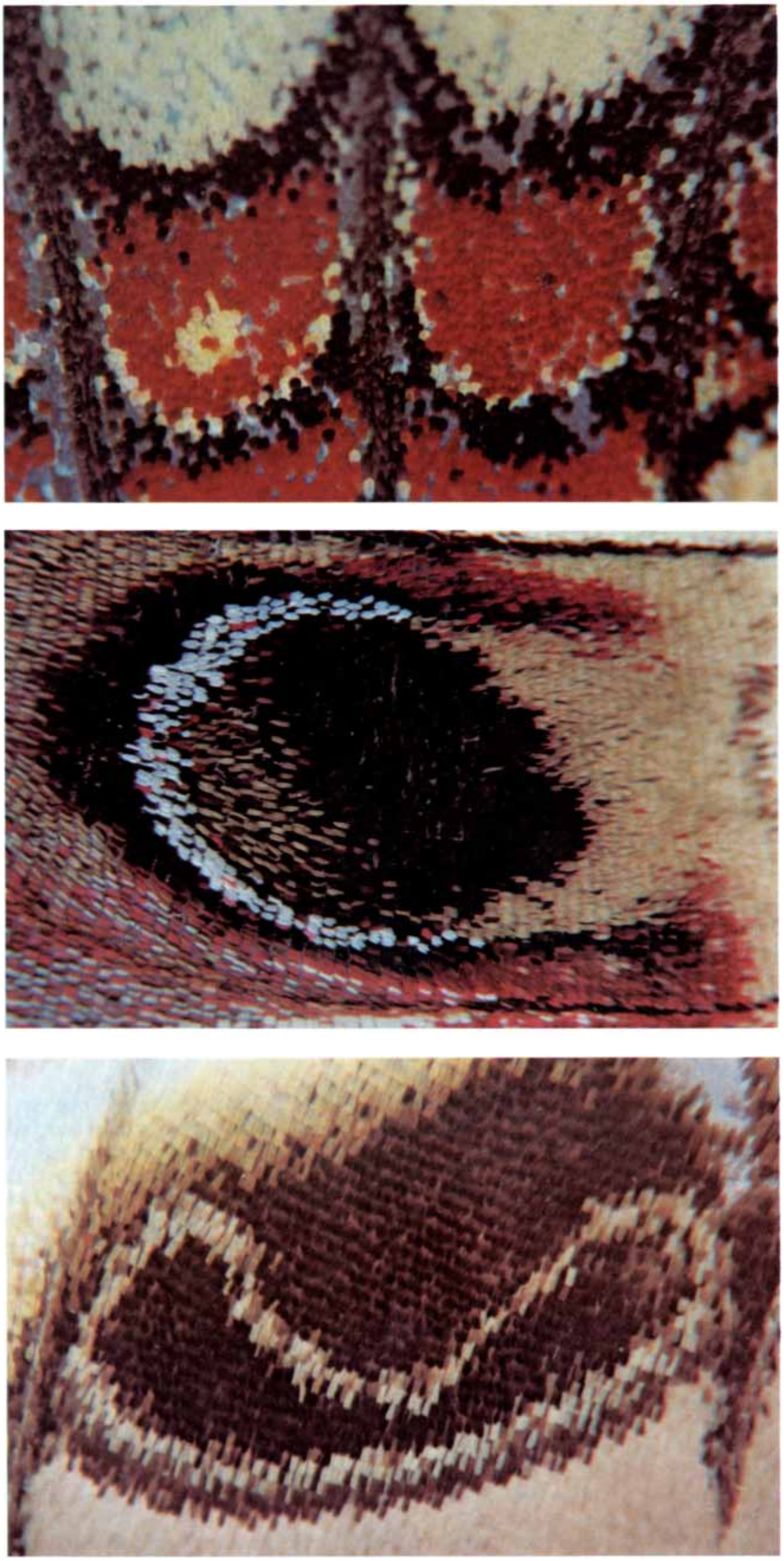
Although true circles are abundant on lepidopteran wings, they are still a minority constituent of the wing patterns. Border ocelli are often stretched into ellipses, bent into crescents or modified into more complex broken-up forms. Similar noncircular forms are observed in the bands of the central symmetry system. An interesting aspect of noncircular patterns is that they are almost always bilaterally symmetrical with the axis of symmetry running down the middle of the wing cell, parallel to the wing veins and through a focus. Noncircular patterns are usually confined to a single wing cell, as normal round ocelli are.

Two conditions are required for the development of circular patterns under the concept of positional information. First, the positional signal must be transmitted at the same rate and intensity in all directions away from the focus. Second, all the cells on the wing must follow the same rules in interpreting the strength of the positional signal. It is clear that at least one of these conditions must be violated when a noncircular pattern develops around a focus.

In its purest form Wolpert's positional-information concept assumes that the positional signal is invariant. A considerable body of observational and experimental evidence supports this view. The idea is that there may exist one positional signal or a few universal types of signal, and that differences in the genetic makeup and prior developmental history of the responding cells determine the pattern that actually develops. If this is the case, it follows that for the development of noncircular patterns the scale cells at different places on the wing must be following different rules to interpret the positional signal from the focus.

A number of processes can be visualized that yield bilaterally symmetrical patterns around a focus. At present I favor a model embodying an interaction of two gradients. The model has the virtues of being visually simple and, most important, testable.

To comprehend the model it is necessary to form a more concrete image of the positional-information signal that emanates from the focus. Think of it as a chemical substance, *P*, synthesized at the focus and diffusing equally in all directions. The three-dimensional concen-



NONCIRCULAR PATTERNS often develop around a focus (a source of positional information for cells that make pigment). From the top, examples appear in the butterflies *Euphydryas editha, Hyalophora cecropia* and *Rothschildia arethusa*. The development of such patterns indicates that the signal from the focus is not interpreted equally by all the cells around it.

tration gradient of P would be a cone. The highest concentration would be at the focus, and the concentration would decline gradually with increasing distance from the focus. The exact shape of the gradient depends on a number of factors, but since it does not affect the principle, I assume that the gradient is linear.

Suppose the scale cells are programmed to synthesize a black pigment wherever P is at a particular critical concentration. The result would be the synthesis of a ring of black pigment. This situation can be modeled by transecting the cone-shaped gradient of P with a horizontal plane and projecting the conic section thereby made onto the surface of the wing. The height of the plane corresponds to the critical concentration at which black pigment is coded for. Different pigments can be represented by different planes, and the width of a pigment ring is proportional to the thickness of the corresponding plane. The planes are then graphic representations of the rules by which the positional-information gradient is interpreted.

To account for patterns that deviate from circularity, given the same positional-information gradient, the scale cells on one side of the focus must be programmed to synthesize a given pigment at a higher concentration of P than scale cells on the opposite side. If a horizontal plane (in the three-dimensional representation) signifies that the gradient of P is interpreted the same way everywhere on the wing, a useful rule would be that the height of a transecting surface is proportional to the concentration of P at which a given pigment will be synthesized. An ellipse would then be generated by a sloping plane.

At this point it becomes tempting to think of such a transecting plane as a gradient of a second chemical substance, Q, made near the base of the wing cell and diffusing distally. It is possible then that pigment is synthesized only where a particular critical ratio exists between the concentrations of the two chemicals. With this concept it would not be necessary to have a different chemical to specify the position of each pigment in a pattern; a different critical ratio of the same two substances would achieve the same effect.

Unfortunately virtually nothing is known about the physiology of the development of noncircular patterns except that foci are involved. At the moment the existence of a second gradient is purely hypothetical, but the hypothesis does provide a direction for future investigations. If such a gradient exists, one should be able to manipulate it by removing its source surgically. It should also be possible with transplanted foci to probe different regions of this second gradient field and see whether the

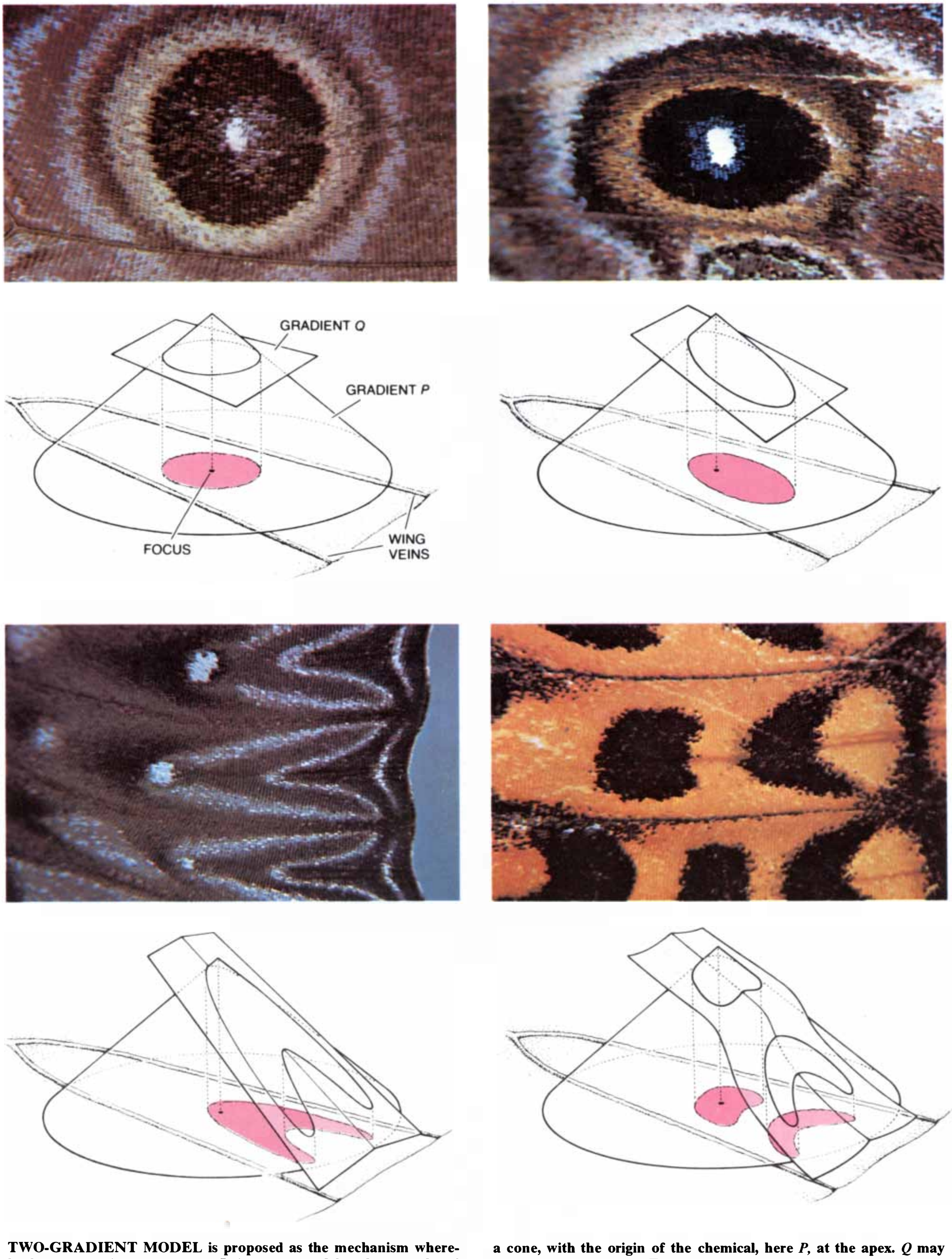
OVERALL PATTERN on the wing of a butterfly or a moth is a mosaic of independent patterns, one in each wing cell. (A wing cell is an area bordered by veins.) The pattern of *Smyrna blomfildia,* a Central American butterfly, is shown at the top and broken down by wing cells at the bottom. Patterns of the upper and lower surfaces of a wing also develop independently.

results match predictions about the shapes of patterns that would be generated around a focus transplanted "upstream" or "downstream."

In many lepidopteran species the border ocelli have lost all resemblance to closed circles or ellipses. They may look like arrowheads, chevrons or crescents, often with the convex side toward the focus and always bilaterally symmetrical with respect to the midline of the wing cell. Such patterns can be explained by the double-gradient model if one assumes that at least one gradient is nonlinear and that the gradient of Q has a ridge or high point running along the midline and sloping off symmetrically toward the wing veins. What this model in effect proposes is that at least two factors figure in the control of the determination of a color pattern and that they will be found to be distributed as the relative values of P and Q.

Experiments published by Kühn in 1926 offer indirect support for such a model. He exposed pupae of the European nymphalid butterfly *Argynnis paphia* to subfreezing temperatures during the time when the color pattern was being determined. The exact physiological effects of such temperature shocks are not known, but they result in an apparent locking in of the color pattern at various stages of development. The sequence of patterns thereby obtained can be accounted for by the two-gradient model if there is a progressive rise in the level of Q accompanied by a gradual spreading of the cone-shaped gradient of P during the early stages of pattern development, when the gradients are first established.

No additional factors appear to be needed to explain any of the thousands of different wing patterns based on the nymphalid ground plan. Individual species-specific patterns therefore seem to come about through the developmental manipulation of only a few variables: the number and exact location of foci; the relative shapes of the two gradients, one generated by the focus and the other somewhere in the proximal region of the wing cell; the choice of pigments in which the pattern is expressed, and the magnitude of the positional signals inherent in the two gradients that code for each of the pigments. Of all these factors only the existence of foci is known today. The small group of us working in this field are still a long way from understanding the nature of the positional gradients and the mechanisms by which they influence the synthesis of specific pigments in lepidopteran scales. At least, however, we now have a conceptual framework in which we can begin to ask sensible questions not only about the genesis of individual color patterns but also about the comparative physiology and evolution of biological patterns in general.



TWO-GRADIENT MODEL is proposed as the mechanism whereby border ocelli are generated. It is represented by photographs of the wing patterns of four species, each with a diagram below it in which the two gradients, *P* and *Q*, are represented as planes transecting a cone. The three-dimensional shape of a linear gradient (a chemical that diffuses equally in all directions in a wing cell) would be a cone, with the origin of the chemical, here *P*, at the apex. *Q* may be linear or nonlinear. The conic section created by the transecting planes is not the pattern on the wing; that arises from the projection of the conic section on the wing. The patterns depicted here increase in complexity. From left to right the species are *Morpho peleides, Smyrna blomfildia, Dichorragia nesimachus* and *Speyeria aphrodite.*

What if you chose as a technical

"At Sun, an HP computer system helps maximize production and minimize equipment failures."

Sun Production Company, a subsidiary of Sun Company, uses an HP 1000 computer system to control pump units and monitor production at its Eliasville, Texas, oilfield.

Marvin Boyd, District Production Manager, says, "We find the HP 1000 an effective management tool. By providing constant surveillance of our operation, it alerts us to malfunctions immediately and shuts units down before damage can occur.

"Because the HP system eliminates overpumping and speeds response time we can reduce operating costs and still achieve maximum production rates. Based on these results, we are now installing additional HP 1000s at Sun's Bennett Ranch and Levelland units."

HP can be your business computer partner, too!

A new kind of personal office computer, the HP 125 is specifically designed to help business people do their jobs better. Qualifications? Powerful management software, impressive graphics capabilities, communicates well with other computers, friendly and flexible. The HP 125 is supported by over 100 software programs and handles word processing just as easily as graphics. All this is supported by HP's service, reputation, and experience.

Computation as personal as your wristwatch.

There's virtually no numerical problem in engineering or statistics beyond the scope of HP's new programmable 11C. Included in addition to the commoner functions are hyperbolics, permutations and combinations, random-number generation. You also acquire the ability to redefine five key functions making a grand total of 15 program labels you can call up at the touch of one key. Subroutines can be nested

Hewlett-Packard computer partner?

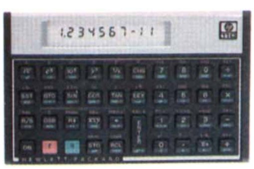
"At GCA, an HP computer tests IC manufacturing processes, and paid for itself in less than a year."

GCA is a leading international manufacturer of semiconductor production equipment. At its Burlington Division, in Bedford, Mass., semiconductor fabrication processes are investigated in a production-oriented R&D lab.

Says Dr. Thomas P. Shaughnessy, Manager of Process Engineering, "Our goal is to understand and quantify process requirements for future products. Our HP 9845 computer system is crucial to the collection, analysis, and graphical display of our experimental data.

"We find the 9845's interfacing capabilities, plus its powerful software options have let us cut manual procedures 10 to 20 percent, saving us $50,000 to $100,000 a year. HP 9845-generated graphics are used for reports to customers, making the computer an effective marketing tool as well."

four deep. Insert-delete editing of program steps is standard. This hand-holdable machine is supported by a library of thousands of applications programs. The HP-11C costs $135*. Also available, the HP-12C. It does for finance what the 11C does for engineering and statistics.



Company-wide computer network connects . . .

The Manufacturers' Productivity Network is a coordinated network of hardware and software that helps you collect, measure, control, process, and distribute information. From drawing board to laboratory, from factory floor to front office, MPN not only helps each area of your company run more efficiently, it ties them all together to give you greater control over your entire operation—and that's the first step toward improving your bottom line.

For your free 24-page, 4-color brochure *The HP's Manufacturer's Productivity Network*, call Dept. 306A toll free, (800) 547-3400 (except Alaska and Hawaii). Or write A. P. Oliverio, Vice-President, Marketing, Hewlett-Packard, 3000 Hanover Street, Palo Alto, CA 94304.

*Domestic U.S. prices only

When performance must be measured by results

HEWLETT PACKARD

00146

The Lining of the Small Intestine

It was long assumed to be a passive tissue. Recent work has shown that the cells of the lining are covered by a membrane that actively digests foods and speeds nutrients into the blood

by Florence Moog

Nutrients appear to enter the body through the mouth, but in a more profound sense nothing becomes part of the body until it has crossed the lining of the small intestine and entered the bloodstream. Most higher organisms, including all mammals, have the topological form of a tube within a tube. The alimentary canal extends from the mouth to the anus, with the body mass lying between this narrow tube and the broader one formed by the skin. The interior of the gastrointestinal tract is thus more closely connected to the outer world than it is to the muscles, nerves, skeleton and blood vessels of the inner organism. The crossing of this boundary is the fundamental event in the assimilation of nutrients.

It was once assumed that most of the work of digestion went on in the lumen, or interior space, of the intestine, and that the intestinal lining served as a passive membrane. It was thought that when food had been digested, nutrients merely diffused across the lining without the active assistance of the lining cells. Recent work in physiology presents a different picture of the lining. It is now clear that the inner surface of the intestine is active both in breaking down and in absorbing nutrients. Cells of the lining have enzymes embedded in their membrane that convert complex sugars into simple ones and break down the peptides produced by protein digestion into their constituent amino acids or into small peptides made up of a few amino acids each.

In addition the membrane of the lining cells appears to be equipped with a diverse array of transport systems. When digestion is complete, these carriers move sugars, amino acids and small peptides into the cytoplasm of the cells; from there the nutrients are absorbed into the blood. For each simple sugar and amino acid molecule there appear to be several transport mechanisms; most of them remain obscure. Elucidating these mechanisms is currently a task of great interest in the physiology of the intestine. No matter which one of several proposed explanations ultimately proves to be the most satisfactory, the concept of the intestinal lining as a passive membrane has been made obsolete. The lining has been shown to be an active tissue that digests nutrients, selects those to be admitted to the blood and speeds their movement into the body.

Where and how digestion and absorption take place are questions that have long interested biologists. Work on these problems has proceeded from crude but illuminating experiments, occasionally involving the experimenter's own digestive tract, to more refined investigations of the microscopic structure of the intestinal lining.

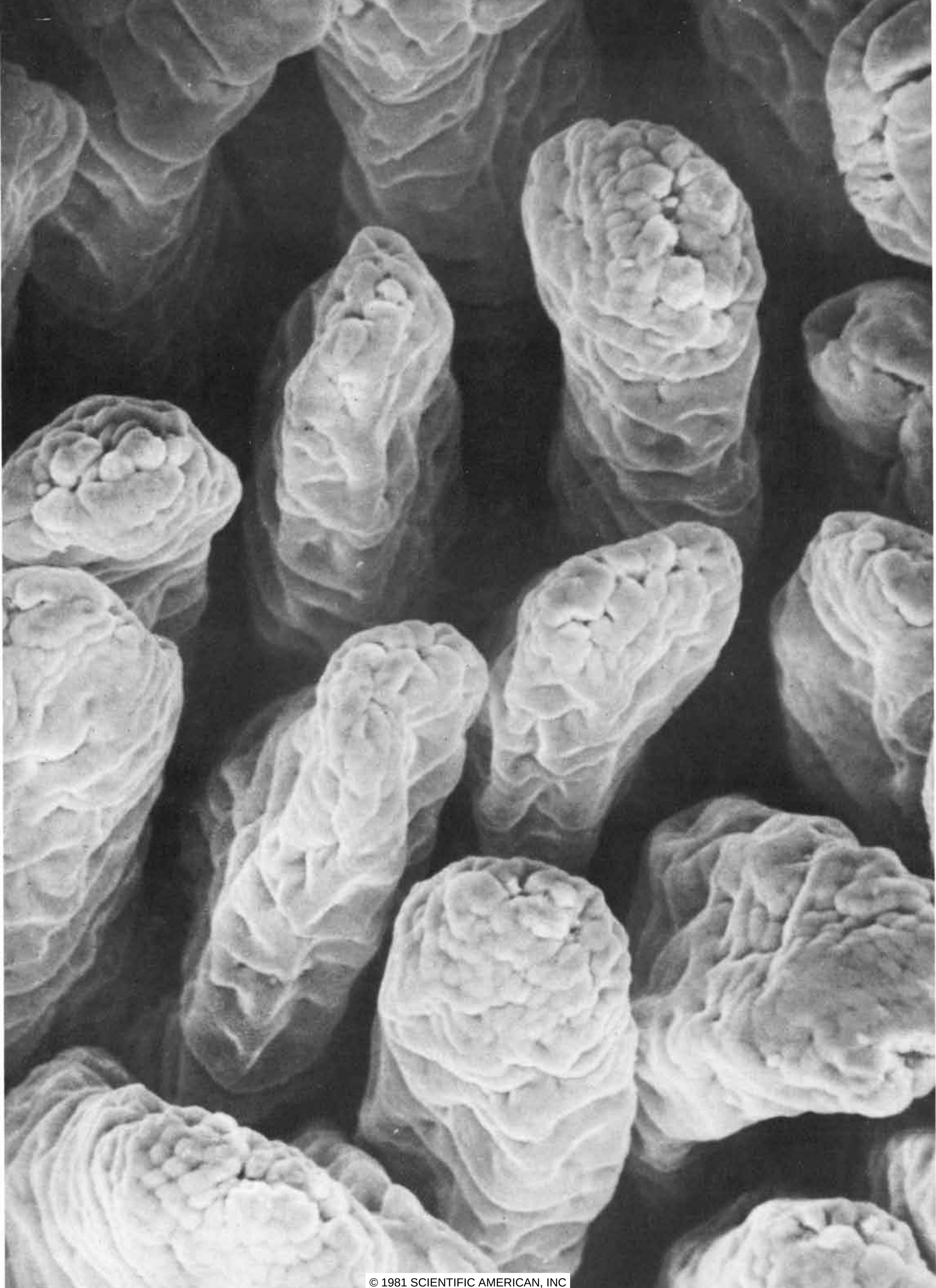
In the 1770's Lazzaro Spallanzani of the University of Pavia made several tests that helped to clarify the nature of digestion. He began by feeding his pet hawk a piece of meat on a string. When the meat was pulled out of the bird's stomach shortly thereafter, it was found to be smaller, having been partially digested. Spallanzani then swallowed three ounces of chicken meat in a cloth bag. When the bag was eventually expelled, most of the meat was gone. In a slightly more elaborate experiment Spallanzani had a hawk swallow a perforated tube containing a sponge. When the sponge was removed, a fluid was extracted from it that was capable of digesting meat.

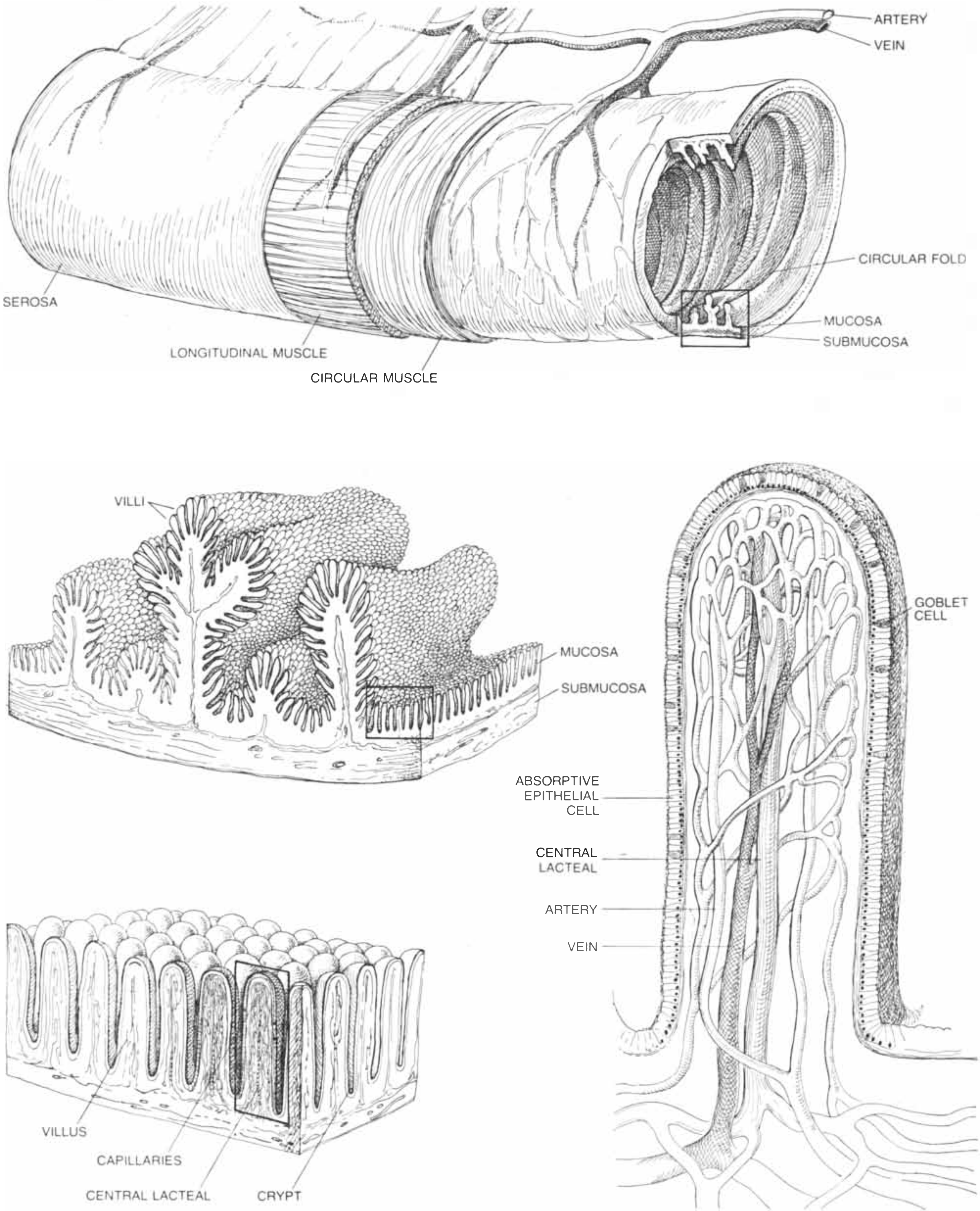
By means of these procedures Spallanzani had established that digestion is begun in the stomach by substances capable of penetrating cloth; he had even isolated some of these fluids in his sponge. By the middle of the 19th century it was known that digestion continues in the small intestine, where the active agents include enzymes secreted by the pancreas. Although it was understood that the small intestine participates in digestion, little was known about how nutrients are absorbed.

One explanation was suggested in the 1850's, when it was discovered that molecules in solution can diffuse through a semipermeable membrane from a region of higher concentration to one of lower concentration. After a meal nutrients might therefore pass from the richly concentrated medium of the lumen into the blood by passive diffusion. In the next two decades, however, it became clear that although passive diffusion might account for some of the absorptive properties of the intestinal lining, it could not explain all of them. By the 1870's it had been found that water can be absorbed into the blood across the intestinal lining from a segment of the intestine that has been filled with blood serum (the liquid portion of the blood). Passive diffusion cannot explain this movement because the concentration of water is roughly equal in blood and serum. In addition it had been noted that the six-carbon sugar molecules utilized in metabolic processes are absorbed more rapidly than those that are not of metabolic importance. Indications of selective absorption of fats and proteins were also observed.

Much of the subsequent research on the small intestine has consisted in uncovering the means by which physiologically useful molecules are digested and then transported across the intestinal

PROJECTIONS CALLED VILLI line the inner surface of the small intestine. Here villi on the duodenum, or upper segment, of the small intestine of a rat are magnified 630 times in a scanning electron micrograph made in the author's laboratory at Washington University. The villi give to fresh intestinal tissue the appearance of velvet. Each villus is covered by an epithelial layer a single cell thick. It was formerly thought that the work of digestion took place exclusively in the lumen, or interior space, of the intestine. Recent work in physiology, however, has demonstrated that much of digestion, as well as of absorption, is carried out by means of molecular structures that are linked to the membrane of the epithelial cells of the villus.





ANATOMY OF THE SMALL INTESTINE is similar in all mammalian species. The wall of the intestine is composed of four concentric layers. Three outer strata—the serosa, the muscularis (made up of two muscular strata) and the submucosa—surround the inner mucosa, which is the most important structure for digestion and absorption. Several sets of projections extending from the intestinal surface greatly increase its absorptive area. In the larger mammals, including man, the largest of these projections are the circular folds. Eight to 10 millimeters in height, the circular folds extend from half to two-thirds of the way around the inner intestinal surface. They are most highly developed in the jejunum, the middle segment of the intestine. The villi extend into the interior space from the folds and the surrounding surface. In the human intestine the villi range in height from about .5 millimeter to about 1.5 millimeters and are packed at a density of between 10 and 40 per square millimeter. Inside each villus is a dense network of blood and lymph vessels to which nutrients are transported when digestion is complete. Between the villi are depressions called crypts that reach down to the level of the mucosa.

lining quickly and against a concentration gradient. Simultaneously, the microanatomy of the lining has been explored. Since about 1960 these two lines of work have converged in the discovery that molecular structures linked to the inner surface of the intestine are important sites of digestion and absorption.

The structure of the small intestine is similar in most mammals. The structure reflects two necessities: the intestine must provide a large surface area for absorption, and it must fit into the small volume of the abdominal cavity. These conditions have been met by the coiling of the intestine and by the development of several sets of projections on the inner intestinal surface that greatly increase its area.

The small intestine extends from the pyloric valve of the stomach to the beginning of the large intestine; in an adult human being its length, on the average, is six meters. This length is divided into three sections. The duodenum descends about 25 centimeters from the stomach and is attached by the membranous tissue called mesentery to the rear wall of the abdominal cavity. The remainder of the intestine is freely movable. The segment following the duodenum is called the jejunum, and it occupies the upper left-hand part of the abdominal cavity; the ileum fills the lower abdomen. In all three segments the wall has the same stratified organization. Three outer layers—the serosa, the muscularis and the submucosa—surround an inner layer called the mucosa, which is the most important stratum in the digestion and absorption of nutrients.

In the larger mammals, including man, three sets of projections extend from the mucosa into the lumen. In order of diminishing size and increasing physiological importance they are the circular folds, the villi and the microvilli. The circular folds are ridges from eight to 10 millimeters high that extend from one-half to two-thirds of the way around the inner circumference of the intestine. The folds begin near the upper end of the duodenum and end at about the middle of the ileum; they are most highly developed in the last part of the duodenum and the initial segment of the jejunum. The folds may increase the intestinal surface by as much as a factor of three.

The intestine of smaller mammals lacks the large folds and has only villi and microvilli. The villi project into the lumen from the folds and the surrounding mucosal surface. These structures, with a density of between 10 and 40 per square millimeter, give the mucosa the appearance of velvet to the unaided eye. The villi vary in height from about .5 millimeter to 1.5 millimeters. Between the villi are depressions called crypts that reach down into the spongy mucosa. Inside each villus is a network of blood vessels and lymph vessels supported by loose, fibrous tissue; it is to these two systems in the core of the villus that nutrients must be transported after digestion.

Each villus is covered by an epithelium one cell thick. Two main types of cells make up the epithelium on the villus shaft: the goblet cells secrete a mucus that lines and protects the intestinal wall, and the absorptive cells carry out the main functions of digestion and absorption. The epithelial cells in the crypt form an unbroken line with those higher up on the villus. They differ from the upper cells, however, in that they are shorter and more nearly cube-shaped and certain markings on their surface are less well developed. Moreover, in many crypt cells but in no cells from the villus shaft it is possible to observe mitotic figures, which are distinctive structures that appear during cell division.

Observations made with the light microscope of the differences between the cells of the crypt and the absorptive cells higher up on the villus suggested how the intestine might secrete a digestive fluid (called succus entericus) and also absorb nutrients: the crypt cells might perform the first task and the absorptive cells the second one. Later work has shown that although the absorptive cells do take in digested nutrients, the relation of the two kinds of cells to the digestive fluid is subtler than was originally supposed.

The visible key to the relation is the mitotic figures that are seen in the crypt. The distribution of the figures indicates rapid cell proliferation, which is unusual in a population of cells in the adult body. In 1958 Charles P. Leblond and Bernard Messier of McGill University demonstrated that the cells of the crypt migrate to the tip of the villus. Leblond injected mice with thymidine that had been labeled with tritium, the radioactive isotope of hydrogen. Thymidine is one of the four nucleotides that constitute the repeating units of the DNA molecule; it is incorporated into the DNA only when the cell is about to divide. Thymidine is the only one of the four nucleotides in DNA that is not also a component of RNA; its uptake by cells is thus a clear sign of DNA replication. A few hours after the injection cells labeled with the radioactive isotope were found only in the crypt. Twenty-four hours after the injection radioactive thymidine was observed in cells halfway up the shaft; after 48 hours the radioactively labeled cells were found at the tip of the villus.

The cycle of proliferation and migration of the epithelial cells of the villus requires roughly the same amount of time in all mammals. When the migrating cells reach the tip of the villus, they are shed into the lumen, carrying with them whatever digestive enzymes they possess; hence the enzymes found in succus entericus are not secreted in the usual sense but are extruded, along with the rest of the content of the absorptive cells, into the interior of the intestine. About 17 billion cells per day are discarded along the length of the human intestine.

The elucidation of the migratory cycle among the cells of the epithelium of the villus stimulated interest in the cells. Further work showed that the absorptive cells and the goblet cells differentiate as they travel up the shaft of the villus. In this process each cell type acquires its characteristic structure.

A fully developed absorptive cell is tall and narrow. The nucleus is situated deep in the cytoplasm, below a broad apical zone richly supplied with mitochondria, endoplasmic reticulum and Golgi saccules; the abundance of these organelles indicates intense oxidative and synthetic activity. Just below its upper surface, or apex, each absorptive cell is bound to neighboring cells by an exceptionally close juxtaposition of cell membranes. Molecules cannot penetrate this tight junction, and so nutrients cannot enter the blood vessels of the villus core by intercellular channels, as was once thought. Below the tight junctions are desmosomes, which are looser connections between cell membranes. The tight junctions and desmosomes bind the epithelium into a single sheet as it traverses the villus.

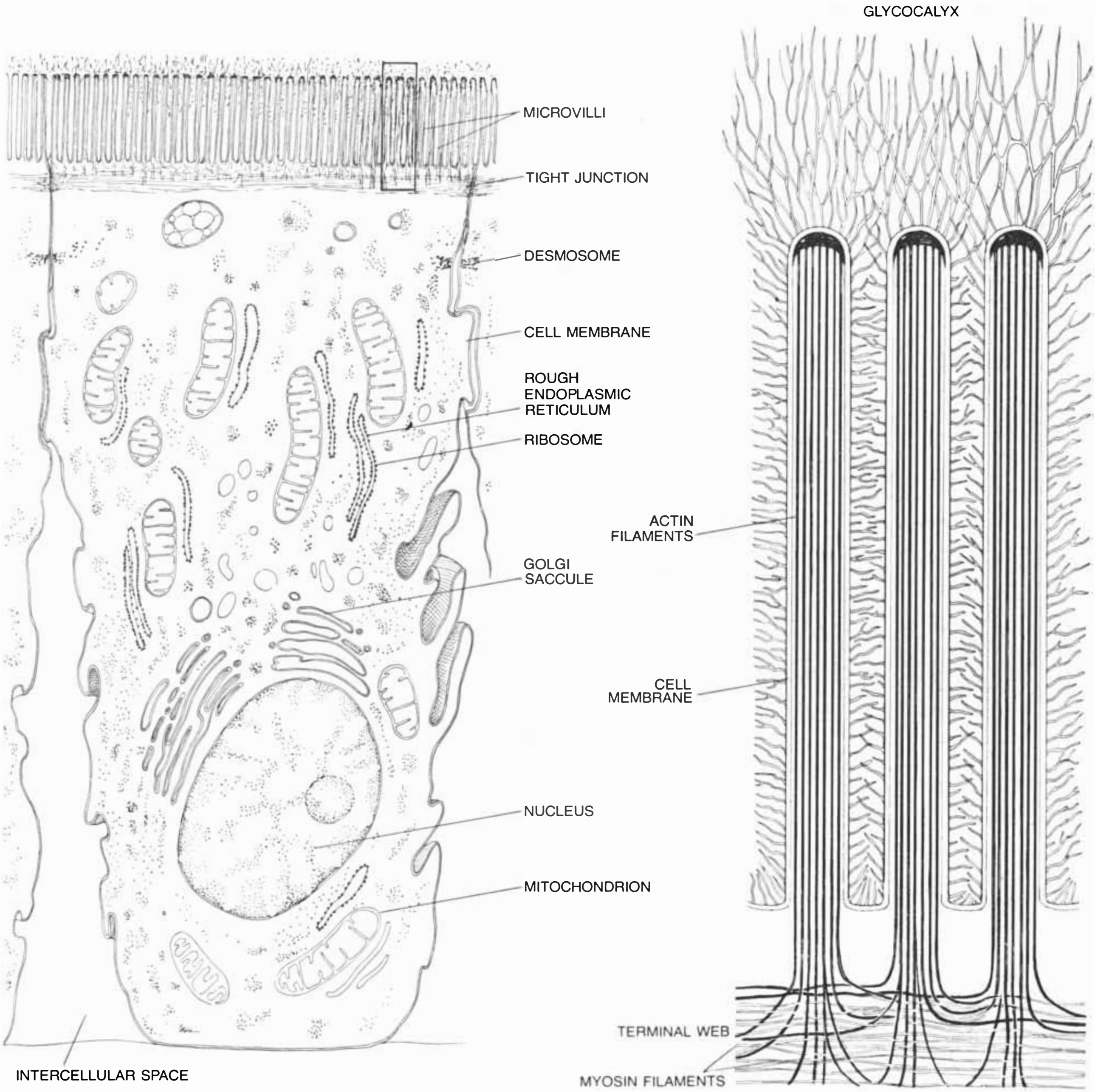
The most distinctive feature of the absorptive cells is a striated border about 1.5 micrometers wide on the top plane of the cell, which faces the lumen. The structure of the border is just beyond the resolving power of the light microscope, and it puzzled histologists until 1950, when Barbara Granger and Richard F. Baker of the University of Southern California showed by means of the electron microscope that the border is composed of a third set of projections. These smallest projections, the microvilli, are about a micrometer long; they are packed at a density of about 200,000 per square millimeter in the human jejunum. Their presence in the brush border, as the striated edge is now called, increases the surface area of the intestine by a factor of about 20. Whereas the area was previously supposed to be about 15 square meters, the intestine is now known to have a surface of some 300 square meters.

In the interior of each microvillus is a core of fibers that maintains the shape of the organelle. The fibers from all the microvilli in each absorptive cell run down to a layer of fibers called the terminal web; here the fibers are interlinked as in a piece of felt. Covering the microvillus is the cell membrane, consisting of two

WWF/Kojo Tanaka/BCL

layers of lipid molecules in apposition. Each lipid molecule comprises a hydrophilic "head" and two hydrophobic "tails" made up of extended fatty acid chains. In the membrane the hydrophilic heads face outward and the tails in the two layers face inward toward one another. The microvillus membrane has a carbohydrate coat that is thicker than the coats of most cell membranes.

It is the membrane covering the microvillus that is the actual site of admittance of nutrients into the body. A series of investigations carried out in many laboratories has shown that an array of enzymes and transport systems is incorporated into the membrane. The enzymes are capable of digesting a wide range of foodstuffs, yielding molecules that are then taken into the cytoplasm of the epithelial cell by transport systems. The molecules are sometimes modified further in the cytoplasm of the absorptive cell before being moved into the



STRUCTURE OF THE ABSORPTIVE CELL of the intestinal lining reflects its specialized function in digestion. The absorptive cell is one of the two main types of cell that constitute the epithelium of the villus. The other type is the goblet cell, which secretes mucus. Cells of both types are produced by cell division in the crypts and assume their mature structure during a two-day migration to the villus tip. Exposed in transit to the action of digestive fluids, some of the membrane covering the absorptive cell may be sloughed off. The cell is abundantly supplied with mitochondria, which provide energy for metabolic processes; with endoplasmic reticulum, where proteins (including digestive enzymes) are assembled, and with Golgi saccules, where carbohydrate side chains are attached to some proteins. Tight junctions and desmosomes bind the absorptive cells into a single sheet as they traverse the villus. The nucleus lies deep in the cell. The most striking feature of the cell is the brush border, a prominent structure on the luminal surface of the cell. The brush border is made up of rows of minute projections called microvilli. In the core of the microvillus is a bundle of fibers composed of protein. The fibers from all the microvilli on each cell extend down to the terminal web, a meshwork of fibers running parallel to the luminal surface of the cell.

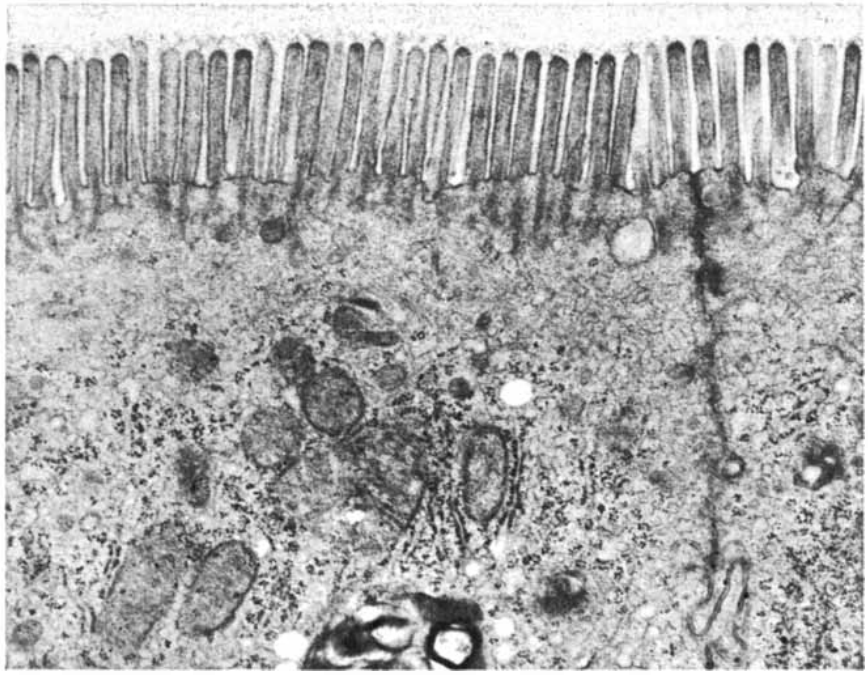
blood vessels in the core of the villus.

Susumu Ito of Harvard University was the first to establish that the carbohydrate-rich outer coat of the membrane, which is known as the glycocalyx, is composed of fine fibers that are continuous with the membrane itself. In the 1940's George Gomori of the University of Chicago and H. Takematsu of Tokyo University had developed a technique capable of showing that an enzyme can occupy a fixed position in a cell. They used their method to demonstrate that the enzyme alkaline phosphatase, which was known to be present in absorptive cells, is confined to the brush border. Alkaline phosphatase hydrolyzes, or splits, a variety of phosphate compounds, including many found in a human diet. In the 1960's it was shown by chemical analysis of tissues combined with electron microscopy that alkaline phosphatase is limited to the microvillus membrane. Alkaline phosphatase and all the other enzymes that are now known to be bound to the membrane of the brush border are glycoproteins; each molecule of such an enzyme consists of a protein with carbohydrate side chains attached to it.

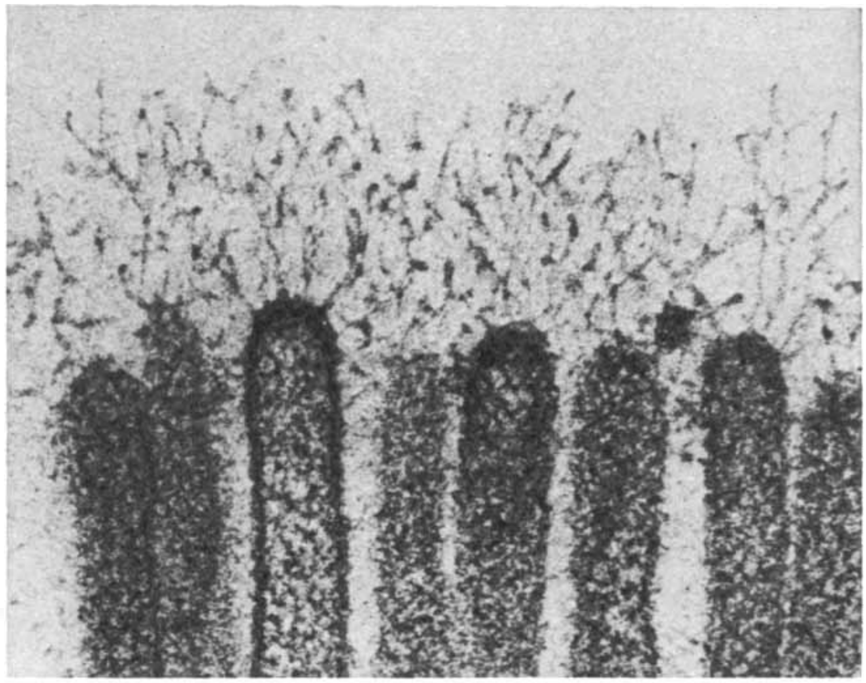
In 1961 David Miller and Robert K. Crane of Washington University in St. Louis devised a technique for isolating strips of intact brush border from homogenates of intestinal mucosa. By isolating the brush border in this way, Miller and Crane were able to analyze it chemically. They found that among its constituents are the enzymes sucrase and maltase, which are disaccharidases, or enzymes that split 12-carbon sugar molecules into six-carbon subunits. Sucrase splits a molecule of sucrose, or cane sugar, into a molecule of glucose and one of fructose. Maltase splits maltose (a derivative of starch) into two glucose molecules. Since then lactase (which breaks down the milk sugar lactose into glucose and galactose) and several other disaccharidases have been found in the brush border. Further work in other laboratories has also revealed the presence in the brush border of aminopeptidases: enzymes that cut the final amino acid from the amino end of a short peptide chain.

That these enzymes are fixed to the microvillus membrane rather than associated in some other way with the brush border was verified during the 1970's. New techniques have made it possible to separate the microvillus membrane from the rest of the brush border. In the electron microscope the membrane fraction isolated from the absorptive cells appears as a set of closed vesicles with knobby protrusions on their outer surface. From these preparations a fraction that has the activity of the digestive enzymes can be extracted.

Active disaccharidases are readily re-



BRUSH BORDER of the absorptive cell is shown magnified 18,000 times in an electron micrograph of the duodenum of a mouse. The structure of the brush border is just out of reach of the resolving power of the light microscope. It was not until 1950 that the border was shown by means of electron microscopy to consist of individual projections, namely the microvilli. The microvilli are about a micrometer long, and they are packed at a density of 200,000 per square millimeter in the human jejunum; they increase the surface area of the small intestine by a factor of about 20. The outer membrane of each microvillus is covered by a glycocalyx, or carbohydrate coat, thicker than that of most cell membranes. The glycocalyx appears here as a light fuzz extending upward from the tips of the microvilli. The dark line at the right is the boundary between two absorptive cells. The electron micrograph was made in the author's laboratory.



GLYCOCALYX extends from the tip of the microvillus into the intestinal lumen. This electron micrograph, made in the 1960's by Susumu Ito of Harvard University, shows microvilli from the ileum of a cat at a magnification of 140,500 diameters. It provided the first convincing evidence that the fibers of the glycocalyx are anchored to the microvillus membrane. It is now known that some of the fibers are parts of the digestive-enzyme molecules embedded in the membrane. The enzymes are glycoproteins, made up of a protein with several attached carbohydrate chains; the carbohydrate portion extends into the lumen. The enzymes of the microvillus membrane do much of the work of digesting sugars, proteins and phosphate compounds.


It flies through the air
with the greatest of ease.
Regal

High Fidelity for Humans:

IT EVEN FINDS Y
STATIONS WH

Finding your favorite station isn't always as easy as tuning to 123.

For example, now that digital station readouts are standard on most receivers you have to memorize the precise call numbers of all your favorite stations. Not an easy job if you have a dozen or so stations you tune in regularly.

That, however, is just one of the many unpleasantries you have to deal with if you own one of today's conventional receivers. On the other hand, it's just one of the many reasons you should own Pioneer's new SX-7 receiver.

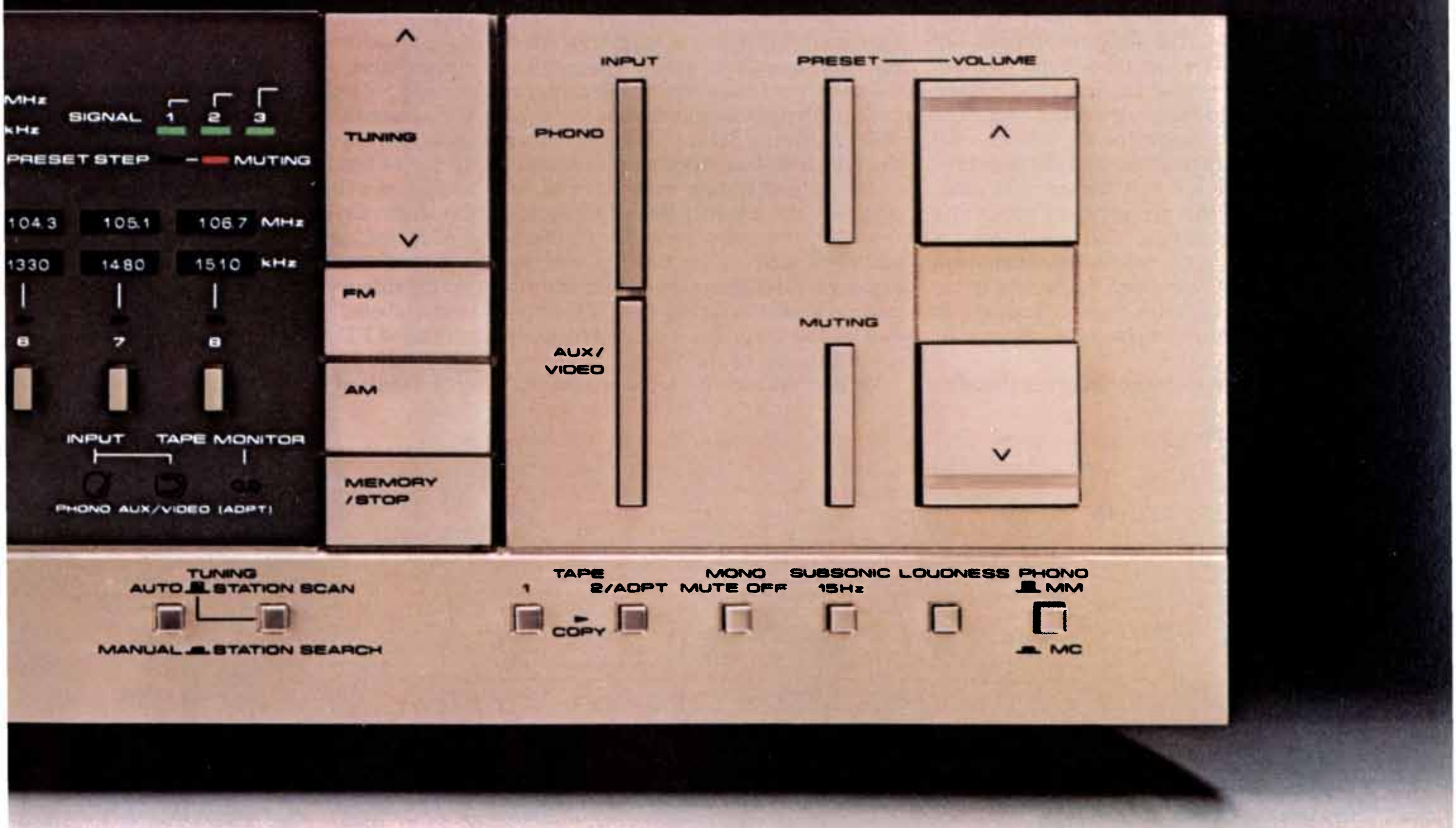
The SX-7 is a product of Pioneer's unique new concept in component design and engineering called *High Fidelity for Humans.* The result is a line of components that are as pleasant to live with as they are to listen to.

For instance, our receiver will commit to memory all your favorite stations. You can preset up to eight AM and eight FM stations. The moment you want to hear one you can recall it instantly.

Should you want to sample a variety of stations without any manual effort, simply press Station Scan. You'll hear five seconds of every strong station on the entire tuning band. If you discover a station you like you simply stop scanning.

Needless to say, not all stations have strong signal strengths. In the past you've had to struggle to tune in those stations with weak signals. The struggle's over. Due to the SX-7's ID Mosfet transistors you can

YOUR FAVORITE
EN YOU CAN'T.



tune in weak stations as quickly and clearly as you can strong stations.

Drift, of course, is another way in which distortion has been allowed to sneak in and prevail where there once was music. The only remedy has been to simply get up and readjust your station. But with the SX-7 you won't have to bother. Because our Quartz PLL Synthesized tuning is designed to make drift totally impossible.

While these technological achievements make our components easy to live with, others just plain make your music sound better.

Our patented Non-Switching™ Push-Pull circuitry is a prime example. It eliminates the distortion created by output transistors as they click on and off, thousands of times a second, in response to music signals. The SX-7's Non-Switching circuits keep our transistors from ever completely switching off, so they don't have to click back on.

If it seems as though the SX-7 has many features you just don't find on other receivers, it's because it does. Which is why we invite you to visit your nearest Pioneer dealer. He'll show you the SX-7, and an entire line of new Pioneer receivers.

They're all designed to let you spend more time enjoying music and less time simply trying to find it.

PIONEER®
We bring it back alive.

leased from the membrane vesicles by papain, a protein-splitting enzyme that ruptures the bonds holding the enzymes to the membrane. The extraction of alkaline phosphatase and the aminopeptidases from the vesicles, however, requires the application of a detergent capable of rupturing the bonds between lipid molecules. These findings suggest strongly that the disaccharidases are bound only to the outer part of the membrane; a protease such as papain can therefore cleave them away from the membrane. Alkaline phosphatase and the aminopeptidases are more deeply inserted, and extracting them requires breaking the bonds between lipid molecules in the interior of the membrane.

S. Maroux and his co-workers at the Institute of Biochemistry and Molecular Biology in Marseilles have gone on to show that the aminopeptidases not only penetrate the membrane but also protrude beyond the inner membrane surface into the microvillus core. The best current theory of the relations between the membrane and the digestive enzymes holds that the protein component of the glycoprotein penetrates to varying depths, depending on the enzyme; the carbohydrate component stretches outward into the lumen, forming the glycocalyx, which appears in electron micrographs as a network of thin fibers.

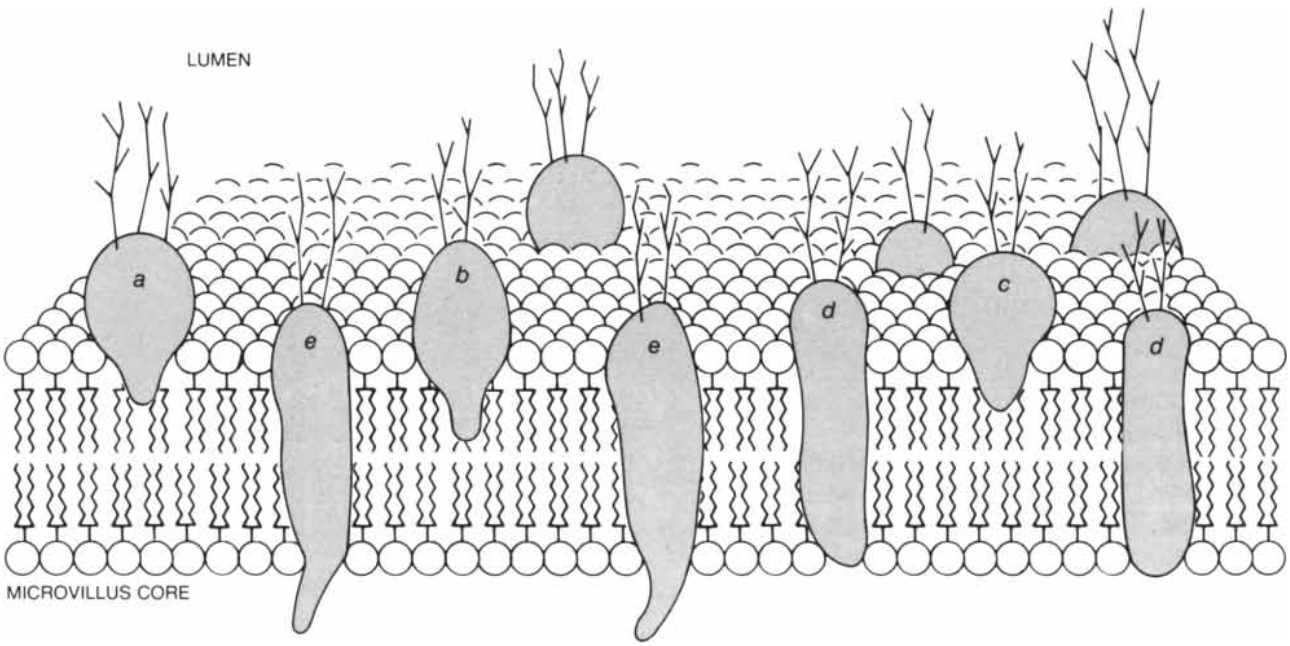
The glycoprotein enzymes accomplish much of the work of digestion directly on the surface of the microvillus. The enzymatic activity of the microvillus membrane, however, varies greatly along the length of the villus shaft. It was once thought that the membrane-bound enzymes and transport systems were synthesized while the absorptive cells were still in the crypts and then were put to use as the epithelium moved toward the tip of the villus. It is now clear that protein synthesis in the absorptive cells continues as the epithelium moves up the villus.

Milton M. Weiser, working in the laboratory of Kurt J. Isselbacher at Harvard, has developed a technique that allows the cells on the outside of the villus to be collected in serial order from the tip to the crypt. The procedure calls for filling a length of intestine with a buffered saline solution followed by a solution containing an agent called ethylenediaminetetraacetic acid (EDTA); over a period of about an hour the cells float free, beginning with those at the tip and ending with those in the crypt. By this method Weiser separated the epithelial cells into nine fractions and confirmed earlier findings that the activity of alkaline phosphatase and sucrase is absent from the cells at the bottom of the crypt but increases steadily toward the tip of the villus.

Weiser has used his technique to elucidate the processes that may underlie this gradient in enzymatic activity; he has shown that six-carbon sugar molecules are incorporated into glycoproteins (including alkaline phosphatase and the disaccharidases) more rapidly on the upper part of the villus shaft than they are near the base. It was known that labeled sugar molecules are taken up by absorptive cells and enter the Golgi saccules within five minutes of being injected into rats. The Golgi saccules constitute the system of membranes in which carbohydrates are added to a newly synthesized protein to form a glycoprotein. When this process is completed, the new glycoprotein molecule is moved to the brush border.

Weiser investigated the rate of sugar uptake by placing cells from the villus in a solution of glucosamine labeled with a radioactive isotope of carbon; glucosamine is an amino sugar that is incorporated into the carbohydrate side chains of some glycoproteins. Cells from the upper half of the villus take up the radioactive marker about twice as fast as cells from farther down on the shaft. Crypt cells take up little of the labeled sugar.

Weiser confirmed this finding by injecting rats with a solution of radioactively labeled sugar molecules; he then employed EDTA to collect fractions of villus cells in order. Again, cells higher up the shaft of the villus had incorporat-

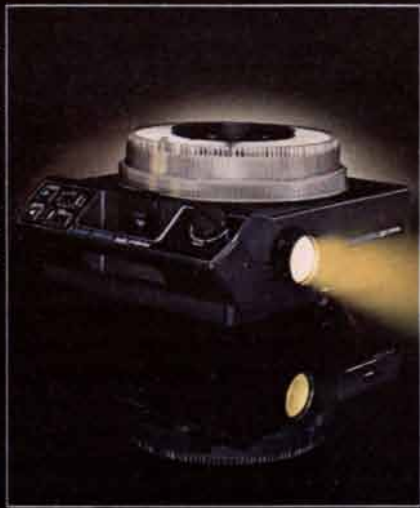


GLYCOPROTEIN DIGESTIVE ENZYMES extend to varying depths into the membrane of the microvillus. The protein component of the enzyme is inserted into the matrix of lipid molecules making up the membrane; the carbohydrate chains protrude into the lumen. The lipids of the membrane are arranged in two apposed layers, with their hydrophilic "heads" facing outward and their hydrophobic "tails," composed of fatty acid chains, facing in. The enzymes attached to the membrane are of several kinds, distinguished by the substances they digest. Each disaccharidase (*a*, *b*, *c*) splits one kind of 12-carbon sugar into its six-carbon subunits. These enzymes protrude from the membrane, which they penetrate only to a limited depth. Alkaline phosphatase (*d*) hydrolyzes, or splits, many of the phosphate compounds in food; the protein segment of the enzyme extends the depth of the membrane. Aminopeptidases (*e*) remove an amino acid from one end of a short peptide chain; they pass all the way through the membrane and into the interior of the microvillus.

SLIDE SHOW.

This Christmas get an entirely new view in slide projection. Our exclusive Slide-Scan™ built-in screen puts viewing slides right at your fingertips. Just remove the standard lens, pull out the Slide-Scan screen, and you're ready to preview, review, or just plain view your slides . . . anywhere . . . without putting up a projection screen!

Other new features include a conveniently located, illuminated control panel. Much longer lamp life (70 hours). And, on some models, a variable-speed auto-timer.



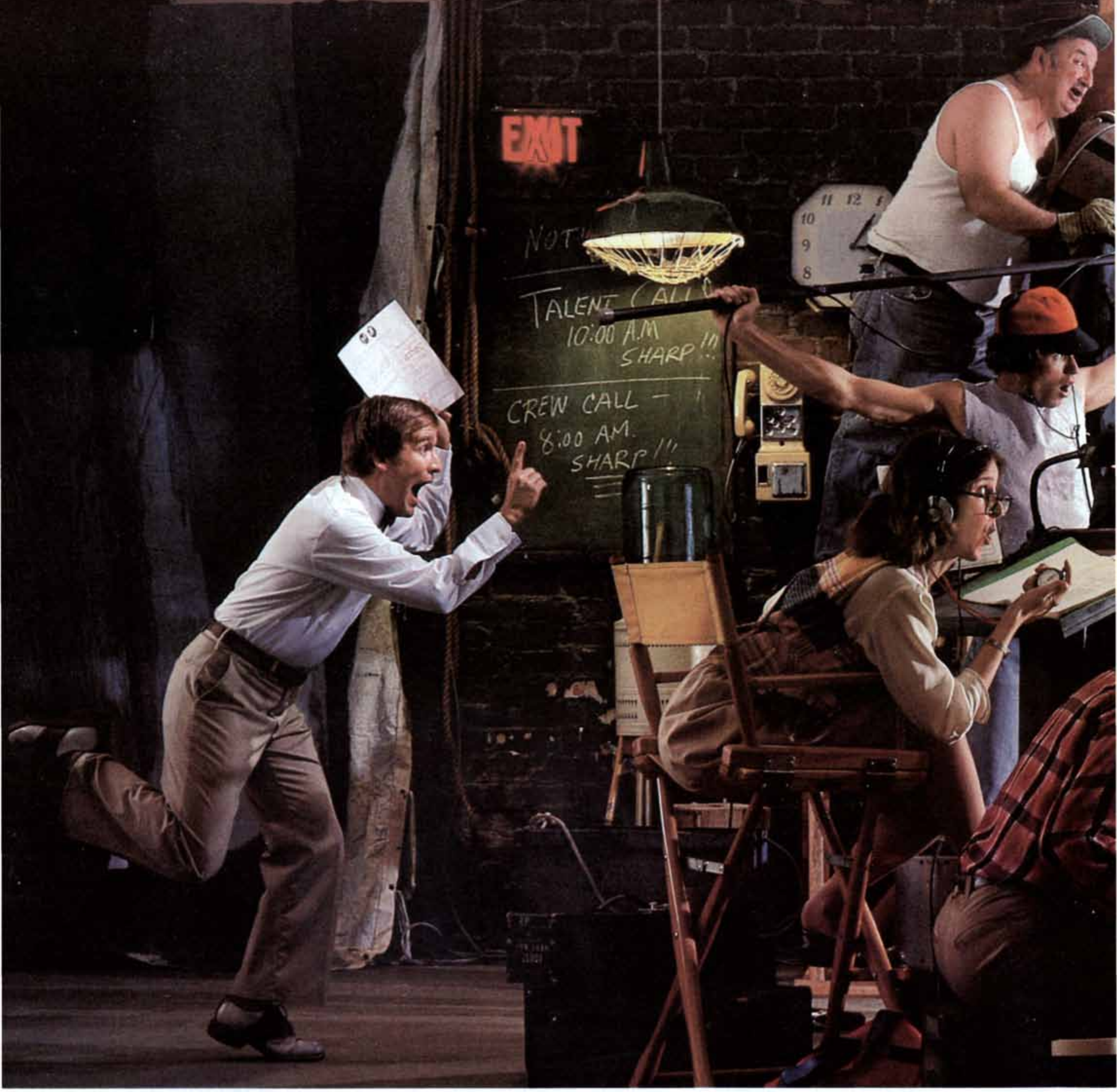
These new Carousel projectors make a terrific gift for anyone who has a 35 mm camera. Because they give you a more uniform corner-to-center illumination. A manual select control, which lets you remove the tray when the power is off. And a rapid, reliable, automatic focus system on the top four models.

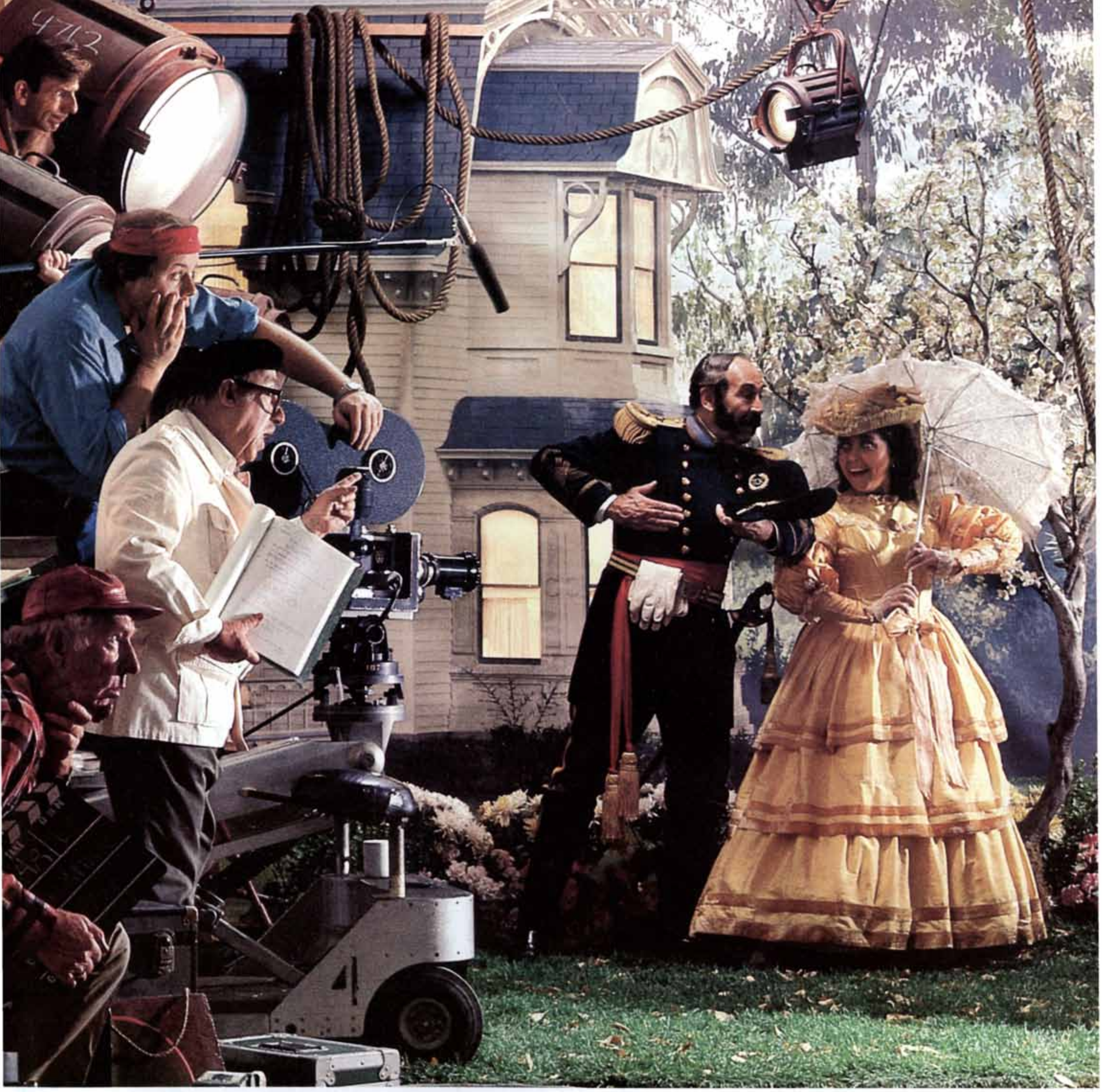
Add all this to regular features like dependable gravity feed, an Ektanar C projection lens, and a dark-screen shutter latch. Plus accessories such as a dissolve control and a sound-slide synchronizer. Then you'll know why the new Kodak Carousel projector makes such a great gift.

KODAK CAROUSEL® PROJECTORS. THE NEW BREED.

SIDE SHOW.

THIS CHRISTMAS KODAK INTRODUCES THE NEW BREED OF PROJECTORS THAT GIVES YOU BOTH.


EXIT
TALENT CALL
10:00 A.M
SHARP!!
CREW CALL –
8:00 AM.
SHARP!!



Qantas perfects Business Class.

Quick check-in at the First Class counter. Priority baggage handling.

Whenever possible, a drink on us at the Captain's Club before boarding.

A special section aboard your Qantas 747. Wide chairs in pairs so you're never more than one seat from an aisle.

Free bar service and in-flight entertainment.

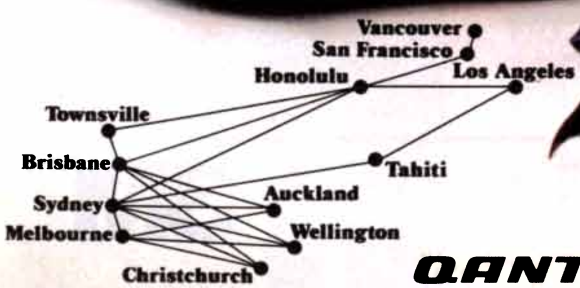
A special menu with a choice of entrées and fine wines, followed by a selection of liqueurs.

Enjoy these and other first-class touches on Qantas Business Class. Twelve flights a week to Australia. Three flights a week to Tahiti.

Ask your Travel Agent about Qantas Business Class. Only $100 or so more than the Economy Class Fare to Australia.

And that's perfect.

"Nobody's perfect, Qantas. You made the bootees too big."



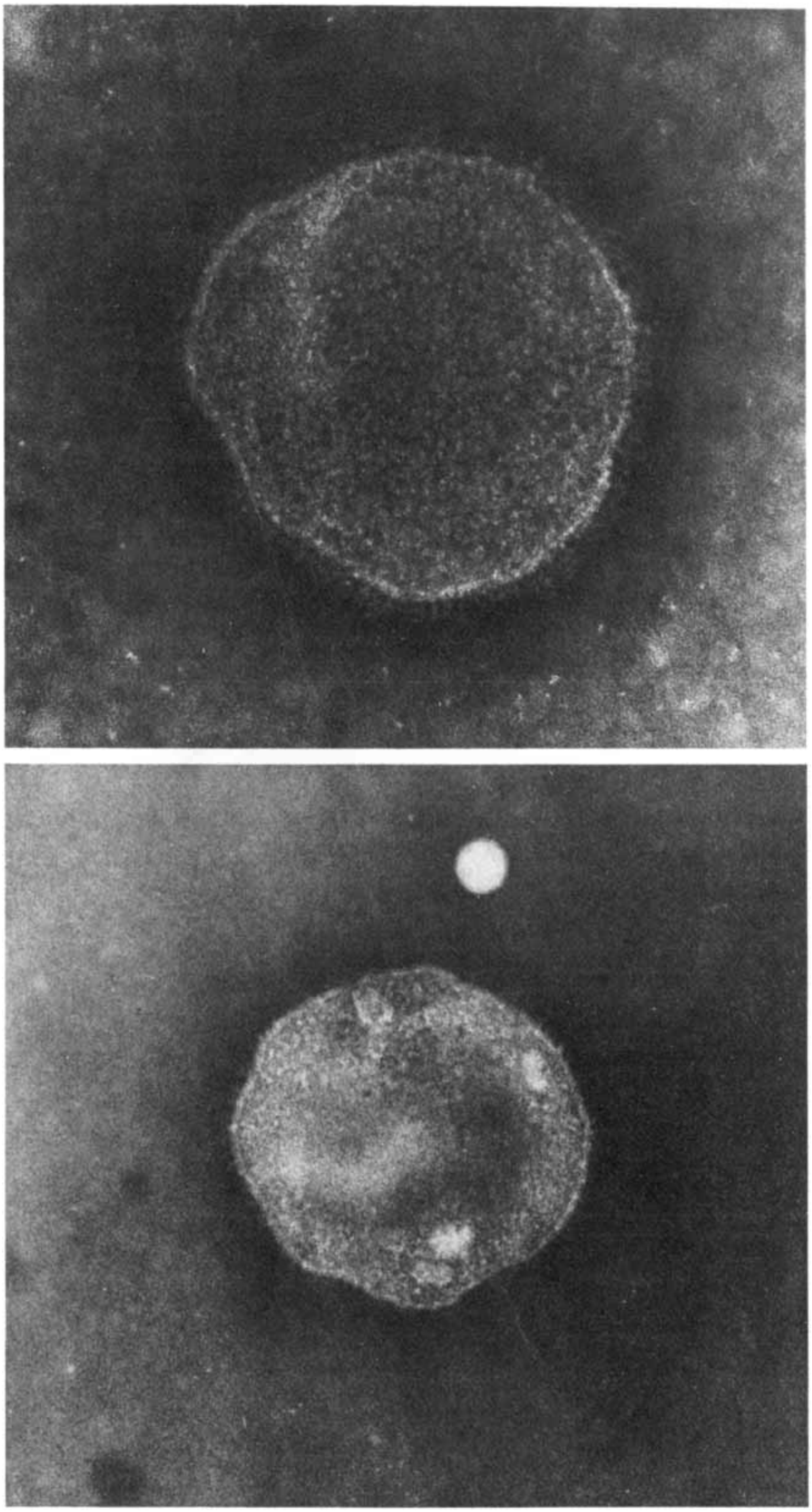
QANTAS The Australian Airline.

ed more of the marker than those near the base had.

The pattern of manufacture of RNA in the cells of the villus also supports the theory of more copious synthesis of protein higher up the shaft. A radioactive isotope of phosphorus injected into rats or chicks is taken up by the cells of the intestinal epithelium, which incorporate it into RNA molecules under construction. Crypt cells incorporate the radioactive phosphorus into both ribosomal RNA and messenger RNA. (Ribosomal RNA forms part of the structural equipment needed for the assembly of all proteins; messenger RNA specifies the sequence of amino acids in an individual protein molecule.) On the shaft of the villus, however, the radioactive phosphorus is incorporated only into the form of messenger RNA called polyadenylated RNA. (This fact was established last year by Alan Morrison and John W. Porteus of the University of Aberdeen.) A specific molecule of polyadenylated RNA is needed for the synthesis of each species of protein; the polyadenylated-RNA molecules are unstable and apparently must be manufactured continually to ensure production of the necessary glycoproteins in the required amounts. Ribosomal RNA, on the other hand, is stable and does not need to be renewed during the passage of the cells up the shaft.

Why is so much energy directed toward protein synthesis and glycoprotein assembly in the absorptive cells, which are short-lived and apparently fully differentiated? The answer lies in the fact that the brush border, exposed to the movement of chyme, or partly digested food, and to digestive enzymes in the lumen, is rapidly destroyed and must be continually repaired as it migrates upward. David H. Alpers of the Washington University Medical School has shown that the bonds holding the superficially situated disaccharidases to the membrane are subject to attack by proteases secreted by the pancreas. The life span of a disaccharidase on the membrane is only a few hours. In addition, as digestion proceeds sections of the membrane itself lift off. Entire microvilli may break into fragments. Only continuous synthesis can maintain the structure that is required for digestion and absorption.

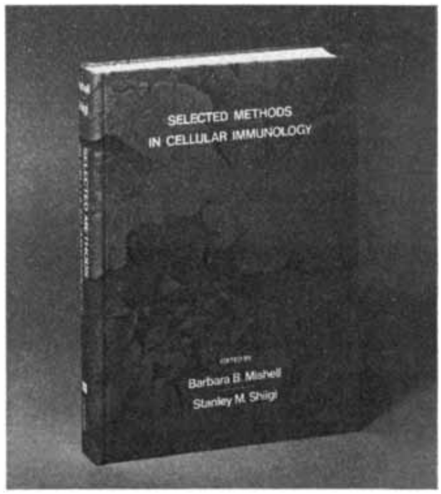
When foods have been reduced to their component molecules by the enzymes in the lumen and by the glycoprotein enzymes attached to the microvillus membrane, the nutrients must be absorbed into the bloodstream in order to be made available to the tissues of the body. Although much progress has been made in clarifying the relation of the glycoprotein enzymes to the microvillus membrane, the mechanisms by which nutrients are transported across the



MEMBRANE FRACTION of an isolated brush-border preparation appears to consist of closed vesicles with knobs made of glycocalyx on their outer surface. In the upper electron micrograph the vesicles are magnified 263,000 times. When the vesicles are treated with papain, an enzyme that splits proteins, the knobs disappear, as is shown in the lower micrograph at a magnification of 237,000 diameters. Active disaccharidases can be freed from the vesicles by means of the papain, which ruptures the bonds holding the disaccharidases to the outer membrane surface. Alkaline phosphatase and the aminopeptidases, on the other hand, can be liberated only by a substance that is capable of disrupting the lipid matrix of the membrane. Both electron micrographs were made by Dominique Maestracci of the University of Montreal.

Selected Methods in Cellular Immunology

Edited by
Barbara B. Mishell, *University of California, Berkeley*
Stanley M. Shiigi, *Oregon Regional Primate Research Center*



"*Selected Methods in Cellular Immunology* by B. Mishell and S. Shiigi is almost all that one could ask from a methods book. . . .This is a book that every laboratory doing or planning to do *in vitro* cellular immunology will want to have, and don't bother finding a spot for it on your shelf; this book will be on the bench being used." —*Cell*

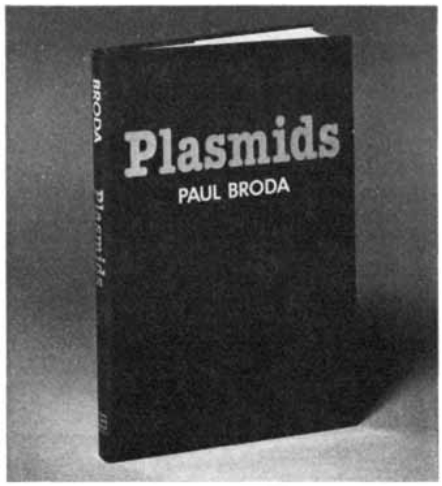
"An excellent book. . . . No laboratory with an interest in cellular immunology should be without it." —*Bioscience*

"An added bonus is that the book is written with such style that it is quite enjoyable to read for general knowledge, as well as for specific procedures."
—*American Society for Microbiology News*

1980, 486 pages, 36 illustrations

Plasmids

Paul Broda, *U.M.I.S.T., Manchester, England*



"An up-to-date, comprehensive and comprehensible guide to the subject was badly needed. . . . Here it is—Paul Broda has provided just the thing, and it is not only newcomers to the subject who will welcome his *Plasmids*." —*Nature*

"This book satisfies a real need for current information on plasmids for university lecturers, as a reference text in libraries and for use in advanced graduate student courses on microbial genetics. The author's writing style makes easy reading and his experimental familiarity with plasmids is clearly displayed." —*Search*

1979, 197 pages, 26 illustrations

Please send me the following books. I enclose payment or credit card information with order. Publisher pays postage and handling.

Copies

______ *Selected Methods in Cellular Immunology*
Hardbound: 1106 $31.95

______ *Plasmids*
Hardbound: 1111 $19.95

Guarantee: Examine these books for fourteen days. If for any reason you are not satisfied, you may return them for a full and prompt refund.

W. H. Freeman and Company
660 Market Street, San Francisco, California 94104

Name ______________________

Address ______________________

City/State ______________ Zip ______

I enclose payment with order.
(California residents add appropriate sales tax.)

Charge my ☐ Visa/BankAmericard ☐ Master Charge.

Account Number ______________________

Expiration Date ______________

Signature ______________________
All credit card orders must be signed. SA 61

W. H. Freeman and Company is the book-publishing affiliate of *SCIENTIFIC AMERICAN*.

membrane are still little understood.

Starch, the main source of carbohydrate in most human diets, is reduced to maltose by amylase in the mouth and in the lumen; amylase is an enzyme secreted by the salivary glands and the pancreas. Some of the amylase molecules may be trapped in the fibers of the glycocalyx along with other pancreatic enzymes. As mentioned above, the maltase attached to the brush-border membrane splits each molecule of maltose into two molecules of glucose.

When it was first established that the glycoprotein enzymes are associated with the membrane, it seemed logical that a molecule of maltase, after splitting the disaccharide, might serve as a carrier for the six-carbon sugars in their movement into the interior of the microvillus. Such a mechanism may account for some of the transport of glucose molecules into the absorptive epithelial cells. It is now known, however, that the maltase on the membrane is capable of splitting far more maltose than this transport system could carry; some of the liberated glucose molecules are released into the lumen.

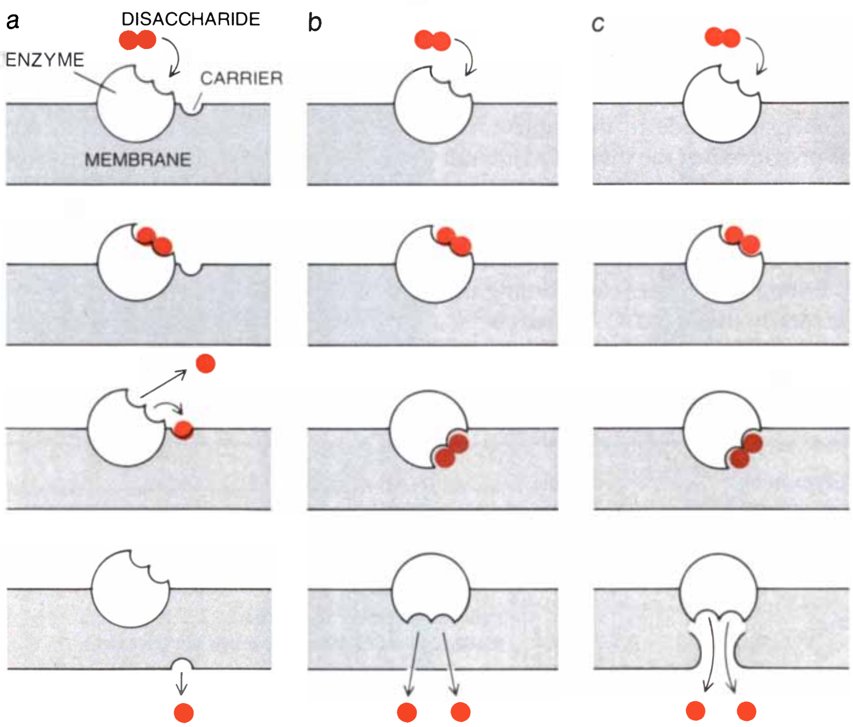
After circulating in the lumen the glucose may return to the surface of a villus. It is assumed that a second transport system then carries the glucose molecules across the membrane. The second system, it is thought, is capable of pumping glucose molecules into the absorptive cell; in this system the movement of glucose molecules is coupled to the transport of sodium ions across the membrane in the same direction, a kind of movement observed in other cells. The parallel flow of sodium ions is capable of providing the energy needed for the accumulation of glucose beyond the concentration present in the lumen. A defect in this additional glucose-transport system is presumed to be responsible for a condition called glucose-galactose malabsorption, characterized by severe diarrhea. Although the outline of the second transport mechanism has been suggested, the exact molecular pathway by which the sugar molecules enter the absorptive cell is not known.

The nature of the pathways by which amino acids and small peptides move into the absorptive cell is even more obscure. Whereas carbohydrates arrive at the microvillus membrane as simple pairs of six-carbon sugars, proteins are partially digested in the lumen to form an assortment of individual amino acids and small peptide chains. In the 1950's it was thought that proteins are completely digested by the proteases secreted by the stomach and the pancreas, leaving only individual amino acids. Actually digestion in the lumen leaves not only free amino acids but also peptide chains of various lengths. Late in the 1960's evidence began to accumulate that dipeptides (molecules made up of two amino acids) can be absorbed directly into the epithelial cells. In recent years the picture of protein absorption has been complicated further: it is now evident that tripeptides, made up of three amino acids, can also proceed directly into the absorptive cell.

The uptake of amino acids appears to require four active transport systems that are clearly independent of one another, although their molecular nature is still not known. These systems correspond to the four classes into which biochemists have resolved the 20 or so amino acids the intestine may have to deal with. The largest class, the 15 neutral amino acids, have their amino ($-NH_2$) and carboxylic ($-COOH$) groups attached to the same carbon atom. This configuration is necessary for recognition by the carrier. The neutral amino acids are transported rapidly but at rates that vary with the affinity of the amino acid for the carrier; the molecules in the neutral class compete with one another for entry into the absorptive cell. The three dibasic amino acids, which have two amino groups, also compete among themselves for a common transport site. There is little cross competition between neutral and dibasic amino acids for entry into the absorptive cell. Glycine, the simplest of the amino acids, and proline and hydroxyproline, in which the nitrogen atom of the amino group is part of a ring structure, have an affinity for a third carrier. Glutamic acid and aspartic acid, which have two carboxylic groups, also have their own system. Whether or not this fourth transport system moves amino acids into the cell against a concentration gradient is difficult to determine, because these amino acids are immediately converted into glutamine and asparagine; hence glutamic acid and aspartic acid do not accumulate in the cell. Transport of amino acids in all four systems requires energy and is coupled to the influx of sodium ions.

Evidence for the existence of the four transport systems for amino acids is provided by rare diseases brought on by a deficiency in the transport of one kind of amino acid but not of the others. In Hartnup disease, for example, the intestine is unable to absorb certain neutral amino acids but handles others without difficulty.

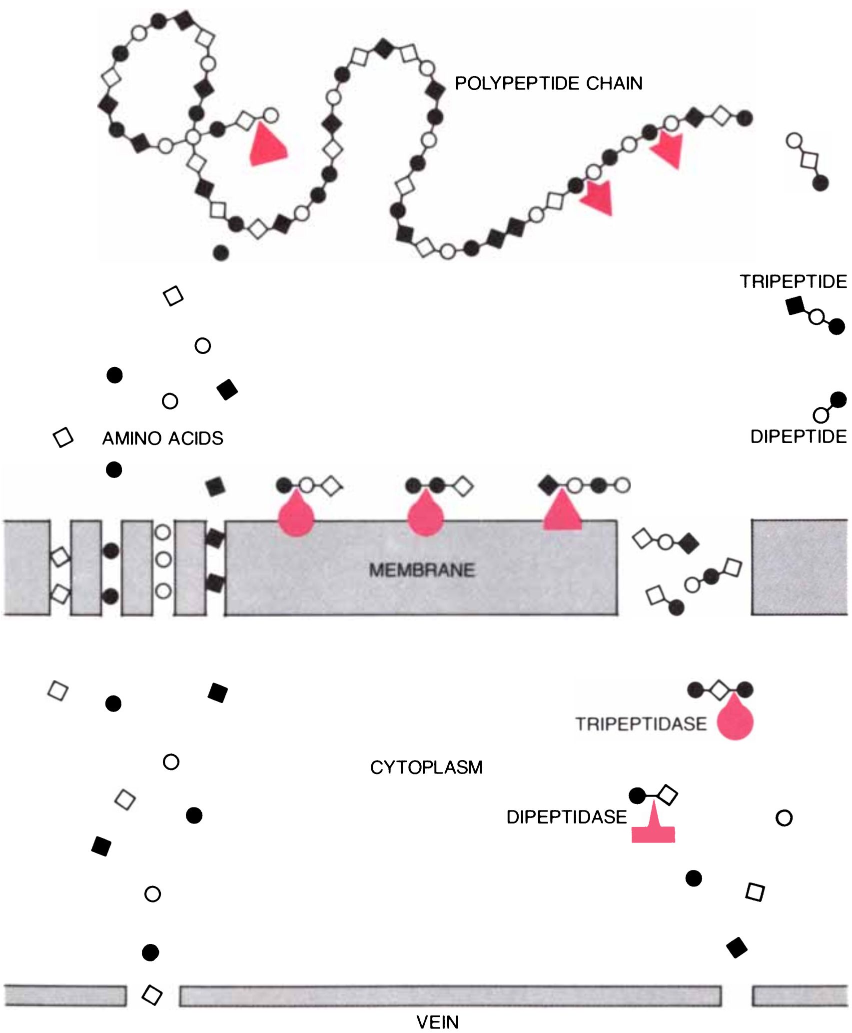
A single transport mechanism separate from those for individual amino acids provides for the movement across the microvillus membrane of dipeptides and tripeptides. When a tetrapeptide arrives at the brush border, a tetrapeptidase may cleave away its final amino



SEVERAL MECHANISMS have been proposed to explain how sugar molecules are transported across the membrane of the microvillus into the cytoplasm of the absorptive cell. The disaccharidase that splits a 12-carbon sugar, such as sucrose, into its six-carbon components is bound superficially to the membrane. If the orientation of the enzyme remains unchanged during the cleaving process, the two six-carbon sugars may be released into the space between microvilli (*a*). They could then be transported across the membrane by a mechanism independent of the glycoprotein enzyme. On the other hand, the enzyme's active site, which holds the sugar molecule, may be rotated inward while the sugar molecule is being split. In that case the six-carbon sugars might be released into the membrane itself (*b*), or they might be liberated into an aqueous channel extending through the membrane into the cytoplasm of the absorptive cell (*c*).

acid, leaving a tripeptide and a free amino acid. The tripeptide may be absorbed directly into the cytoplasm, or it may be split by a second enzyme into a dipeptide and a free amino acid. The enzymes that split tripeptides are distributed about equally between the membrane and the cytoplasm. Enzymes that cleave dipeptides into two amino acids are predominantly in the interior of the cell. Few dipeptides are split on the membrane; instead most of them are transported intact into the absorptive cell.

The transport of peptides rather than amino acids has several advantages to the organism. The competition among amino acids for places in the four carriers of individual molecules is avoided when intact peptides travel across the lipid membrane. The presence of peptide carriers minimizes the energy expended to counter concentration gradients; dipeptides and tripeptides are quickly cleaved inside the cell, so that the concentration of the peptide chains does not rise to a level higher than that in the lumen. In addition a deficiency in one of the four carriers of individual amino acids may be compensated for by the peptide transport, since the molecules that would have entered the cell by the defective transport system enter as components of peptide chains. The molecular structure of the carrier systems for peptides, as of those for individual amino acids, has not been made clear.



AMINO ACIDS AND SHORT PEPTIDES, the products of the digestion of proteins, are transported across the membrane of the microvillus by an array of transport systems. Proteolytic enzymes in the lumen break down proteins into single amino acids and peptide chains. The microvillus membrane has four systems for transporting individual amino acids into the cytoplasm of the absorptive cell. All amino acids with similar chemical structures compete for entry by means of the same transport system. The movement of amino acids into the cell is coupled with a parallel flow of sodium ions, a process that yields the energy needed to accumulate amino acids beyond the concentration present in the lumen. Aminopeptidases embedded in the membrane cleave the final amino acid from short peptide chains. When the chain has been shortened, it may circulate back into the lumen, or it may move directly across the membrane. A separate mechanism exists for the transport of dipeptides and tripeptides. Once small peptides are across the membrane, they are all reduced to amino acids by the action of peptidases in the cytoplasm. This scheme emerged in the 1960's and 1970's; although its general outlines are well established, the molecular structure of the carriers is in most cases not known.

A further gap exists in knowledge of the internal structure of the microvillus. Until the early 1970's it was widely supposed that the microvillus is a rigid structure serving only to expand the surface area of the membrane. In 1975 Mark Mooseker, who was then working at the University of Pennsylvania, demonstrated that an isolated brush-border preparation can contract under certain circumstances. The necessary conditions include exposure to calcium ions and to adenosine triphosphate (ATP), the molecule that takes part in the linked reactions that provide energy for most cellular processes.

When a microvillus contracts, the fibers in the core slide down toward the underlying terminal web. Mooseker and Lewis G. Tilney, also of the University of Pennsylvania, had shown that the brush border contains actin and myosin, the proteins that slide along each other in the contraction of muscle. In muscle tissue a third protein, meromyosin, mediates the interaction of actin and myosin. Mooseker and Tilney applied meromyosin to the microvillus to determine whether the interaction is similar to the one observed in muscle. They found that it is, and they proposed that in the contraction of the microvillus the actin filaments in the core reach down to horizontal strands of myosin in the terminal web. Meromyosin, they suggested, facilitates the contraction of the microvillus by enabling the actin to ratchet along the strands of myosin.

In the endeavor to explain how and why the microvillus contracts four additional proteins have been isolated. The most intriguing of these is calmodulin, a protein similar to troponin, which binds calcium ions and also regulates skeletal-muscle contraction. Calmodulin is known to have analogous functions in nonmuscle cells. Calmodulin may help to regulate the contraction of the microvillus, but there is another possible explanation for its presence. The calmodulin is concentrated in the microvillus shaft rather than in the terminal web, where the actin and the myosin interact. Because calcium is one of the substances that must be absorbed through the intestinal lining, Mooseker has suggested that calmodulin forms part of a transport system for calcium ions. In carrying the ions the protein may migrate between the microvillus core and the terminal web. If this hypothesis proves to be correct, it may advance understanding of the transport systems both in the membrane and in the cytoplasm of the absorptive cells. Such transport systems now represent one of the most fruitful areas of research into the lining of the intestine.

TENBA INC. believes that there is only one way to build a bag for the serious photographer. It must be easy to work out of, light to carry, protective of its equipment, and completely reliable. We take no shortcuts, and make no compromises in quality.

Why You Should Invest In Tenba Quality

- **Quick, positive, one-hand access and closure**
- **Everything at your fingertips**
- **More gear in less space**
- **Your interior-arranged as you like it**
- **Complete shock-proof padding**
- **Rugged dependable materials and construction**
- **Lightweight**
- **Complete weather protection**

TENBA is a Tibetan word meaning *strong, unshakable* and *reliable.* From its first bag, TENBA, INC. has upheld the commitment to its name.

The goal of the TENBA Pro-pak is to facilitate you, the photographer, in reaching your objectives. We consider your photographic achievements, in part, a measure of our success.

For complete information see your dealer or write or call TENBA. Dept. SA 11C

Construction

1 Weather flap over zipper **2** #10 coil nylon zipper **3** Outside raised top compartment **4** Inside top flap pocket **5** Waterproof cotton-backed vinyl liner **6** #10 coil nylon zipper **7** Welded D-ring **8-9** Box-X bar tack **10** Non-skid pad **11** Small D-rings (one in each corner, rear two are for attaching Waist Belt, and/or Back Pack harness. All four can be used to lash tripod to bottom.) **12-13** 3000 lb test nylon web **14** Wood liner **15** Closed-cell ETHAFOAM **16** 1000 lb test nylon web **17** 1000 denier CORDURA nylon **18** Bar tacks **19** Velcro pocket closure **20** 1000 lb test nylon web **21** Tourniquet buckles **22** Box-X bar tacks **23** Nylon adjustable tuck catch **24** #5 molded nylon zipper **25** Bar tacks **26** Top double stitched onto bag and ends reinforced with bar tacks **27** Velcro hook **28** Velcro loop **29** Mid divider **30** Side divider

The Quality Bag For The Working Professional

Baked Apple.

Last Thanksgiving, a designer from Lynn/Ohio Corporation took one of the company's Apple Personal Computers home for the holidays.

While he was out eating turkey, it got baked.

His cat, perhaps miffed at being left alone, knocked over a lamp which started

a fire which, among other unpleasantries, melted his TV set all over his computer. He thought his goose was cooked.

But when he took the Apple to Cincinnati Computer Store, *mirabile dictu*, it still worked.

A new case and keyboard made it as good as new.

Nearly 1,000 Apple dealers have complete service centers that can quickly fix just about anything that might go wrong, no matter how bizarre.

So if you're looking for a personal computer that solves problems instead of creating them, look to your authorized Apple dealer.

You'll find everything well-done.

The personal computer. apple®

For the authorized dealer nearest you, call (800) 538-9696. In California, call (800) 662-9238. Or write: Apple Computer Inc., 10260 Bandley Dr., Cupertino, CA 95014.

A 16th-Century Basque Whaling Station in Labrador

Less than 100 years after Columbus there was a thriving Spanish whaling industry in Canadian waters. For the first time evidence of it has been found in archives, on the shore and on the bottom

by James A. Tuck and Robert Grenier

The east coast of Canada has been a fishing ground for Europeans at least since 1497, when John Cabot completed the first known voyage to the Gulf of St. Lawrence. Cabot said he encountered cod in such teeming numbers that the fish could be caught simply by lowering a basket over the ship's side. Canada's Atlantic waters soon became a mecca for European fishermen, notably those from the Bay of Biscay.

In the 16th century Biscayan sailors, Basques in particular, came not only to fish for cod but also to hunt whales. This they did from shore stations for months at a time, trying (rendering) the blubber, storing the whale oil in casks and sailing home with their valuable cargo at the onset of winter. The Basques conducted their annual hunt at a place they called Grand Bay. Until recently, however, little evidence of this thriving industry had come to light.

In the early 1970's archival research in Spain began to turn up information on the whale fishery at Grand Bay. Subsequent archaeological work, stimulated by the 16th-century documents, has revealed the remains of one of the Basque shore stations on the southern coast of Labrador. It has also led to the discovery of a wrecked Basque ship, its contents remarkably well preserved, in the waters near the station. Here we shall tell something of the archival research (which is still in progress) that gave rise to the archaeological discoveries and report what has been found at the Labrador site over the past four seasons. The onshore archaeology has been done by workers from the Memorial University of Newfoundland; the offshore work has been conducted by the Marine Excavation Unit of Parks Canada (the Canadian equivalent of the U.S. National Park Service). Since last year what has grown into a multidisciplinary project that may require another three seasons' work to complete has been supported by Parks Canada, the Canadian Social Sciences and Humanities Research Council, the Government of Newfoundland and Labrador, the Public Archives of Canada and the Canadian Conservation Institute of the National Museums of Canada.

The archival research in Spain is being done by Selma Huxley Barkham, a student of Basque history. Her project has called for the translation and analysis of literally thousands of notarial documents, such as contracts, charters, lawsuits, promissory notes and wills, that contain references to the Grand Bay fishery. Mrs. Barkham has found these documents, many of them long forgotten, in various Basque archives, particularly those at Oñate. Among the clues she has uncovered are Basque place-names for various locations in the Grand Bay area. The documents also record numerous instances of whaling success and failure. For example, the Parks Canada reconnaissance that found the wrecked ship was undertaken on the basis of a 1565 account of the loss of a vessel, the *San Juan,* shortly before it was due to return to Spain with a cargo of whale oil.

Mrs. Barkham's studies have already provided evidence of the profitability of 16th-century whaling in Canadian waters. One document shows that in 1571 Doña Marina de Urançu insured a fifth of a cargo of New World whale oil for 2,000 ducats. This suggests that the value of the entire cargo was 10,000 ducats. That figure in turn is about the cost of two large galleons that were purchased at Seville in the same year. Therefore it was possible in a single successful season of whaling for a shipowner (or a group of joint shipowners) to recover the cost of a new vessel and to make a handsome profit as well.

The documents also provide at least a sketch of the whaling operation itself. The ships usually sailed for Grand Bay in the spring. One charter of 1566 mentions a departure as early as April 25, but later sailings were commoner. Once at the shore station the whalers pursued their prey in shallops: small boats typically some five meters long. They continued the hunt until ice conditions forced their return to Spain in December or perhaps early in the new year. Often they had to leave in a hurry because of the sudden development of pack ice or landfast ice. In some years escape from the ice was impossible and the whalers had to spend the winter at the station. The mid-1570's were particularly severe; Mrs. Barkham's research shows that the whalers wintered involuntarily at Grand Bay both in 1574–75 and in 1576–77.

The overwintering of 1576–77 caused several deaths. A few may have been due to the cold or the lack of provisions, but at least one, which did not occur until June, 1577, may be attributable to scurvy or some other deficiency disease. The victim was Juan Martinez de Larrume, as we learn from his last will and testament (the earliest such document known to have been written in Canada). In spite of the hazards of small-boat whaling most of the recorded deaths at Grand Bay were the result of overwintering rather than of hunting accidents.

Mrs. Barkham's research has yielded much other information about the techniques of shore-based whaling and about the whalers' life. The preferred prey was the krill-eating baleen-whale species *Balaena mysticetus,* known to the English as the right whale (possibly because it was the right whale to take: it was a slow swimmer and it also floated after death rather than sinking and being lost). The whales were taken as it is so vividly described in *Moby-Dick:* the whale was harpooned, and after it had weakened sufficiently it was killed with a lance. (In 1566 Captain Miguel de Cerain shipped 100 harpoons and 24 lances to Grand Bay; harpoons were more easily lost than lances.) The whale was then

INNER HARBOR of Red Bay, Labrador, in the foreground, is seen from the north. The Strait of Belle Isle lies in the background, beyond Saddle Island and the smaller Penney's Island. Basque whalers based themselves here, 3,000 miles from Spain, in the 16th century.

MIDSECTION OF WRECK, probably the vessel *San Juan* that sank in 1565, was located by marine archaeologists in 1978 off Saddle Island. The photograph shows the keelson, three frames, the floor of the starboard side, a reference grid, ballast and collapsed barrels.

FALLEN ROOF, made of red tiles, is examined by archaeologists from the Memorial University of Newfoundland. In the ruins, near the north shore of Saddle Island, were found pottery, cutlery and other artifacts made in Spain during the latter half of the 16th century.

BLEACHED WHALE BONES litter a section of beach along the western side of Red Bay harbor. The archaeological groups found the remains of at least 30 right whales (the species *Balaena mysticetus*) here and more whale bones submerged off Penney's Island.

towed to the store station for the flensing of its blubber. When the animal had to be towed against the current and the wind, the work must have been hard; one document indicates that on occasion the dead whale would be moored to await favorable conditions. In 1575 Nicholas de la Torre was accused of letting his men flense a whale that had been killed by Joan Lopez de Rey's crew and left moored to an offshore rock. The lawsuit that resulted went on for 20 years; it was ultimately concluded by the widows of the two whalers.

Of particular interest with respect to daily life at the shore stations are data on the provisions shipped to Grand Bay. Mrs. Barkham has thus far translated two lists of provisions. Foodstuffs included bacon, salt cod (even though cod could be caught in New World waters), sardines, wheat, broad beans, peas, olive oil, mustard seed, garlic, sherry, cider and large quantities of ship biscuit. The whaling equipment (in addition to the shallops, harpoons and lances) included several sizes of hemp line and knives for flensing and mincing the blubber. Medicines for the ship's barber, in those days more of a surgeon than a haircutter, are listed but are not specified.

Imported building materials included nails, "earth" (presumably clay) and roof tiles. Lumber is not mentioned, and so it seems safe to assume that the shore-station buildings were made of local timber, cut and hewn with the axes brought from Spain. (Captain de Cerain also shipped 16 axes to Grand Bay.) How these buildings were constructed is largely ignored in the documents. Replacement roof tiles were in demand; in 1563 Francisco de Florriaga shipped a "sufficient quantity of tiles for the repair of the cabins" exposed to the elements in the preceding winter. Only a single oblique reference suggests how such a "cabin" was used. Giving testimony in a lawsuit, a harpooner, Simon de Azcoita, states that the place where he held a conversation with a companion was "beside the cabin that they had made for the boiling down of the whales they kill."

Where was Grand Bay? Mrs. Barkham's study of place-names has led her to the conclusion that the Basque whalers had given that name to the Strait of Belle Isle, the narrow passage between northernmost Newfoundland and the southern Quebec-Labrador coast that leads from the Atlantic to the Gulf of St. Lawrence. One place mentioned in the Basque documents was Samadet or Semadet, a name that seems to equate with East St. Modeste, a former settlement a few miles west of Red Bay on the Labrador side of the strait. Another Basque place-name was Xateau or Chateo, which seems to equate with today's Chateau, also a settlement on the Labrador side. The number of references to a third place, variously called Butus, Buytes, Boytus and Buitres, suggests that it was one of the most important Basque shore stations.

Red Bay is one of the better natural harbors on the Labrador side of the strait. It includes an inner basin surrounded by steep hills 300 to 500 feet high. It is sheltered from the outer bay by a low island with a hill at each end that British Admiralty surveyors aptly named Saddle Island. A visit to Red Bay seemed indicated.

Accordingly in 1977 Mrs. Barkham and one of us (Tuck), accompanied by Walter Kenyon, a representative of the Royal Ontario Museum, and others, visited the little town of Red Bay. We found the beaches and garden plots of the town littered with thousands of fragments of red tile. The same was true of the shores of Saddle Island, where there were a number of stone walls bearing traces of a solidified black substance. Samples of the substance were later identified as the residue of burnt animal fat. We concluded that the tile fragments were the remains of Basque roof tiles. The clinkerlike black material on the stone walls suggested that this was the place where whale blubber had been tried into oil.

Our findings were promising enough to attract funding for a month of exploratory excavations the following year. The work of that first season was concentrated at three areas on Saddle Island that appeared to have lain virtually undisturbed since their abandonment some 400 years earlier. On a small terrace overlooking the harbor we uncovered a fallen red-tile roof. Associated with this evidence of a former structure were pottery, glassware, objects of iron and other artifacts that placed the date of the structure at some time in the second half of the 16th century.

Our second exploratory excavation was in a waterlogged area, the kind of place whose wetness might preserve organic materials. Our test trench exposed bushels of wood chips, the by-product of shaping locally felled treetrunks. Also uncovered were parts of barrel hoops and barrel bindings, fragments of leather and a few bits of oak, a tree that is not native to Labrador.

Our third site was where the animal-fat residues had been found, a place we already thought of as the tryworks, or oven area. A small test trench revealed a layer of well-preserved roof tiles buried under wave-tossed sand. The substantial stone wall that had attracted our attention on our first visit stood near the middle of what were obviously the remains of a large structure.

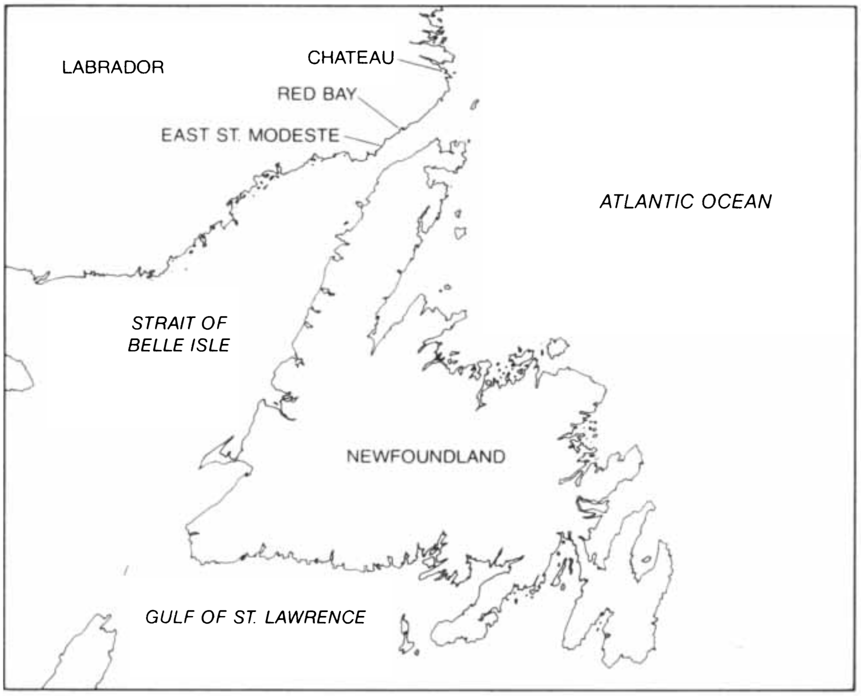
The Memorial University group had left Red Bay, more than satisfied with the season's results, before the Parks Canada divers made 1978's most spectacular single discovery. Only 30 meters off the north shore of Saddle Island they located the remains of a sunken ship loaded with a cargo of wood casks. The hoops and bindings of the casks exactly matched those we had excavated from the waterlogged site. Furthermore, the vessel fitted the description of a wreck recorded in the Basque archives. It was Mrs. Barkham's knowledge of the wrecked vessel, the *San Juan*, that had led to the Parks Canada underwater reconnaissance.

Over the next season, in addition to opening up new areas of Saddle Island, the onshore archaeologists mapped, photographed and removed the shattered tiles of the fallen roof uncovered in 1978 at what was designated Area *A*. The zone outlined by the tiles was 14 meters long and eight meters wide, so that the structure sheltered by the roof may have been only slightly smaller. Evidently it had been built at least in part of fieldstone. Other stone walls on the island characteristically have a core composed of rubble and small stones and are faced with larger stones, carefully fitted together. Here, however, only a line of rubble suggests where the front, or north, wall of the structure had stood. There was no need to build a back wall because the structure was sited so that its south side was formed by a nearly vertical outcrop of bedrock. (Incorporating such natural features in building construction is a common Basque practice.)

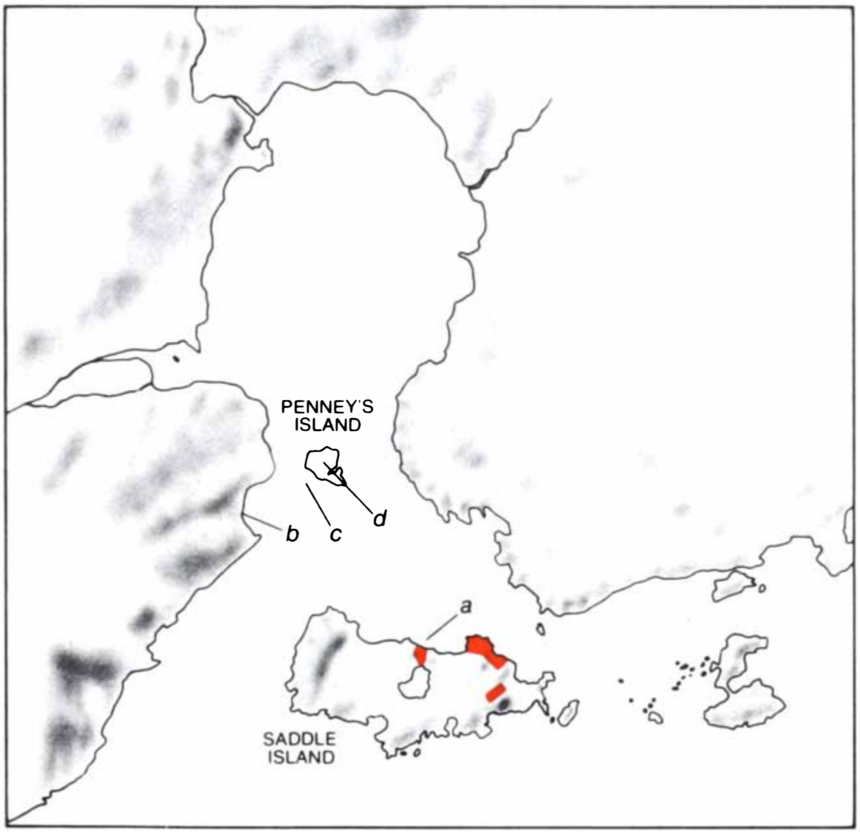
The shorter east and west walls, if they ever existed, have left no trace. Perhaps the structure was open at each end, or perhaps these walls were wood and soon rotted away in the island's acid soil. It is clear that the roof frame was wood; hundreds of hand-forged nails attest to it. One cannot, however, determine whether it was a simple shed roof or a gabled one. Both styles are common in Basque construction.

Whatever the style of the roof, the structure sheltered two small hearths that were excavated during the 1980 season. They proved to contain charcoal, burnt animal bones, fragments of pottery, lead shot and bits of waste lead from casting. It was evident that besides providing heat the hearth fires served for cooking and for other homely tasks such as casting shot.

In and around the structure (particularly at the west end, which seems to have been something of a refuse dump) many artifacts were uncovered. The most numerous were bits of broken pottery; they totaled more than 2,000. Many could be identified as parts of storage jars, cooking vessels, pitchers, porringers and plates. Shards representative of some of the smaller pitchers, porringers and jars bore blue, green or gold decorations applied over a white



STRAIT OF BELLE ISLE, the entrance to the Gulf of St. Lawrence separating Newfoundland from the Quebec-Labrador coast, was known to the Basque whalers of the 16th century as Grand Bay. East St. Modeste may have been a shore station the Basques called Semadet and Chateau a station they called Xateau. The Red Bay station they may have called Butus.



RED BAY AREA is shown in greater detail. Most of the excavating has been done in the parts of Saddle Island marked in color; the sunken vessel lies at *a*. The bone deposit shown on page 182 is at *b*, northwest of Saddle Island. Other whale bones lie off Penney's Island, as does the wreck of a smaller vessel (*c*). Ovens for trying (rendering) the blubber were found ashore at *d*.

glaze. Such majolica ware (named after the island of Majorca, a major center of production) was commonly used in (and exported from) Spain in the 16th century. Glassware was far less abundant; only the fragments of a stemmed wineglass and of a large decanterlike bottle decorated with a raised design of wishbones were unearthed.

In addition to the nails and spikes for the timber framework of the structure, the excavators found a number of other metal artifacts. Among them were several iron knives, one with an elaborately carved wood handle and a bronze pommel shaped like a crown. Taken together with the hearths and the domestic character of the pottery, the knives suggest that the structure had been a dwelling. The impression is reinforced by a scattering of other finds: a wood rosary, a coin and a key. As the 1979 season closed it seemed plausible to believe that individuals of relatively high status had lived here.

Work the following season at an adjacent Saddle Island locale, Area *E*, altered this opinion. A structure uncovered in the area was in a location almost exactly like that of Area *A*, and most of the artifacts unearthed—including pottery, glassware and other domestic objects—were also much the same. Among them, however, were several tools indicating that the occupants of the dwelling were coopers. The tools include two adzes, one so well preserved that its cutting edge is still sharp, a hoop driver and several head vises (for setting the upper head piece of a barrel in place as the staves were assembled). These finds threw new light on two bits of metal found earlier in Area *A* that were not identifiable until the Canadian Conservation Institute made X-ray plates of them. One bit proved to be the tongue of a plane and the other a fragment of a saw blade. It therefore seems probable that the residents of both areas had been engaged in woodworking.

All doubt about this was soon dispelled as far as Area *E* is concerned. Immediately to the north of the structure ground water had accumulated in natural depressions in the bedrock. When these waterlogged hollows were excavated, they were found to contain the refuse of barrelmaking so well preserved that it might have been thrown there only years earlier rather than centuries. Here were barrel parts that had been discarded because of some imperfection in material or flaw in manufacture. Oak and beech staves and head pieces matched those recovered from the wrecked ship, as did hoops made of split alder and withy barrel bindings. There were bits of wood left over from the shaping of the round head pieces, trimmed hoop ends, willow fragments too knotty or too small to use and even

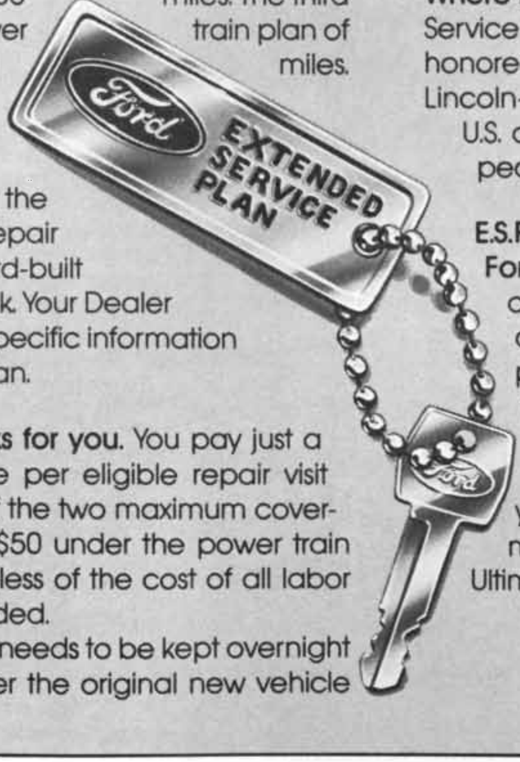
Picking options for your new car? Don't forget Ford's Ultimate Option. The only one that could pay for itself.

FORD'S ULTIMATE OPTION–E.S.P.

EXTENDED SERVICE PLAN

Ford Motor Company's Extended Service Plan is designed to offer you long-term protection against rising service costs.

E.S.P. offers long-term car repair protection. There are three plans to choose from. You can purchase a maximum coverage plan of 3 years/36,000 miles or 5 years/50,000 miles. The third plan is the power train plan of 5 years/50,000 miles.



What E.S.P. protects. E.S.P. guards against the cost of major repair bills on your Ford-built car or light truck. Your Dealer can give you specific information about each plan.

How E.S.P. works for you. You pay just a $25 deductible per eligible repair visit under either of the two maximum coverage plans, or $50 under the power train plan ... regardless of the cost of all labor and parts needed.
And if your car needs to be kept overnight for repair under the original new vehicle warranty or the repair of E.S.P.-covered components thereafter, you're provided a rental car reimbursement allowance of up to $20 per day (excluding mileage) for up to five days.

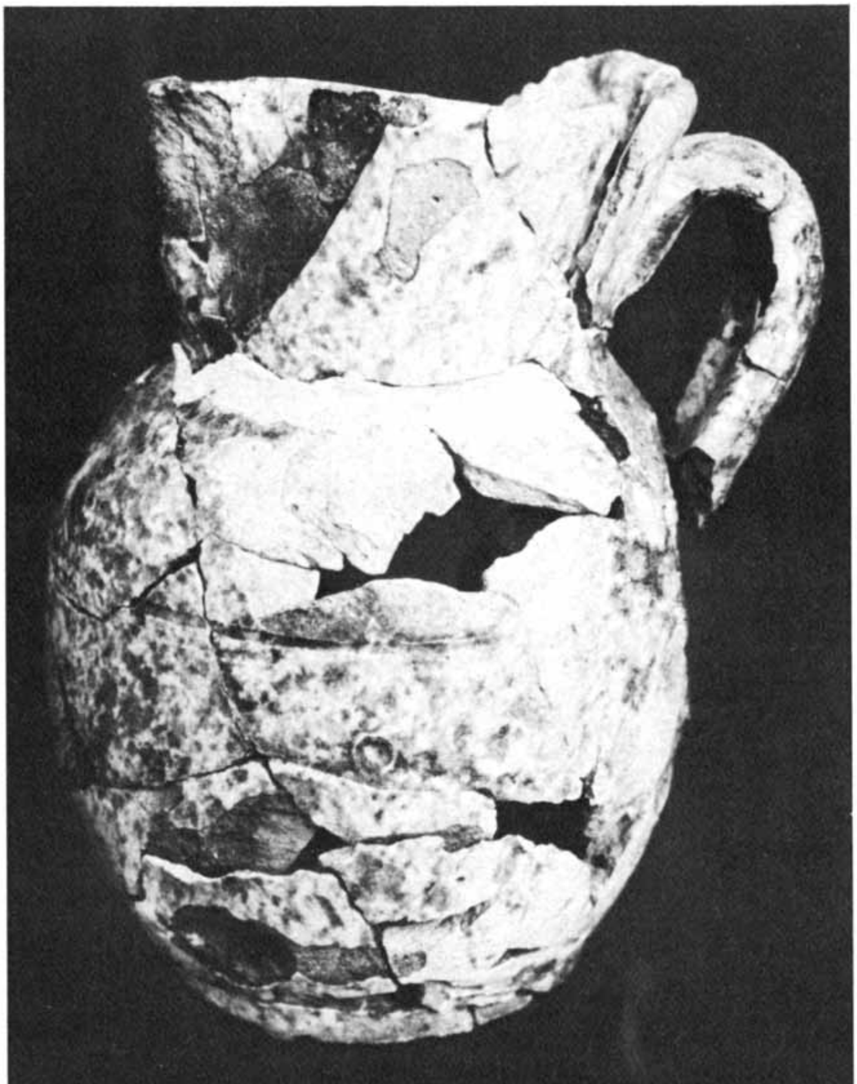
How E.S.P. helps at resale time. For a small fee, the remaining coverage is transferable to the new owner. This can be a strong selling point.

Where E.S.P. is in effect. The Ford Extended Service Plan is the one and only plan that's honored by more than 6400 Ford and Lincoln-Mercury Dealers throughout the U.S. and Canada. That can mean real peace of mind wherever you are.

E.S.P. offers low-cost protection for your Ford or Lincoln-Mercury vehicle. The cost of E.S.P. is surprisingly low. And it can be added to your monthly payment if you finance your new vehicle. When you consider the low cost of protection on your initial investment, and how long you'll probably keep your car, it just makes good sense to buy Ford's Ultimate Option ... E.S.P.

Ask for Ford Motor Company's Extended Service Plan by name.

BARRELHEADS AND STAVES raised from the hold of the ship sunk off Saddle Island were so well preserved during their immersion in cold Labrador waters that they belie their 400-year age. The marks of the split-alder hoops that bound the staves together are still visible.



SMALL PITCHER, one of a number of pottery objects made of distinctively Spanish majolica ware, was unearthed at the site of a shore-station cabin where barrelmakers worked.

shavings and sawdust. Unquestionably the Area *E* dwelling was a combined residence and cooper's workshop, and the general similarity between it and the dwelling in Area *A* suggests the same role for both structures. Still, our 1979 assessment may not have been entirely mistaken; perhaps coopers, whose barrels were certainly essential to the entire enterprise of shipping whale oil, enjoyed a high status at the shore stations.

Between the coopers' quarters and the shore were the ovens for trying blubber. One of them, with a stone wall that still stands nearly a meter high, was excavated during the 1980 season. The wall, which ran for 10 meters parallel to the shore, formed the backbone of the oven. On its shore side had stood the trying caldrons and on its land side was the work area. The wall itself was carefully built of local stone and ship ballast, secured by a mortar made of clay (presumably the "earth" listed among the imported construction materials). When the clay became impregnated with whale oil, it formed an almost cement-like bond between the stones.

What was the trying process? We possess no descriptions of the Basque techniques, but these same whalers taught the art of whaling to the English and Dutch mariners who established the whale fishery on the Arctic Ocean island of Spitsbergen. Accounts of the procedure at the Spitsbergen stations survive. The whales were towed or winched close to the shore for flensing. (In the shallow water in front of the ovens on Saddle Island there are piles of rock ballast and carefully mortised timber slipways. These may have been "cutting in" stages where the whales were flensed.) The strips of blubber were minced, loaded in barrels and hauled up a timber ramp to caldrons set on the seaward side of an oven wall. At Spitsbergen circles of brick supported a row of copper caldrons, each with a capacity of 48 to 54 gallons (roughly one barrel of minced blubber). The bricks not only supported the caldrons but also formed fireboxes for them. The tops of the caldrons stood some two meters above the ground. (On the seaward side of the wall at Saddle Island are the remains of four fireboxes constructed of rock and clay.)

One observer of the Spitsbergen tryworks, Captain Thomas Best, reported that it was customary to place an unseaworthy but still watertight shallop on the landward side of the oven wall and partly fill it with water. The oil from the caldrons was then ladled into the water by men standing on a scaffold built along the wall. The water served to cool the hot oil and partially purify it; any dross would settle to the bottom of the boat and the whale oil floated.

On the landward side of the wall at Saddle Island were some remains of a similar ladlers' scaffold, preserved by saturation. The work platform had been built mainly of local timbers, but several oak planks with peg holes in them showed that the Basque builders had also used old boat or ship timbers. Instead of an old boat hull to hold the water for cooling the oil the ladlers had set up 24-gallon tubs, made out of half barrels, on the scaffold. The staves, head pieces and hoops found on top of the spruce and oak timbers of the scaffold were again a perfect match with the barrel parts aboard the wreck. We also uncovered roof tiles and parts of a wood framework, evidence that the ladlers had worked under shelter and that the caldrons too were protected from the elements.

The document that led to the search for a sunken ship off Saddle Island records the proceedings of a lawsuit. It includes the statement that late in 1565 the vessel *San Juan,* fully loaded with a cargo of whale oil, was preparing to sail for Spain when it was lost "for want of cable." The phrase suggests that a sudden storm had snapped the heavily laden vessel's moorings and driven it aground, where eventually it broke up. In 1978 the divers of the Marine Excavation Unit had spotted ship timbers protruding through the layer of bottom silt at a depth of 10 meters close to the north shore of Saddle Island. The first thing they saw as they surfaced after inspecting the timbers was the test trench the Memorial University excavators had cut through one of the ovens ashore.

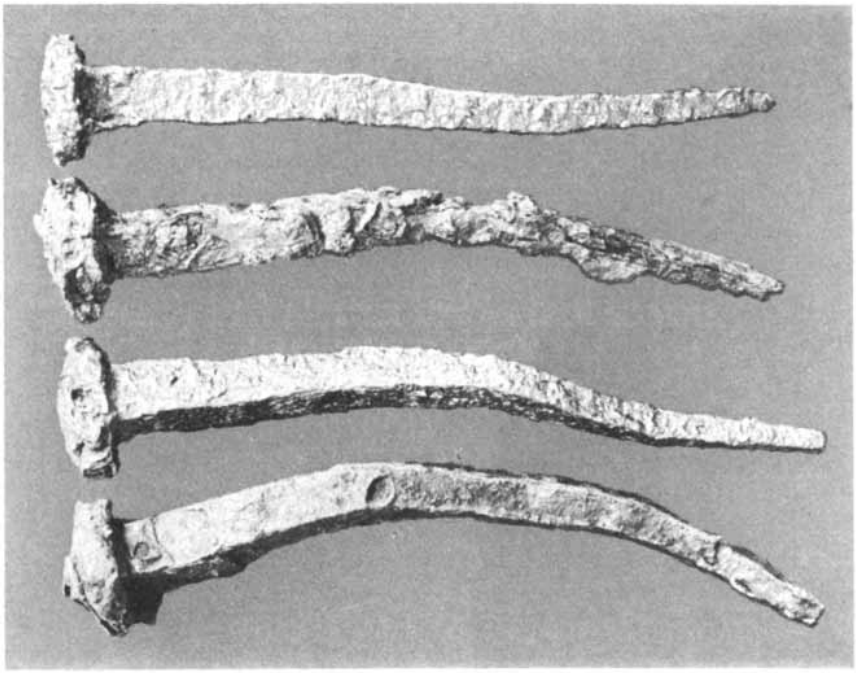
The fact that the wreck lies so near a Basque shore station raised the hope that it is indeed one of the Basque ships and perhaps even the *San Juan*. Preliminary work by the Marine Excavation Unit has strengthened this hope. Most of the ship's timbers are oak, a fact that rules out its being a later, locally built craft. To this must be added the fact that the wreck contains quantities of oak and beech barrel staves, and head pieces and barrel hoops matching those found in the waterlogged area ashore. Whether or not this was the *San Juan,* it was clearly some vessel that was associated with the 16th-century Basque whale fishery at Red Bay.

In the seasons since 1978 the Parks Canada archaeologists have done major work at Red Bay in cooperation with the Government of Newfoundland and Labrador. Their efforts have been concentrated on two main tasks: excavation of the wreck itself (including the digging of a "shore trench" running from the stern of the wreck in the direction of Saddle Island) and a limited survey of the harbor area in general. The harbor survey, a joint effort of both archaeological groups, has found one beach area to the west that is littered with skulls, vertebrae, ribs and other bones of right whales. It has also uncovered the remains of at least two more ovens on the northwest side of Penney's Island.

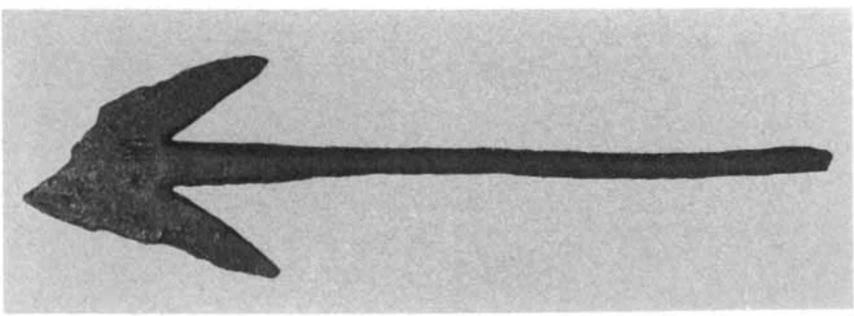
Apart from the whale bones the only evidence of Basque activity at the beach site, both on the surface and in one test trench, was a few fragments of roofing tile. Off the shore of Penney's Island, however, the divers located two more concentrations of whale bones. It appears that after the whales had been stripped of their blubber for trying ashore their carcasses were simply allowed to sink to the bottom.

Off the western end of Penney's Island the divers came on the remains of a second sunken vessel, considerably smaller than the one off Saddle Island. Its construction and fastenings suggest that the vessel is an old one, but only further study will show whether or not it is contemporaneous with the Basque fishery at Red Bay.

In addition to the shore trench three other exploratory trenches were excavated underwater to determine the limits of the wreck. One extended north of the ship's bow and the other two were directed east and west, that is, to port and to starboard. The wreck lies on its starboard side and so structural remains and artifacts were found widely dispersed in that direction. The floor of the harbor consists of silt, rock and fragments of shell. In digging the shore trench the underwater group uncovered a thick layer of wood chips identical with the ones found by the thousands ashore: waste from the trimming of local timber. Since the chips had sunk to the bottom rather than floating away, one may speculate that they represent a secondary deposit of refuse that had



IRON SPIKES, 17 to 18 centimeters long, were among the hundreds of hand-forged nails and spikes used by the Basque whalers in building their cabins and oven sheds on Saddle Island.



HARPOON POINT, originally 40 centimeters long, is another item of whalers' equipment found at Red Bay. The letter *M,* probably an owner's mark, resembles a 16th century Basque monogram. Harpoons, lances and flensing knives were among the supplies shipped to Butus.

already become soaked with water at a construction site.

Above the layer of chips the divers found a second layer of organic material that in places was as much as 20 centimeters thick. It consisted of the bones of birds, land mammals, whales and, most abundant of all, cod. The fishbones are mainly skulls and fragments of vertebrae; the absence of the other bones suggests that the Basques split the fish for salting and drying. This find may explain something that had puzzled the Memorial University group on its first visit to Saddle Island: a series of circular platforms consisting of stone and gravel and measuring about two meters in diameter. They could well have been places where the salted cod was stacked after drying. In the shore-trench deposit the whale bones lay among the cod bones, indicating that the Basques whaled and fished simultaneously. Artifacts from the shore trench include the ubiquitous fragments of roof tile, shards of pottery, bits of rope, a number of remarkably well-preserved leather shoes and the partial remains of one or two of the whalers' small boats.

The major objective of the underwater investigation is the wreck itself. This is not only because of the quantity of the remains and their excellent quality of preservation but also because the ship's timbers can yield unique information about the construction of 16th-century merchant vessels. According to the documents studied by Mrs. Barkham, the *San Juan* sank in waters shallow enough for the crew to salvage part of the cargo, some of the victuals for the voyage home and some of their personal belongings. Thus if the Saddle Island wreck is indeed the *San Juan*, it can be assumed that some parts of the sunken vessel's upper structure remained above water, perhaps for several seasons. In any event the action of the winter ice would have accelerated the collapse of the hull. Deposits of silt and refuse, including pieces of the vessel itself, have built up on both sides of the wreck. This protective blanket, in combination with the low temperatures of Labrador waters, has preserved both the wreck and the part of its cargo that remained aboard.

The underwater group has now excavated a series of two-by-two-meter squares, chiefly in the midsection of the wreck. These excavations have uncovered the entire width of the vessel below

FIELDSTONE-AND-RUBBLE WALL was built parallel to the shore at Saddle Island. The kettles for trying the blubber stood on the shoreward side of the wall. The trench on the landward side held fragments of the platform where men stood to ladle oil out of the try kettles.

NOW WE KNOW WHAT IT'S LIKE TO BE ALONE.

It wasn't always this way.
It used to be hard to decide among all the 35mm SLR's.
Then we created the new Minolta XG-M. A camera so extraordinary it stands alone in its class. With an unrivaled combination of creative features.

First, it's automatic. So it's easy to get sharp, clear pictures. You just point, focus, and shoot. It even has electronic features that keep you from making mistakes.

As your skills advance, you'll appreciate advanced features like metered manual and exposure override for full creative control.

To further separate ourselves from the competition, we built in the option of professional motor-drive. Something normally found only on more expensive cameras.

It lets you shoot a blazing 3.5 frames per second. So you can catch a baseball as it comes off the bat. Or halt a horse leaping a hurdle.

But to fully grasp the XG-M's advanced design, you have to hold it.

The body feels rugged yet light. With a built-in textured grip that's sculpted to fit snugly in your hand. And oversized controls that make it easy to adjust to changing conditions.

As your creative potential develops, you'll have access to over 45 interchangeable, computer-designed Minolta lenses. As well as the Minolta system of SLR accessories.

All in all, the XG-M is a remarkable achievement. But then, we have over 50 years of remarkable achievements to draw on.

The new Minolta XG-M.
Now we know how it feels to be alone.
But we're not complaining.

WAIT 'TIL YOU SEE HOW GOOD YOU CAN BE.

MINOLTA

For more information, write Minolta Corporation, 101 Williams Drive, Ramsey, N J 07446
Or see your Minolta dealer. In Canada: Minolta, Ontario, L4W 1A4.

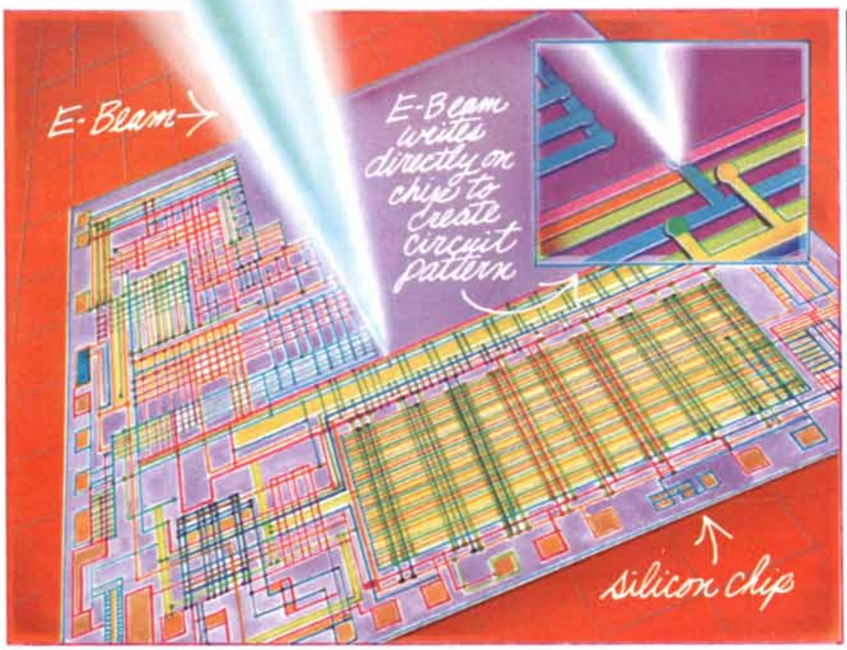
Product appearance and/or specifications are subject to change without notice the waterline forward and aft of the estimated position of the mainmast, and parts of the bow and stern as well. Almost from the start of this work the divers came on the remains of whale-oil barrels, crushed but virtually complete. Some were randomly dispersed but others were found where they had been stowed on their side, wedged with ballast rocks and short billets of wood to prevent shifting. The barrels would have held about 54 gallons of oil. By the end of the 1980 season the underwater archaeologists had raised and recorded more than 10,000 barrel fragments.

The ship's keel was some 15 meters long; on the basis of this dimension her cargo capacity can be estimated at about 300 tons. The extreme rake of the bow and the stern, the smooth curve of the stempost and the design of the transom are all compatible with 16th-century naval architecture. Other features, however, reflect older shipbuilding technology. For example, the after end of the keel rises upward in a "knee" and is fastened to the sternpost in a manner similar to that followed in the construction of the medieval merchant vessels known as cogs. (Marine archaeologists recently uncovered one of these vessels in the harbor of Bremen.)

Except for the barrels only a few artifacts have as yet been found in the wreck. An anchor was uncovered some 22 meters east of the keel; it is now being treated for the effects of submersion. The capstan, one of the few that are known to date to the 16th century, and several rigging blocks have been brought up. Two sections of bilge-pump tubing have been located but not raised. Perhaps the most interesting single artifact is a two-meter wrought-iron *verso,* or swivel gun. These breech-loading weapons, mounted on the ship's rail, were commonly called "murderers" because they often did as much damage to those who fired them as to those they were aimed at. In any event another puzzle is presented by the fact that on examination the *verso* was found to be loaded. One wonders why.

Individual ship's timbers are being raised so that tracings and scale drawings can be made. As the excavation proceeds the location of each structural member and even of fragments of wreckage is precisely recorded. The information will be used to build a scale model of the wreck as it rests on the harbor bottom. When the excavation is complete, work can begin on a second scale model, with the various components reassembled to show the shape of the ship's hull. This reassembly will increase understanding of the shipbuilding design and technology responsible for the oceangoing vessels that carried to the New World 16th-century European explorers, colonists and fishermen.

OUR RESEARCH IN E-BEAM LITHOGRAPHY WILL MAKE OUR 1.25 MICRON VLSI CHIP LOOK LIKE HORSE AND BUGGY TECHNOLOGY.



WE'RE USING E-BEAMS TO DEVELOP VERY LARGE SCALE INTEGRATED CIRCUITS WITH SUB-MICRON FEATURE SIZE.

At Honeywell, solid state electronics is a key technology. We've already built one of the largest in-house semiconductor operations in the United States. But we're fast expanding our current capabilities to meet the growing demand for integrated circuits with extremely small feature size. Recently, our research in very large scale integrated circuits helped us win a contract for Phase 1 of the Department of Defense's VHSIC program.

In Phase O of VHSIC, we produced a bipolar microcircuit with 1.25 micron feature size and more than 7000 gates on a 200 X 250 mil chip. But that's only the first step. We're now pushing optical lithography to its limits, fabricating high-density, near-micron microcircuits. With the help of our advanced design automation system, we are developing a bipolar chip with 30,000 gates and a CMOS chip with 300,000 transistors.

At the same time, we're developing CMOS and bipolar microcircuits with 0.5 micron feature size. But, the micron barrier will not be broken with conventional optical lithography. Sub-micron technology will require significant advances in four major fields: electron-beam lithography, dry etching, ion implantation and low temperature processing.

For the last six years, Honeywell has been conducting e-beam research. Currently, computer controlled e-beam equipment is used to print the masks that are used in near-micron circuit printing. In August, 1980, we began using an e-beam system for direct writing of very large scale integrated circuits in the development laboratory at our Solid State Electronics Division.

One of the problems in e-beam direct writing is the proximity effects from high-energy electrons colliding with the atoms of the resist. At our Corporate Technology Center, we're working on an improved polymer resist that will not degrade during direct write.

Dry etching techniques, like plasma etching and reactive ion milling, are another focus of Honeywell research. Because dry etching techniques work vertically, they are preferred to wet etching, which tends to undercut the layers of the resist. At Honeywell, we're now in the process of designing our own reactive ion milling equipment.

We've taken tremendous strides in our VLSI capabilities. But there are still many possibilities in sub-micron chip technology yet to be explored.

If you are interested in learning more about Honeywell's research and development of VLSI, you are invited to correspond with Dr. J. M. Daughton, Vice President, Solid State Development Center. If you have an advanced degree and are interested in a career in systems analysis, solid state electronics, sensors, design automation, or material sciences, please write to Dr. K. C. Nomura, Vice President and General Manager, Solid State Electronics Division. Both may be reached at this address: Honeywell, 12001 State Highway 55, Minneapolis, MN 55441.

Honeywell

Although Honeywell engineering is world-wide, the bulk of corporately-funded research and applied research is done in Minneapolis. The most recent Quality of Life Study conducted by Midwest Research Institute shows Minnesota to be one of the best places in the country to live and work considering cultural, social, economic, educational and political factors.

This ideal environment is further enhanced by Honeywell's affiliations with universities across the country. We have an ongoing program of seminars with Berkeley, MIT, Stanford, Carnegie Mellon, the University of Illinois, Cornell, Purdue, Oregon Graduate Center, and the University of Minnesota.

THE AMATEUR SCIENTIST

The pleasures of the pinhole camera and its relative the pinspeck camera

by Jearl Walker

The staggering variety of sophisticated cameras now on the market obscures the fact that quite acceptable photographs can be made with nothing more than a pinhole between the film and the object being photographed. The same is true with the optical complement of a pinhole: a "pinspeck," which is a small, circular spot placed between the film and the object. In discussing pinhole photography I shall be following the work of Kenneth A. Connors of the University of Wisconsin at Madison and Matt Young of the National Bureau of Standards. The novel idea of pinspeck photography comes from Adam Lloyd Cohen of Loyola University in Chicago.

Pinhole photography relies on the passage of light through a small hole in an opaque screen. The light falls on a piece of film to construct an image of the object being photographed. Images from pinholes were mentioned by Aristotle, explained in principle by Leonardo da Vinci and analyzed formally by Lord Rayleigh. Simplicity is only one of the advantages the pinhole camera offers over a camera with a lens.

When an object is photographed with a pinhole camera, each point of the object facing the camera casts a corresponding spot of light onto the film (or photographic paper). The composite of the spots is the image recorded by the film. If the image is to be clear, the adjacent spots should not overlap. Therefore they should be as small as possible.

Part of the design of a pinhole camera lies in choosing a size for the pinhole and a distance between the hole and the film that will make the individual spots of light cast on the film remain discrete and as bright as possible. In principle the pinhole can be of unlimited size. The larger it is, however, the farther it must be placed from the film and the larger the film must be, so that in practice the size of the hole is limited. There is also a theoretical limit on how small a pinhole can be.

Consider a camera in which the pinhole is a reasonable distance (several centimeters) from the film. Suppose the pinhole is relatively large (too large for the distance to the film) and the camera faces a distant point source of light. The rays of light from the source arrive at the pinhole essentially parallel to one another and to the central axis that passes through the hole and is perpendicular to the screen. The radius of the circular spot of light cast on the film is equal to the radius of the hole. Since the pinhole is large, the spot of light is large. If many point sources of light were being photographed, the spots on the film would overlap and the individual point sources would not be recognizable.

The size of the spot of light from a single point source is reduced if the size of the pinhole is reduced. The scope of this improvement is limited, however, because eventually the pinhole is so small that the light passing through it diffracts into an interference pattern. A point source of light creates not a single small spot of light on the film but a circular pattern consisting of one central bright spot surrounded by dimmer rings. If the size of the pinhole is reduced further, the diffraction pattern gets larger, with a resulting loss of resolution in the photograph.

The optimum radius for the pinhole is related to the distance between the hole and the film. The relation can be shown by a theoretical argument that depicts light as being in wave form. Imagine that the pinhole and the screen have been removed. A light wave from a point source travels through the plane formerly occupied by the screen.

Consider a family of circular zones on that plane and concentric on the camera's central axis. The zones are distinguished by their path lengths to a point at the center of the film that is also on the central axis. The distance between the central zone and the center point on the film is the distance that was between the pinhole and the film. The second zone is farther from the center point by half a wavelength of light, the third by an additional half wavelength and so on.

All the zones send light waves to the center point, but because of the differences in the path lengths the waves interfere when they arrive. For example, the wave from the second zone arrives half a wavelength out of phase with the wave from the central zone. If the amplitudes of the two waves were equal, the waves would cancel each other. Indeed, if all the contributions were equal in amplitude, they would all cancel at the center point.

In reality, however, the amplitudes are not equal, as is shown in descriptions more precise than those I can supply here, so that the cancellation is only partial. The net amplitude of the light wave at the center point turns out to be half the amplitude the central zone would have contributed on its own. Because the brightness of the light is related to the square of the amplitude this result means that the brightness at the center point is a fourth of what it would be if only the central zone were contributing light.

One of the purposes of a pinhole is to block all the zones except the central one. (Some investigators say it blocks all but the first two zones.) A pinhole of optimum size allows only the central zone to send light to the center point on the film. With a pinhole of that size the spot of light at the center point will be bright and small with a good distribution of light. If the pinhole is smaller than this optimum size, only part of the central zone contributes light at the film. The spot of light is dimmer and the distribution of light is poorer. If the pinhole is too large, the additional zones in it decrease the brightness of the spot and increase its size.

What one seeks, then, is not a particular size of pinhole but rather a particular relation between the size of the pinhole and the distance from the hole to the center point on the film. When an object to be photographed is relatively distant, the optimum radius of the pinhole is approximately equal to the square root of the product of the wavelength of the light and the distance between the pinhole and the film.

From this relation a focal length for the pinhole can be defined. The pinhole acts as a lens in the sense that it concentrates an image of an object. The focal length is approximately the wavelength of the light divided into the square of the radius of the pinhole. The spot of light on the film is small and bright, with a good distribution of light, when the film is distant from the pinhole by the focal length. Then only the central zone fills the pinhole and contributes light to the center point.

Suppose the object is close. If you photographed it through a lens, you could calculate the proper distance between the lens and the film by applying what is called the thin-lens equation, which states that the inverse of the distance between the lens and the film should be equal to the inverse of the lens's focal length minus the inverse of

PRESENTING YOUR NEXT TWO AUTOMOBILES. THE 2-WHEEL/4-WHEEL DRIVE EAGLE.

IT'S A SWITCH FROM EVERY CAR IN THE WORLD.

You are looking at the automotive breakthrough of the decade. The new American Motors Eagle. The only automobile in the world that gives you both 2-wheel drive fuel economy...*and* the traction and security of full-time 4-wheel drive. It's like having two great cars in one—with just the flick of a switch.

32 HWY EST **23** EPA EST MPG*

WHICH MEANS

that now you can have driving control even in deep snow... or driving rain. Where the roads are steep... or unpaved...or non-existent.

And no matter where the road leads you, you won't have to leave the comforts behind. Eagle's soft reclining seats, plush carpeting, power steering, power brakes, and fully independent front suspension will see to that.

SECURITY...FOR YOUR WALLET.

The confidence of owning an Eagle isn't just a matter of traction. It's knowing you're backed by the *only* full 12-month/12,000 mile warranty available *anywhere*. And shielded by Ziebart® Factory Rust Protection. Surrounded by galvanized steel in 100% of its exterior body panels. And covered by our exclusive full 5-year No-Rust-Thru™ Warranty.

We don't know a car company in Japan, Germany *or* the USA that builds and stands behind its products like American Motors.

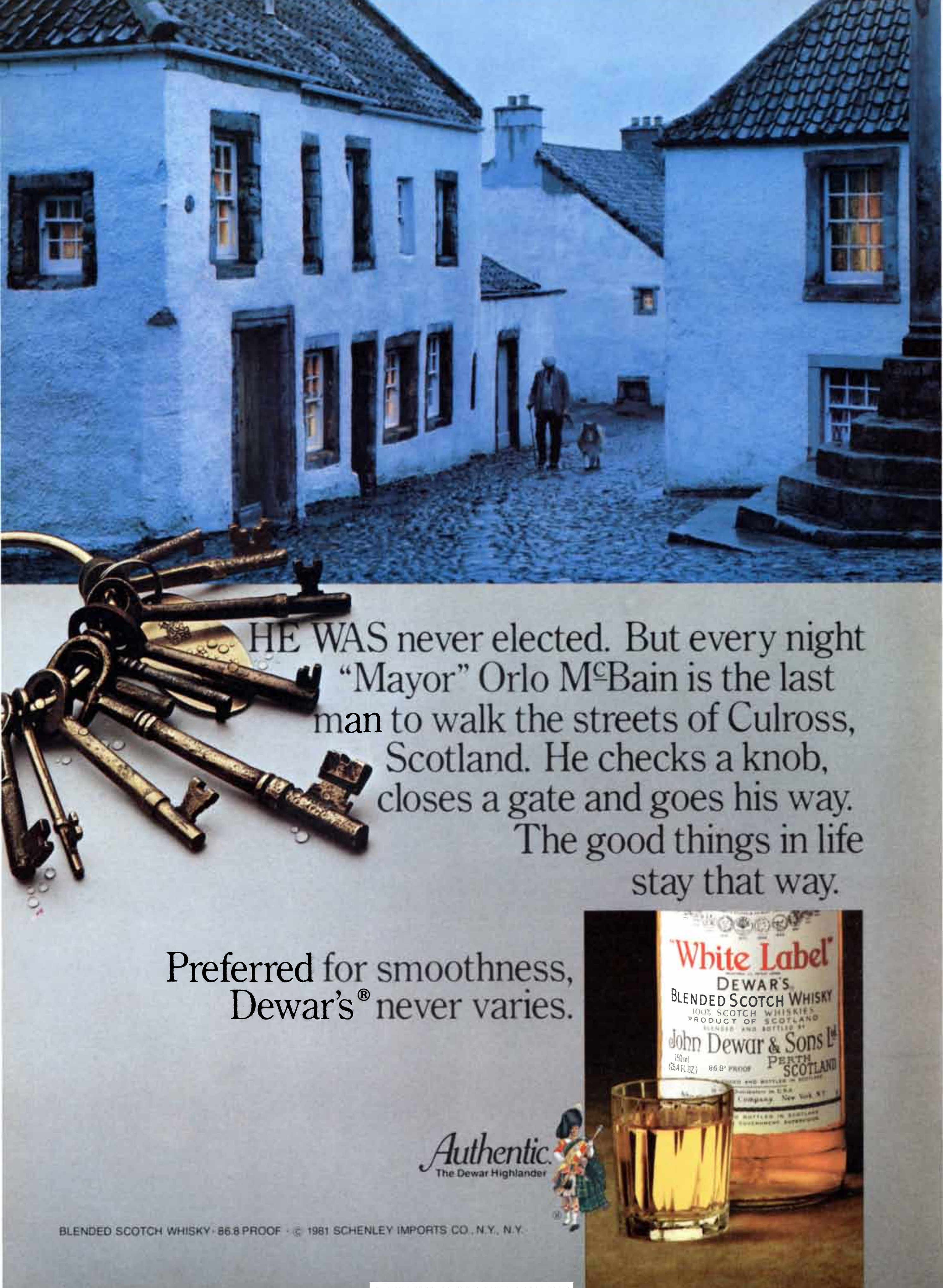
SO GET IN YOUR 2-WHEEL DRIVE CAR

and get to your nearest American Motors dealer for *two* test drives in the revolutionary 1982 Eagle:

First in 2-wheel drive, then *(Click)* in 4-wheel drive. It's a switch from every car in the world.
Available in 2-door or 4-door models.

THE 2-WHEEL/4-WHEEL DRIVE
EAGLE
FROM AMERICAN MOTORS

*Optional 5-speed stick. Use these figures for comparison. Your results may differ due to driving speed, weather conditions and trip length. California figures lower.


HE WAS never elected. But every night "Mayor" Orlo McBain is the last man to walk the streets of Culross, Scotland. He checks a knob, closes a gate and goes his way. The good things in life stay that way.
Preferred for smoothness, Dewar's® never varies.
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
John Dewar & Sons Ltd
PERTH SCOTLAND
Authentic.
The Dewar Highlander
BLENDED SCOTCH WHISKY · 86.8 PROOF · © 1981 SCHENLEY IMPORTS CO., N.Y., N.Y.

the distance to the object. The same relation holds for a pinhole if the focal length is defined in the way I have described. Thus a pinhole camera can be focused in order to make a photograph with the best resolution.

For example, if the object is far away, the best position for the film is at the focal length of the pinhole. If you walk toward the object and thereby decrease the distance between it and the pinhole, you must increase the distance between the film and the pinhole in order to maintain the optimum resolution. Such an adjustment may not be very practical, since in a pinhole camera the distance between the pinhole and the film is usually fixed. Instead you could replace the pinhole with a smaller one so that the focal length is smaller.

In practice neither adjustment is made because the resolution in the photograph is usually acceptable even if the size of the pinhole and the distance between the pinhole and the film are suboptimal. If you photograph a scene in which objects are at a large range of distances from the camera, most of them will be acceptably in focus in the photograph. This large depth of field is a characteristic of the pinhole camera.

From what I have said you could calculate either the appropriate size for the pinhole or the distance between the pinhole and the film once one of them has been picked. How do you make the first choice? Practicality bears on the answer: you do not want a pinhole camera that is several meters long. You also want to turn out a finished photograph that has as much detail as you would see looking directly at the scene. The desire for resolution is the starting point in the initial choice of conditions for the camera.

The limit of resolution of your eye is measured in terms of angle. Suppose your field of view encompasses two points. You can distinguish them as long as the angle between them is larger than a certain minimum value, approximately .001 radian. If the angle is smaller, you see only a single, blurred object. For example, if two adjacent points are separated by one millimeter and are one meter away from you, they would be just at the limit of your ability to resolve them. A camera with that degree of resolution would be sufficient; improving its resolution would add nothing.

Assume for the sake of demonstration that the final photograph is the same size as the film and that it will be viewed at a distance equal to the distance between the pinhole and the film. You are to photograph two adjacent point sources of light whose angular separation is at the limit of resolution (.001 radian) of your eye. The camera should cast two spots of light on the film that barely touch or overlap slightly. When the photograph is viewed, you will just be able to resolve the spots. The angle between the spots in your field of view can be calculated by dividing the diameter of the pinhole into the wavelength of the light. Suppose the wavelength is 500 nanometers (about in the middle of the visible range). If the angle for the limit of resolution is .001 radian, the pinhole's radius should be .25 millimeter.

Once this choice has been made the optimum distance between the pinhole and the film can be calculated (by the relation I have already set out) to be 12.5 centimeters. If you made the pinhole twice as large and adjusted the distance of the film from the hole accordingly,

A photograph made with pinhole optics by Kenneth A. Connors the resolution of a photograph from the camera would be twice as good. If the size of the photograph and the distance at which you view it are unchanged, however, you would not be able to see the improvement. Moreover, the camera would now be 50 centimeters long (in order to have the proper distance between the pinhole and the film) and larger film would be needed to capture all the light from the pinhole. Clearly the improvement is not worthwhile.

When the pinhole is larger than it should be, the poorer resolution can actually add erroneous detail to the photograph. This effect, called spurious resolution, results from the overlap of the images from several adjacent objects. Young's demonstration of the spurious resolution of three vertical bars appears in the illustration at the bottom left on the opposite page.

Most lens systems cause a linear distortion in an image recorded on film. For example, a square object might appear to have slightly curved sides. Most modern cameras incorporate corrections for the problem. One of the advantages of a pinhole camera is that it is virtually free of linear distortion.

The pinhole camera does have several types of aberration, including chromatic aberration. Since the optimum radius of the pinhole (and thus its focal length) depends on the wavelength of light, the camera cannot be optimized for more than one wavelength. The resolution for that wavelength can be optimized but the resolution for the other wavelengths in white light will be poorer.

The result with color film is a blurring of the edges of an image and perhaps some noticeable color along the edges. With black-and-white film only the blurring of the edges is visible. One way to eliminate the chromatic aberration is to use black-and-white film with a color filter placed in front of the pinhole. Optimize the size of the pinhole and the distance of the film from the pinhole for the wavelength passed by the filter.

A wide-angle view made by John M. Franke's camera with a glass hemisphere behind the pinhole

All other colors are eliminated and the edges of the image are blurred less by chromatic aberration.

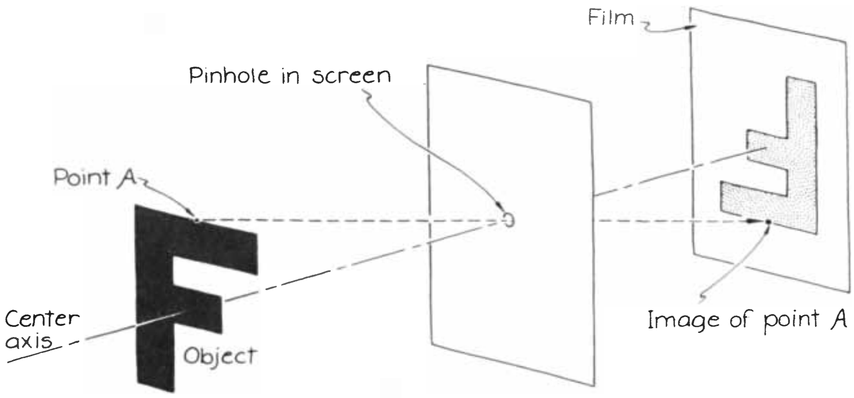
Another aberration with the pinhole camera is astigmatism. It arises when an object being photographed lies off the central axis of the pinhole. The shape of the pinhole perpendicular to the object is elliptical rather than circular. If the object is a point source of light, an elliptical spot is cast on the film. In addition the place on the film where the spot falls will not be at the proper distance from the pinhole. If the center of the film is put at the proper distance from the pinhole, any other point on the film is too far from the hole, which means that the resolution is not optimum anywhere but at the center.

A severer problem with the pinhole camera is its low light-gathering ability. Since the aperture is usually tiny, relatively long exposures are required. For example, if the film is at the pinhole's focal length and if that distance is a few centimeters, the f number of the camera is approximately 200. Although the small aperture makes the system slow, it is responsible for the camera's large depth of field.

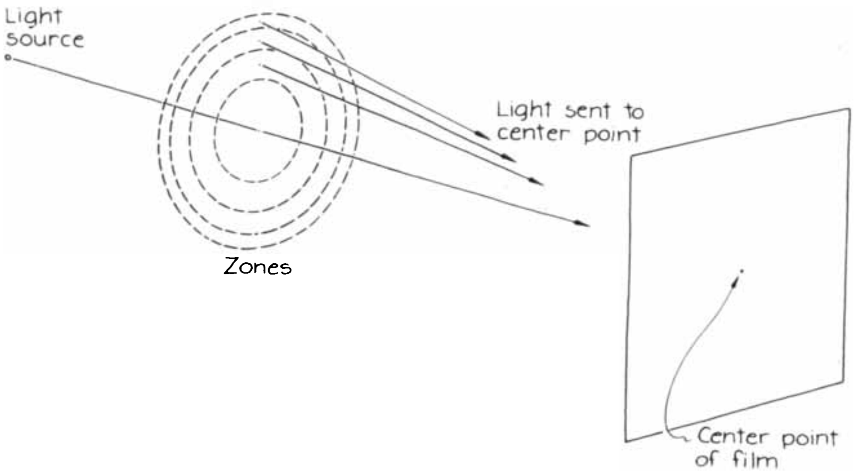
For several reasons the intensity of the light cast on the film is nonuniform. Suppose two point sources of light are being photographed, one on the central axis and one off it. The light from the off-axis point source encounters a pinhole that is effectively elliptical. Therefore less light travels through the hole from the off-axis source than from the point source on the axis. In addition the light forming the off-axis spot must travel farther to reach the film and so spreads more, thereby arriving at the film with less intensity. Moreover, this light reaches the film at an angle that further spreads the exposure over more of the film, reducing the intensity even more. These losses over the width of the film establish a practical limit to the field of view.

Another limit to the field of view is that an object sufficiently off the central axis may not reflect light to the film unless the film is quite wide or fairly close to the pinhole. The usual solution to the problem has been to move the film closer to the pinhole so that a wide-angle photograph can be made. The trouble is that this stratagem reduces the resolution of the photograph because the film is no longer at the right distance from the pinhole for optimum resolution.

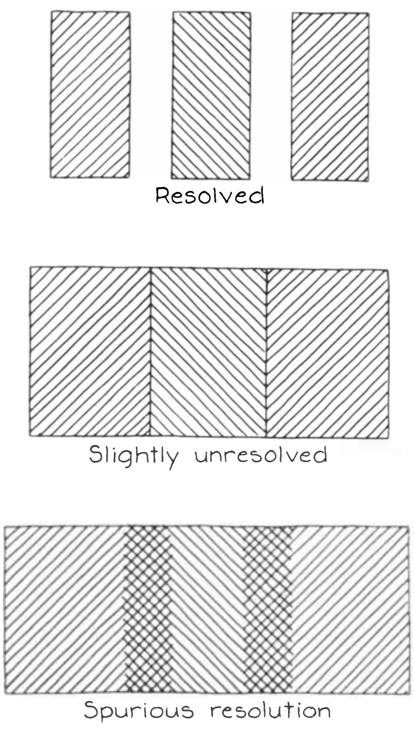
Another way to increase the field of view is to design a film holder that is hemispherical with respect to the pinhole. Then any light entering the pinhole, even light from an object that is almost 90 degrees off the central axis, will reach the film. Another result would be less of a decline in the exposure far from the center of the film because the light would always strike the film per-



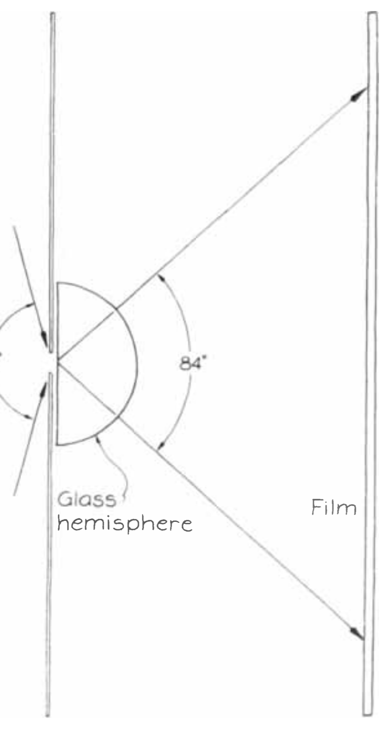
An arrangement for pinhole photography



The zones contributing light to the center of a film



Three types of resolution



Franke's setup for wide-angle pictures pendicularly. The resolution of the objects off the central axis would also be improved, since all sections of the film would be at the same distance from the pinhole. Unfortunately a hemispherical film holder is not very practical. A cylindrical one might be an adequate compromise.

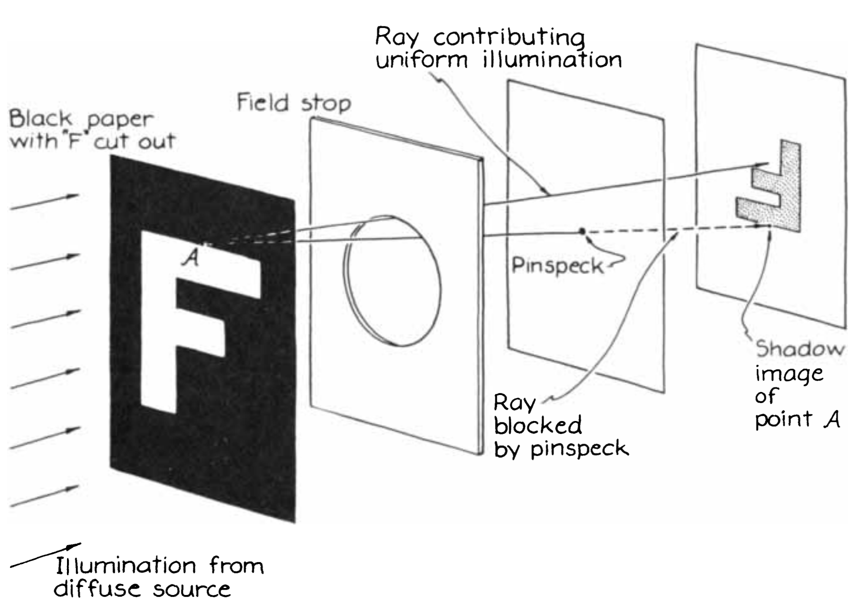
Another solution to the problem was invented by John M. Franke of the National Aeronautics and Space Administration's Langley Research Center. Franke positions a glass hemisphere just behind the pinhole of a camera in which the film is held in a normal flat plate. As the light passes through the pinhole and into the glass it is refracted. The full field of view, which occupies an angle of 180 degrees, is reduced to a cone of light occupying an angle of 84 degrees. When the light emerges from the glass, it is perpendicular to the surface of the glass. Hence the angle of the cone of light is unaltered. The reduction in the angle from 180 to 84 degrees enables Franke to position the film at an appropriate distance from the pinhole and still make a wide-angle photograph with a field of view of approximately 180 degrees.

Franke's glass hemisphere is made from BK-7 glass and is 25.4 millimeters in diameter. Its index of refraction is about 1.5. The diameter is not critical, but different results are obtained with glass that has a different index of refraction. You might like to experiment with other glasses or even with plastic of good quality. If you want a field of view of 180 degrees, you will encounter some distortion of the image toward the edges of the photograph.

You can form a pinhole in several ways. Take care to make a circular hole with smooth edges. Young has made clean pinholes in brass shim stock 50 micrometers thick. He mounts a sewing needle in a milling machine and then with the machine's vertical feed forces the needle through the thin brass sheet. He puts a freshly smoothed lead block under the brass to prevent distortion of the sheet. After removing the burrs on the edge of the hole he reams it with a needle point and cleans it again.

Connors uses brass shims .001 or .002 inch thick. Thicker plates are undesirable because the hole is then more of a cylinder and generates more internal reflection of the light rays. A square piece of the shim is placed on firm cardboard or smooth soft wood. With a needle point Connors gently pushes a dimple into the center of the shim piece, being careful not to push the point entirely through. He turns the piece over and rubs the small mound on the back of the dimple with a fine emery cloth until it is removed. He repeats the procedure, perhaps as many as 15 times, until a hole appears and gets large enough for the shaft of the needle to go through it. He has previously measured the diameter of the needle shaft with a microscope that has a graduated reticle, and so he now knows how large the pinhole is. If he wants a pinhole that is smaller than his smallest needle, he stops the enlargement process before the needle fully enters the hole.

Once the pinhole is complete Connors cements the shim to a thicker brass sheet (.005 inch thick) for support. The pinhole lies over a 1/4-inch hole drilled in the thicker piece. The side of the assembly that is to face the film is painted with a flat black to diminish any reflections of light inside the camera. Some people think the interior of the pinhole should also be blackened, but Connors does not want to degrade the symmetry of the hole he made, and so he paints only to within a millimeter or two of it.



Adam Lloyd Cohen's setup for pinspeck photography

Connors notes that a pinhole should be kept free of dust. He stores his pinhole assembly in a plastic bag until the assembly is needed. Periodically he examines the pinhole with a microscope to check for any degradation of the symmetry resulting from dust.

The assembly of the brass sheet and the shim can be mounted on virtually any type of lighttight box. I have seen pinhole cameras made with cereal boxes. Working in a darkroom, the photographer mounts a piece of photographic paper at the back of the box and slides on the lid. A piece of black tape is put over the pinhole to prevent light from entering the box prematurely. When everything is ready, the tape is pulled back from the hole for the exposure and then put back over the hole. Although such a camera functions as a pinhole camera, it has two disadvantages: only one photograph can be made before the camera is returned to the darkroom, and the removal and repositioning of the tape might shake the box too much, blurring the photograph.

I chose to follow a procedure outlined by Young. On the base of his 35-millimeter camera he mounted an extension tube, which is available for most cameras with removable lenses. At the outer end of the tube he attached his pinhole assembly. Lacking an extension tube, I used a cardboard mailing tube that I attached to my camera base with several layers of black tape. The advantage of this type of pinhole camera is that an entire roll of film can be exposed. Since my camera is a single-lens reflex model, I could actually see a dim image of the scene I was to photograph before I made a picture.

Whereas in pinhole photography light passes through a hole to create an image, in Cohen's pinspeck photography a pinspeck casts a negative image of an object. His setup is the optical complement of the pinhole. The screen and hole are replaced with a small obstacle of circular cross section. Now all the light that would travel through a pinhole is blocked. All the light that would have been blocked by the screen reaches the film, forming a negative image. The final pinspeck photograph is similar to the pinhole photograph except that bright and dark areas are exchanged.

The image cast by Cohen's pinspeck does not depend on the diffraction of light because the pinspeck is too large to give rise to a significant diffraction pattern. The image is created by the simple blocking of light rays from the object. Any particular spot on the film records the shadow of a section of the object lying on a straight line extending from the spot through the pinspeck and to the object.

A photograph made with a pinspeck displays poorer contrast than a pinhole photograph because the pinspeck ar-

SCIENCE/SCOPE

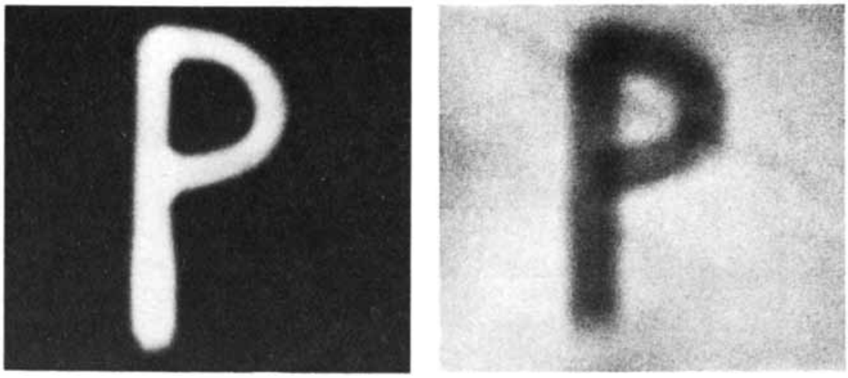
<u>A new window material for infrared sensors to see through</u> has shown to be highly resistant to damage from nuclear radiation. The material, produced using a reactive atmosphere process recently developed for oxide material, is a glassy silica called fused cristobalite. In tests at Hughes the material suffered no damage when exposed to gamma radiation of 1 million rads. Conventional fused silica, though known to be one of the materials least affected by radiation, is heavily discolored by doses even 100 times smaller. Fused cristobalite has slightly different physical characteristics from other fused silicas, such as a higher melting point, but it maintains the same high optical quality.

<u>An all-optical digital computer has been demonstrated</u> at Ohio State University using a liquid-crystal light valve. This unique Hughes device accepts optical images and replicates them on a completely separate light beam from an arc lamp or laser. The device uses technology similar to that of liquid-crystal watches. Optical equivalents to electronic logic gates and flip-flops were constructed with the light valve much as transistors are used in an electronic system. Computers that use photons instead of electrons would be smaller and faster.

<u>In a major advance in computer-aided design (CAD)</u> that eliminates the need for breadboard models, Hughes engineers have created software that checks integrated circuits. The CAD module, called VISTA, tests all hardware from transistors through logic gates, registers, chips, or an entire system. It can test circuit functions and timing parameters, and can help develop and verify system software and test algorithms. The module includes a virtually new simulation language based on PL/1. The CAD module will be used in developing very large-scale integrated circuits for missiles and other military electronics.

<u>Hughes Research Laboratories need scientists</u> for a whole spectrum of long-term sophisticated experiments. Advanced research programs include three-dimensional microelectronics, digital picture processing, space optics, solid-state devices, fiber optics, integrated optics, integrated circuit design, and electro-optical materials. For immediate consideration, please send your resume to Professional Staffing, Dept. SE, Hughes Research Laboratories, 3011 Malibu Canyon Road, Malibu, CA 90265. Equal opportunity employer.

<u>Measuring millimeter-wave impedance is easier now</u> by using a new IMPATT-sourced millimeter-wave reflectometer instead of slotted lines and hybrid impedance bridges. Because of its unique design, the Hughes device can sweep the full 75-to-110 GHz bandwidth automatically without time-consuming point-to-point measurements. It uses such advanced technology as broadband isolators and detectors to reduce the effects of unwanted reflections and mismatch error, a 1-KHz square wave modulator that delivers increased sensitivity and dynamic range, and a rotary vane attenuator for constant attenuation with frequency.



Cohen's photographs by pinhole **(left)** *and pinspeck* **(right)** *of a* **P** *cut in paper*

rangement allows nearly all the light from a scene to reach the film. Most of it is a uniform illumination that is of no value and merely reduces the contrast of the image. It is the remainder of the light, the nonuniform portion, that carries the information about the object. Contrast would be improved if somehow the uniform illumination were diminished or the proportion of the nonuniform light carrying the information about the object were increased.

The variation of light can be increased if the pinspeck is positioned closer to the film, but just as with pinhole photography this arrangement decreases the resolution of the photograph. Cohen says he does sacrifice some of the resolution to achieve enough contrast in the photograph to create a recognizable image.

Some of the uniform illumination results from parts of the scene that are not important to the photograph. To decrease this unimportant illumination Cohen places a field stop (a screen with a hole larger than the pinspeck) in front of the pinspeck. The hole is large enough for the extreme parts of the object to illuminate the edges of the film but small enough to prevent the rest of the scene from reaching the film.

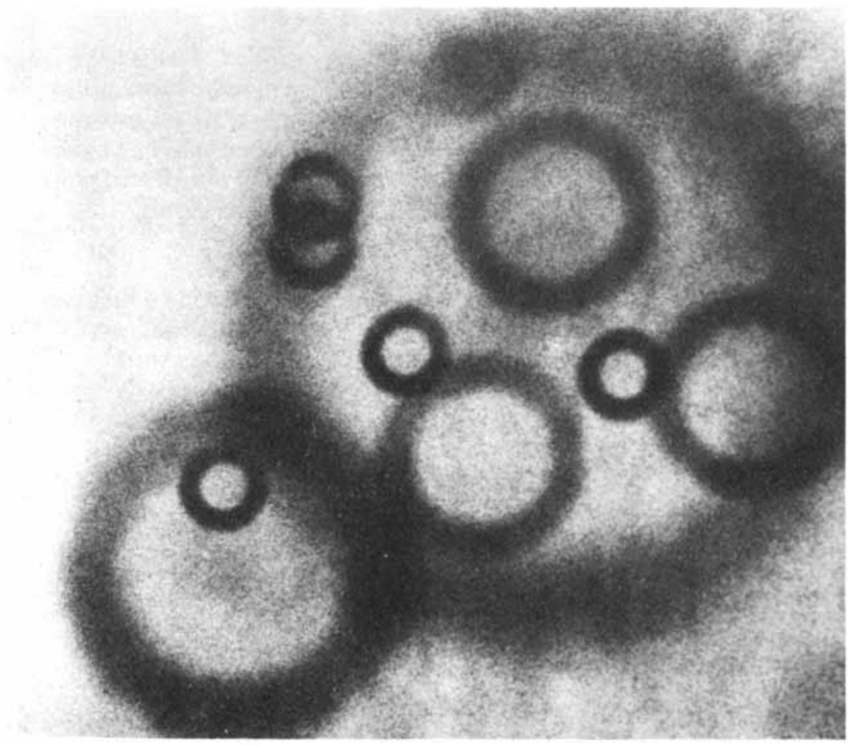
The coloring of pinspeck images can be strange. If a small collection of colored objects is photographed, the image of each object will probably be of a different color from that of the object. The change depends on the combined colors of the objects in the collection. If the combination is white, each color in the collection is switched to its complement in the photograph. For example, a red object will form an image in a color that is the subtraction of red from white (because the pinspeck blocked the red from the object). Therefore the color of the shadow is cyan, the complement of red. Correspondingly, a green object creates a magenta shadow.

Some of the properties of a pinhole camera are displayed equally well by a pinspeck camera. The field of view is large, the adjustment to the magnification of the camera is made by changing the distance between the pinspeck and the film, and there is no linear distortion. Astigmatism can be avoided with a pinspeck camera if the pinspeck is spherical. Then any light traveling from the object to the film intercepts an obstacle with a circular cross section even if the object is well off the central axis of the camera.

One other difference in the two types of photography is that a series of pinholes aligned between the object and the film will not produce a photograph but a series of pinspecks will. The screens perforated with the pinholes keep the light from falling on the succession of pinholes closer to the film. In contrast the pinspecks barely interfere with one another's ability to form an image of the object. The illustration at the bottom left is a photograph Cohen made from a series of pinspecks that he positioned between a single bright ring and the film. Each pinspeck produces its own negative image of the ring.

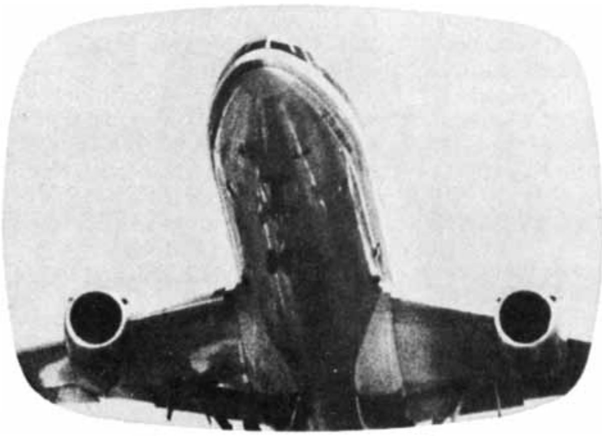
If you would like to make pinspeck photographs, Cohen offers the following suggestions. For a pinspeck place a dot of black paint on a piece of clear glass or acetate. The shape of the dot is not critical. Instead of paint you could paste on a small circular dot. (I find such dots in office-supply stores. They are for labeling purposes.) Cohen recommends that the dot not be too small or the contrast in the photograph will be too low. The scene photographed should have high contrast so that the photograph will also. You could begin your experiments with pinspeck photography by cutting figures in a black, opaque sheet of paper and then illuminating the sheet from behind with a diffuse source of light.

Much more can be learned about pinhole and pinspeck photography than I have set out here. A description of Cohen's work will appear soon in *Optica Acta* under the title of "Anti-pinhole Imaging." Some of the most thorough work on pinhole photography, both experimental and theoretical, can be found in a series of papers by Connors in *Interest,* a journal he edits. It is available from him at the School of Pharmacy, University of Wisconsin, Madison, Wis. 53706. His recent papers have dealt with the conditions for optimum resolution and definition, the calibration of a camera for contrast control and the relation between a pinhole camera and the optics of a zone plate.

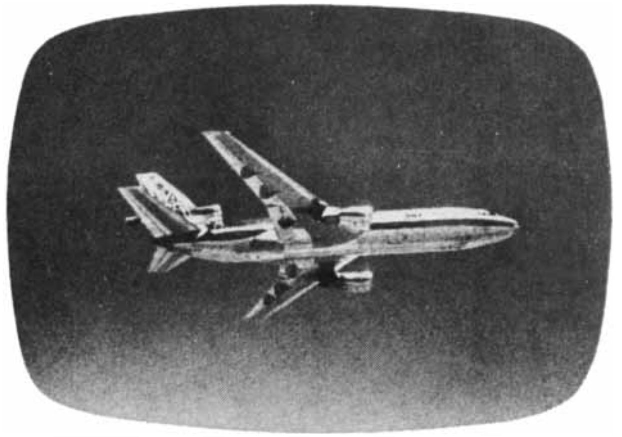


A single bright ring photographed by Cohen through a series of pinspecks

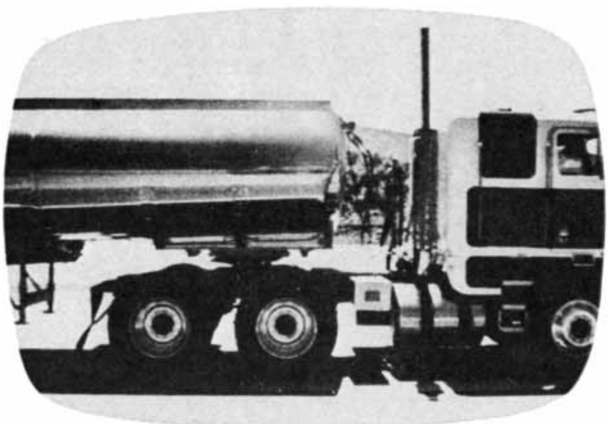
Lightweight aluminum saves energy in transportation.



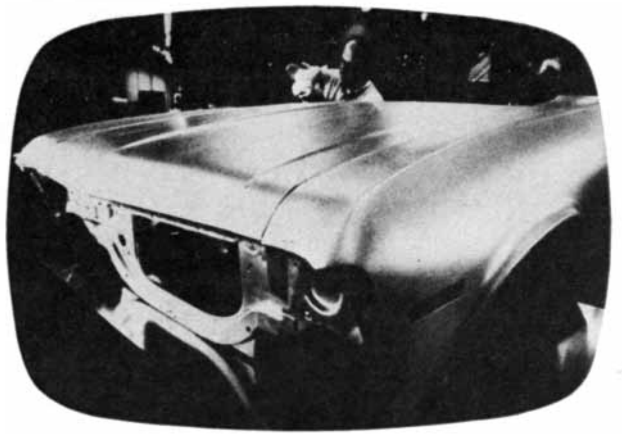
If you've ever wondered how something so big got off the ground, part of the answer is aluminum.



Aluminum's high strength and light weight helped to make modern aviation possible.



Aluminum is important to *most* forms of transportation. Because taking the weight off planes, trucks . . .



and now cars, saves fuel.

Aluminum in transportation.
It's a better way to get there—today.

ALCOA

We can't wait for tomorrow.
For more information write Alcoa,
523-L Alcoa Building, Pittsburgh, PA 15219.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

METAMAGICAL THEMAS

SIMPLE MATHEMATICAL MODELS WITH VERY COMPLICATED DYNAMICS. Robert M. May in *Nature,* Vol. 261, No. 5560, pages 459–467; June 10, 1976.

ITERATED MAPS ON THE INTERVAL AS DYNAMICAL SYSTEMS. Pierre Collet and Jean-Pierre Eckmann. Birkhauser Boston, Inc., 1980.

UNIVERSAL BEHAVIOR IN NONLINEAR SYSTEMS. Mitchell J. Feigenbaum in *Los Alamos Science,* pages 4–27; Summer, 1980.

THE CENSUS OF 1980

HISTORICAL STATISTICS OF THE UNITED STATES: COLONIAL TIMES TO 1970. Bureau of the Census, U.S. Government Printing Office, 1975.

POPULATION. Philip M. Hauser in *Social Statistics in Use.* Russell Sage Foundation, 1975.

STATISTICAL ABSTRACT OF THE UNITED STATES. Bureau of the Census, U.S. Government Printing Office, 1980.

THE MICROCHANNEL IMAGE INTENSIFIER

CHANNEL ELECTRON MULTIPLIERS. Three special issues of *Acta Electronica:* Vol. 14, No. 1, January, 1971; Vol. 14, No. 2, April, 1971; Vol. 16, No. 1, January, 1973.

SECONDARY ELECTRON EMISSION MULTIPLIERS AS PARTICLE DETECTORS. W. E. Baumgartner and W. K. Huber in *Journal of Physics E: Scientific Instruments,* Vol. 9, No. 5, pages 321–330; May, 1976.

MICROCHANNEL PLATE DETECTORS. Joseph L. Wiza in *Nuclear Instruments and Methods,* Vol. 162, pages 587–601; 1979.

INDUCIBLE REPAIR OF DNA

E. COLI RECA PROTEIN-DIRECTED CLEAVAGE OF PHAGE λ REPRESSOR REQUIRES POLYNUCLEOTIDE. Nancy L. Craig and Jeffrey W. Roberts in *Nature,* Vol. 238, No. 5742, pages 26–30; January 3, 1980.

DNA-DAMAGING AGENTS STIMULATE GENE EXPRESSION AT SPECIFIC LOCI IN *ESCHERICHIA COLI.* Cynthia J. Kenyon and Graham C. Walker in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 77, No. 5, pages 2819–2823; May, 1980.

RECA PROTEIN PROMOTES HOMOLOGOUS-PAIRING AND STRAND-EXCHANGE REACTIONS BETWEEN DUPLEX DNA MOLECULES. Stephen C. West, Era Cassuto and Paul Howard-Flanders in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 78, No. 4, pages 2100–2104; April, 1981.

PURIFIED *LEXA* PROTEIN IS A REPRESSOR OF THE *RECA* AND *LEXA* GENES. John W. Little, David W. Mount and Celeste R. Yanisch-Perron in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 78, No. 7, pages 4199–4203; July, 1981.

RINGS IN THE SOLAR SYSTEM

THE RINGS OF SATURN. James B. Pollack in *Space Sciences Review,* Vol. 18, No. 1, pages 3–93; October, 1975.

DISCOVERING THE RINGS OF URANUS. James L. Elliot, Edward Dunham and Robert L. Millis in *Sky and Telescope,* Vol. 53, No. 6, pages 412–416, 430; June, 1977.

JUPITER'S RINGS. Tobias Owen, G. Edward Danielson, Allan F. Cook, Candice Hansen, Virginia L. Hall and Thomas C. Duxbury in *Nature,* Vol. 281, No. 5731, pages 442–446; October 11, 1979.

ENCOUNTER WITH SATURN: VOYAGER 1 IMAGING SCIENCE RESULTS. Bradford A. Smith, Lawrence Soderblom, Reta Beebe, Joseph Boyce, Geoffrey Briggs, Anne Bunker, Stewart A. Collins, Candice J. Hansen, Torrence V. Johnson, Jim L. Mitchell, Richard J. Terrile, Michael Carr, Allen F. Cook II, Jeffrey Cuzzi, James B. Pollack, G. Edward Danielson, Andrew Ingersoll, Merton E. Davies, Garry E. Hunt, Arnold Masursky, Eugene Shoemaker, David Morrison, Tobias Owen, Carl Sagan, Joseph Veverka, Robert Strom and Verner E. Suomi in *Science,* Vol. 212, No. 4491, pages 163–191; April 10, 1981.

THE FLUID PHASES OF MATTER

THE STRUCTURE OF LIQUIDS. J. D. Bernal in *Scientific American,* Vol. 203, No. 2, pages 124–134; August, 1960.

STATES OF MATTER. David L. Goodstein. Prentice-Hall, Inc., 1975.

THEORY OF SIMPLE LIQUIDS. Jean Pierre Hansen and Ian R. McDonald. Academic Press, 1976.

WHAT IS "LIQUID"? UNDERSTANDING THE STATES OF MATTER. J. A. Barker and D. Henderson in *Reviews of Modern Physics,* Vol. 48, No. 4, pages 587–671; October, 1976.

THE COLOR PATTERNS OF BUTTERFLIES AND MOTHS

POSITIONAL INFORMATION AND PATTERN FORMATION. Lewis Wolpert in *Current Topics in Developmental Biology,* Vol. 6, pages 183–224; 1971.

WING PATTERN FORMATION IN LEPIDOPTERA: A MODEL. H. Frederik Nijhout in *The Journal of Experimental Zoology,* Vol. 206, No. 2, pages 119–136; November, 1978.

ONTOGENY OF THE COLOR PATTERN ON THE WINGS OF *PRECIS COENIA* (LEPIDOPTERA: NYMPHALIDAE). H. Frederik Nijhout in *Developmental Biology,* Vol. 80, No. 2, pages 275–288; December, 1980.

PATTERN FORMATION ON LEPIDOPTERAN WINGS: DETERMINATION OF AN EYESPOT. H. Frederik Nijhout in *Developmental Biology,* Vol. 80, No. 2, pages 267–274; December, 1980.

THE LINING OF THE SMALL INTESTINE

MORPHOLOGY OF THE EPITHELIUM OF THE SMALL INTESTINE. J. S. Trier in *Handbook of Physiology, Section 6: The Alimentary Canal,* edited by John Field. American Physiological Society, 1968.

TRANSPORT ACROSS THE INTESTINAL MUCOSAL CELL: HIERARCHIES OF FUNCTION. D. S. Parsons and C. A. R. Boyd in *International Review of Cytology,* Vol. 32, pages 209–255; 1972.

THE INTESTINAL CELL SURFACE: PROPERTIES OF NORMAL, UNDIFFERENTIATED, AND MALIGNANT CELLS. Kurt J. Isselbacher in *The Harvey Lectures,* Vol. 69, pages 197–221; 1973–74.

FIFTEEN YEARS OF STRUGGLE WITH THE BRUSH BORDER. Robert K. Crane in *Intestinal Absorption and Malabsorption,* edited by T. Z. Csáky. Raven Press, 1975.

A 16TH-CENTURY BASQUE WHALING STATION IN LABRADOR

LES BASQUES DANS L'ESTUAIRE DU SAINT-LAURENT, 1535–1635. René Bélanger. Les Presses de l'Université du Québec, 1971.

LOS BALLINEROS VASCOS EN CANADA ENTRE CARTIER Y CHAMPLAIN. Selma Huxley Barkham. Real Sociedad Bascongada de los Amigos des Pais, 1979.

DIVERS FIND SUNKEN BASQUE GALLEON IN LABRADOR. Selma Huxley Barkham and Robert Grenier in *Canadian Geographical Journal,* Vol. 97, No. 3, pages 60–63; July, 1979.

THE AMATEUR SCIENTIST

PINHOLE IMAGERY. M. Young in *American Journal of Physics,* Vol. 40, No. 5, pages 715–720; May, 1972.

FIELD-WIDENED PINHOLE CAMERA. John M. Franke in *Applied Optics,* Vol. 18, No. 17, pages 2913–2914; September, 1979.

Beyond quartz, the world's most precise tuning system, lies a new ability to expand sound.

Imagine you're in a room with Technics SA-828 receiver. What you hear is beautiful stereo. Then you activate Technics variable Dimension Control. Incredibly, the sound begins to move. The stereo image widens to the point where the music begins to surround you. You're intrigued by its richness and depth. You're enveloped by a new experience in sound. That's the wonder of the patented technology in Technics Dimension Control.

Just as wondrous is quartz synthesis, the world's most precise tuning system. That's how the SA-828 quartz synthesizer eliminates FM drift as well as the hassle of tuning. You can even preset and instantly retrieve 7 FM and 7 AM stations, all perfectly in tune.

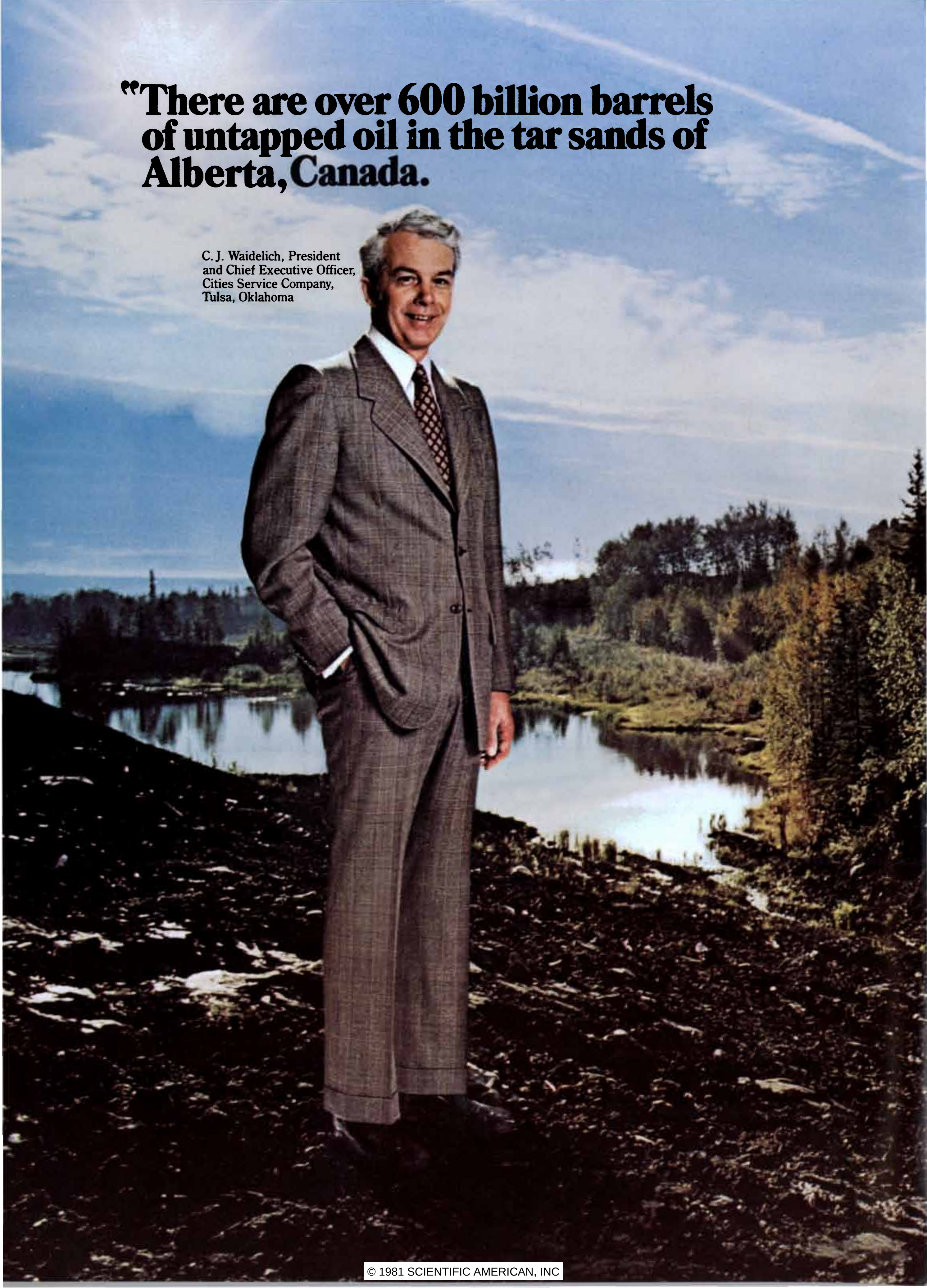
Another perfect example of Technics technology is our synchro-bias circuitry. What it does is constantly send minute amounts of power to the amplifier transistors. And since they can't switch on or off, switching distortion is eliminated.

And when it comes to power, the SA-828 has plenty: 100 watts per channel minimum RMS into 8 ohms from 20Hz to 20kHz with no more than 0.005% total harmonic distortion.

The SA-828 goes on to show its sophistication with a super-quiet phono equalizer, soft touch program selectors, fully electronic volume control, and a Dimension Control display that doubles as a power level meter.

Technics SA-828 is part of a full line of quartz synthesized receivers. Hear it for yourself. Beyond its quartz synthesizer lies a new dimension in sound.

Technics
The science of sound


"There are over 600 billion barrels of untapped oil in the tar sands of Alberta, Canada.
C. J. Waidelich, President and Chief Executive Officer, Cities Service Company, Tulsa, Oklahoma


And Cities Service is on top of it."
Two years ago Cities Service, along with other industry and government participants, made the Syncrude tar sands project here in Alberta, Canada, a commercial reality, with current production of over 100,000 barrels of high-grade synthetic crude oil each day.
But that project extracts tar sands close to the surface. However, 90 percent of these untapped resources lie at depths of 150 feet or more. Too deep for surface mining.
Now Cities Service, in joint venture with other companies, is working on three pilot projects to heat the deep tar sands in place, releasing the bitumen from the sands and allowing it to be pumped to the surface.
One method, for example, is to dig a pit down to the level of the deep tar sands. Drill horizontal holes into the side of the pit, inject steam into the holes which alters the viscosity of the bitumen, allowing it to flow through the holes for recovery.
How important is this effort? The deep tar sands contain an estimated 550 billion barrels of oil. And up to 20% of that amount may be recovered with these techniques—up to 110 billion barrels... almost one-third the proven reserves of the Middle East. And the more oil North America has, the less control OPEC will have over world oil production and prices. In addition, the advanced technology we develop in Alberta, combined with increased capital investment, could help convert America's own 130 billion barrels of tar sands and heavy oil deposits into synthetic crude oil.
So putting steam into the tar sands of Alberta could put more steam in our country's ability to produce more energy.
CITIES SERVICE COMPANY
On top of the problem, part of the solution.


IMPORTED
Seagram's
V.O.
CANADIAN WHISKY