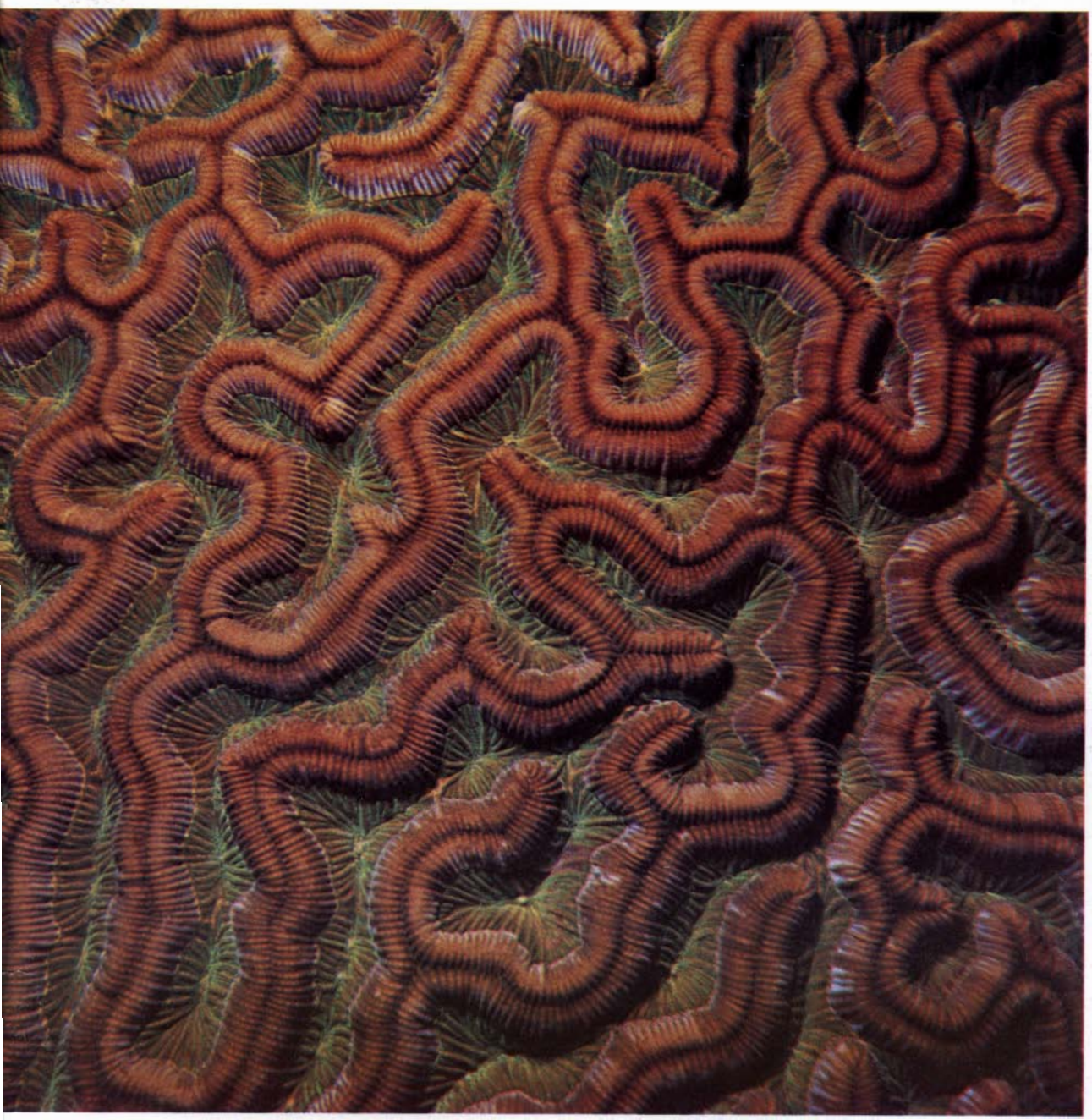
SCIENTIFIC AMERICAN

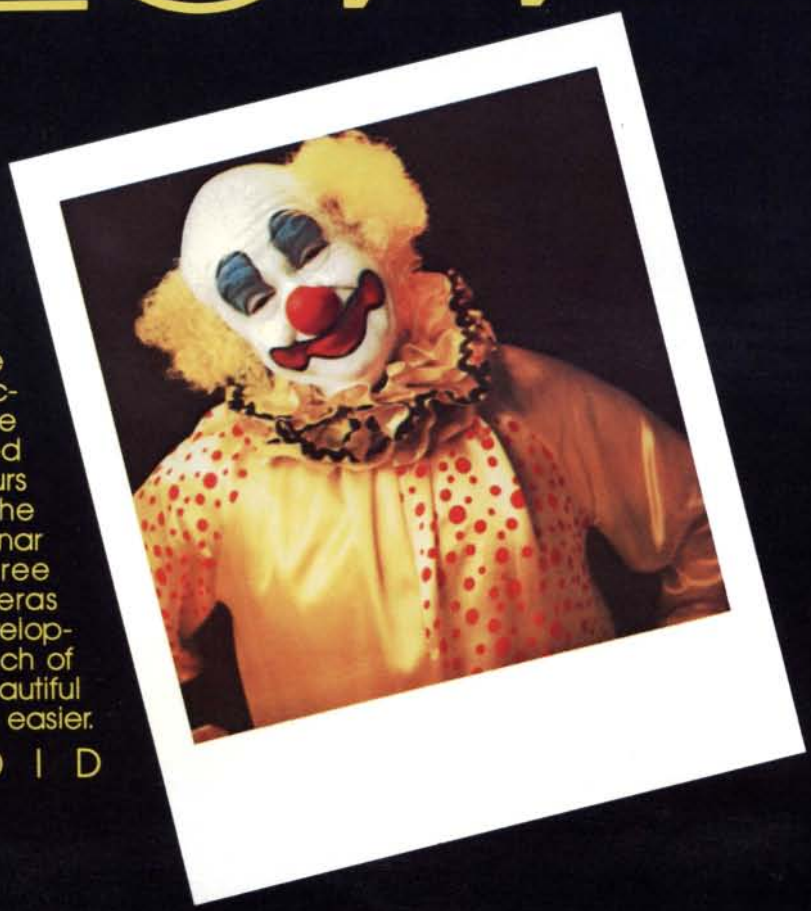


CORAL

$1.50

August 1979

SX-70 YELLOW



How's this for yellow? Bright. Cheerful. And notice how crisp the red, white and blue colors are, too. This picture was made with the same SX-70 film used by millions of amateurs and professionals. The camera: the SX-70 Sonar OneStep, one of three OneStep Land cameras that hand you a developing picture at the touch of one button. SX-70: beautiful color has never been easier.

P O L A R O I D

How a phone call solved the mystery of the sandy teacups.

Based on an actual call made to the toll-free 24-hour Whirlpool Cool-Line® service.

(Telephone Rings)

Cool-Line Consultant: Whirlpool Cool-Line. May I help you?

Woman: I just bought a Whirlpool dishwasher and I keep finding sand in my teacups. Can you help me?

Consultant: That's why I'm here. Now, about the sand. Are the rest of your dishes clean?

Woman: They're fine. My husband's a Mexican food freak. Even pans with baked-on refried beans get clean. But where did the sand come from?

Consultant: What does the sand look like?

Woman: Like...sand. In a puddle of water that didn't drain out of the teacup.

Consultant: If you're seeing "sand," it could be your dishwasher detergent hasn't dissolved. Do you have a cup with some "sand" in it now?

Woman: Right here by the phone.

Consultant: Does the "sand" look like detergent?

Woman: You mean this is detergent?!?

Consultant: Look closer.

Woman: It does look like detergent. So why didn't it dissolve?

Consultant: Check your water temperature. At your dishwasher, it should be at least 140°. If it's okay, then I suggest you buy a fresh box of detergent. Dishwasher detergent sometimes has a very short shelf life and doesn't dissolve completely when it's old. And make sure you load your teacups properly, so all the water drains out.

Woman: Wow. You really helped. Sorry I bothered you, but at least I didn't have to call a repairman. Thanks for your time.

Consultant: Glad I could help.

This is the kind of two-way communication we've been having with our Whirlpool Cool-Line service for the past eleven years. It's just one example of the continuing concern we have for customers who purchase quality Whirlpool appliances.

If you ever have a question or problem with your Whirlpool appliance, call our toll-free 24-hour Cool-Line service at 800-253-1301. In Alaska and Hawaii, dial 800-253-1121. In Michigan, call 800-632-2243. If our Cool-Line service can't help, we have Whirlpool franchised Tech-Care® service representatives all over the country who can.

Whirlpool
Home Appliances
Quality. Our way of life.

41% BETTER EPA ESTIMATED GAS MILEAGE IN OUR MOST EXCITING PONTIAC PHOENIX EVER.

What a difference a model year can make.

That's all it's taken for our all new 1980 Pontiac Phoenix to score 41 percent better EPA *estimated* MPG over its 1979 counterpart. (Based on a comparison of 1979 and 1980 Pontiac Phoenix models with standard powertrains.* The highway percentage increase is 42 percent.)

24 **37**

EPA ESTIMATE MPG **HWY ESTIMATE**

It's all the more exciting when you consider these 1980 Phoenix mileage estimates.

Remember: Compare the circled *estimated* MPG with that of other cars. Your mileage may vary depending on speed, trip length and weather. And your actual highway mileage will probably be less than the highway estimate. Standard powertrain not currently avail. in Calif.

But great mileage is just the beginning of the excitement. Phoenix for 1980 has been dramatically redesigned from the ground up. Choose from the rakish new five-passenger Phoenix Coupe and the sporty and versatile Phoenix Hatchback.

You get an exciting new chapter of Pontiac driving fun. With front-wheel drive. Rack and pinion steering. And MacPherson strut front suspension.

Another exciting thought. You get even more standard features than on our well-equipped '79 Phoenix. Such new standards as radial tuned suspension, integral body-color bumpers, column-mounted dimmer/washer switch, dome lamp switches on all the doors, load floor Hatchback carpeting and much more.

That's more Pontiac.

So don't let the excitement pass you by. Buy or lease the all new Phoenix Coupe and Hatchback at your Pontiac dealer's. It's going to be quite a year.

*Phoenix is equipped with GM-built engines produced by various divisions. See your dealer for details.

MORE PONTIAC TO THE GALLON

GM

Established 1845 **SCIENTIFIC AMERICAN** August 1979 Volume 241 Number 2

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017

executive health

the report that briefs you on what to watch

Volume XIV, Number 10 • July, 1978 • Pickfair Bldg., Rancho Santa Fe, Calif. 92067 • Area 714:756-2600

New research findings . . .

ON WALKING . . . nature's own amazing *"anti-age antibiotic"!*

Not running, not jogging, but *walking* is your most *efficient* form of exercise and the *only* one you can *safely* follow *all* the years of your life!

Today, after some 2000 years, modern medical researchers are demonstrating (somewhat to their surprise) that Hippocrates, "the father of medicine" was right . . . walking is man's best *"medicine"!*

Take one example. At a three-day conference on *Exercise in Aging—Its Role in Prevention of Physical Decline* (held October 27-29, 1977, at the National Institute of Health, Bethesda, Maryland) researchers from across the United States, Canada and Western Europe presented papers on this (until now) largely neglected area of research. As their papers were presented, these important points of agreement emerged:

(1) Walking is the most *efficient* form of exercise . . . and the only one you can safely follow all the years of your life.

(2) Exercise can enable your body to maintain a *vital reserve* which has a protective effect during stress.

(3) Exercised *bones* do not demineralize. As a result they are far less likely to break or lose their range of motion.

(4) Exercised *lungs* still exhibit the emphysema-like changes of age, but are far less diminished in their capacity compared to the lungs of sedentary people.

(5) Exercised *cardiovascular systems* show a similar maximum preservation of function.

(6) The benefits of exercise in preventing or correcting obesity are striking.

(7) Late-onset diabetes is almost entirely reversible by exercise if you are overweight.

(8) Daily exercise permits greater food intake and better blood circulation, thus improving each body cell's nourishment while preventing obesity.

(9) The physically impaired, particularly the arthritic, can perhaps benefit the most from exercise . . .

How Long Will You Live?

Ignorance about how to take care of yourself can be the mistake that kills you in your 50's or 60's . . . "the dangerous years"!

Discover what research scientists now know about how to take the best care you can of the only body you will ever have.

Do as so many thousands of executives do. Subscribe to Executive Health Report. The members of our ***Editorial Board*** are among the world's most distinguished authorities on *preventive medicine*. Their wise advice can help you not only live longer but enjoy those extra years! (Note their high qualifications.)

Executive Health Report is not sold on newsstands but only by private subscription at $24 a year in the U.S.A. and its possessions, Canada and Mexico. Individual reports (back issues) $2.00 per copy. All other countries $26 by surface mail, $30 by air mail. (Only International Money Order or check cashable on U.S. bank will be acceptable.)

Subscribe now under this unusual introductory offer:

(1) Your choice of any three of the reports listed below ($2.00 each) FREE!

(2) If you are not satisfied with your first issue, your money will be *promptly* refunded.

Please study the reports listed here and circle your three choices:

New research findings On Walking . . . nature's own amazing "anti-age antibiotic"! Not running, not jogging, but walking is your most efficient form of exercise and the only one you can safely follow all the years of your life!

Linus Pauling, Ph.D.: For the best of health, how much vitamin C do you need? People who take the optimum amount of vitamin C may well have, at each age, only one quarter as much illness and chance of dying as those who do not take extra vitamin C.

New research findings On Zinc . . . the amazing metal so essential to your health.

George C. Griffith, M.D.: On those irregular Heart Beats (Arrhythmias). Some mean little or nothing, but others warn your heart is in trouble. Have you had any irregular heart beats?

On The Syndrome Of Longevity. Key factors dependably present ("constants") in the way of life of men who have outlived other men born when they were by a generation!

Roger J. Williams, Ph.D., D.Sc.: On Your Startling Biochemical Individuality. Some amazing facts about your body you need to know if you want to understand yourself (and other people) better.

Jeffrey A. Gray, Ph.D.: On Anxiety and hope to cope with it!

John K. Lattimer, M.D., Sc.D.: On that treacherous gland, your prostate — especially as you approach 50.

Linus Pauling, Ph.D.: What About Vitamin E? Eminent investigators now suspect it may be one of the key factors to help resist disease and slow the aging process.

On "The Horizontal Exercise." New research findings about sex and how to keep your middle age young!

The truth about the right diet and exercise to lose weight safely and stay slender for life.

Go easy gentlemen, too much social drinking damages your liver (even though you may feel well, eat well, and never get tight!).

On stress and the executive. Dr. Hans Selye, world-famous endocrinologist, explains why hard work is the secret of living a long, satisfying and successful life.

If you get a heart attack, what then? Dr. Paul Dudley White gives you some amazing case histories of long life!

Travel risks travel agents don't talk about! Things you should know and do if you want to fly the world and come home healthy!

Alton Ochsner, M.D.: On "The Chair Disease" . . . Why blood clots in your veins are a little-realized hazard of "desk-bound" executives.

Stephen R. Elek, M.D. On "The Hurry-up Disease." Why it may be a key factor that triggers heart attacks as early as age 40.

James F. Toole, M.D.: Can An Aspirin A Day Keep A Stroke Away?

If you sit at a desk all day . . . How to avoid a Pot-Belly and Double-Chin (or get rid of them for good!)

Jan Koch-Weser, M.D.: On Systolic Hypertension, more common than diastolic hypertension and as dangerous. What it is, what it does, and what to do about it.

Linus Pauling, Ph.D.: On Vitamin C and Cancer. Recent studies show that vitamin C has a large life-extending effect for patients with advanced cancer and suggest a similar large effect for earlier stages of the disease.

Leonard Hayflick, Ph.D.: On Getting A Good Night's Sleep! What research scientists now know about this mysterious process we so calmly call "sleep."

Albert Szent-Gyorgyi, M.D., Ph.D.: On A Substance That Can Make Us Sick (If We Do Not Eat It!) One of the world's most honored scientists, winner of the 1937 Nobel Prize for Physiology and Medicine explains the fascinating paradox of vitamins.

New Discoveries: On The Dangers of Not Getting Enough Magnesium in the Food You Eat. Chronic deficiency can cause heart disturbances ranging from premature beats to deadly ventricular fibrillation with its useless fluttering of the heart.

Please use the coupon below under our special money-back guarantee.

EXECUTIVE HEALTH, Pickfair Bldg., Rancho Santa Fe, CA 92067
Gentlemen: Enclosed is my check for $_____ for a year's subscription to Executive Health to start with this month's issue. I have circled the three $2.00 reports I am to receive free. It is understood that if I am not satisfied with my first issue, my money will be refunded. NOTE: In filling out coupon below please include *all* information necessary to insure delivery. Failure to do so risks postal loss of letter.

Firm/Name (Please print clearly)_____

Street & No./Route_____ **Suite or Apt.**_____

City, State, Country_____ **Zip**_____ **SA-2**

Now this doctor can search through a medical library half a world away, and never leave her office.

Like doctors anywhere, this doctor in Paris sometimes comes across a case that's hard to diagnose.

But these days she can turn to the National Library of Medicine in Washington, D.C.—through a computer terminal right in the office.

In less time than it takes to tell about it, she's able to search through over *half a million* medical references stored in the Library's computers.

To help her find out what ails her patient.

This wealth of current medical knowledge is being brought to doctors, from France to Hong Kong, through the Universal Data Transfer Service (UDTS).

It's a data transmission network created by the people of ITT World Communications.

And today it's being used not only by medical people, but by people in business, finance, the sciences, agriculture, mathematics, marketing, whatever.

UDTS is surprisingly *economical,* too.

All of which makes it not only a remarkable tool, but something just as important.

A practical tool.

The best ideas are the ideas that help people. ITT

THE COVER

The photograph on the cover is a close-up of a living colony of the brain coral *Diplora,* situated on Elkhorn Reef in Small Hope Bay off Andros Island in the Bahamas. The fissures of the coral skeleton are lined with living polyps, which retract into the grooves during the day, as is shown here. At sunrise and sunset the polyps emerge to feed, immobilizing minute crustaceans and other zooplankton with the stinging cells on their tentacles. The remarkable ability of the living polyps to accrete enormous skeletons of calcium carbonate, thereby building up a massive wave-resistant structure, depends on their intimate partnership with symbiotic algae called zooxanthellae (see "Corals and Coral Reefs," by Thomas F. Goreau, Nora I. Goreau and Thomas J. Goreau, page 124). In spite of the low fertility of tropical seas, coral reefs harbor an enormous diversity of plant and animal species and are probably the most complex ecosystem.

THE ILLUSTRATIONS

Cover photograph by Douglas Faulkner

Page	Source
18–24	Ilil Arbel
41	Raymond Devoret, Laboratory of Enzymology, C.N.R.S., Gif-sur-Yvette
42	Ilil Arbel
43	Dirk Bootsma, Erasmus University, Rotterdam (*top*); Ilil Arbel (*bottom*)
44	Ilil Arbel
46	Raymond Devoret, Laboratory of Enzymology, C.N.R.S., Gif-sur-Yvette
47–49	Ilil Arbel
50–54	George V. Kelvin
55	Ned R. Sauthoff
56–57	George V. Kelvin
58	George V. Kelvin (*top*), Ned R. Sauthoff (*bottom*)
59–61	George V. Kelvin
63	Andrew M. T. Moore
64	Andrew Tomko
65	Andrew M. T. Moore
66–67	Alan D. Iselin
68–69	Andrew Tomko
70	Alan D. Iselin
125	Douglas Faulkner
128	Peter Arnold, Inc.
129	Thomas F. Goreau
130–131	Tom Prentiss
132	Tom Prentiss (*top and bottom right*), Robert K. Trench (*bottom left*)
133–135	Tom Prentiss
136	Thomas F. Goreau
139	William Herbst, Wesleyan University
140–142	Walken Graphics
145	Huberto Gerola and Philip E. Seiden, Thomas J. Watson Research Center, International Business Machines Corporation
146–156	Jerome Kuhl
159	LCR Graphics, Inc.
160–166	Gabor Kiss
168–169	LCR Graphics, Inc.
170–178	Gabor Kiss
181	Fritz Goro
182–184	Patricia J. Wynne
185	Joan I. Morrell, Rockefeller University
186	Patricia J. Wynne
188–191	Ben Rose
192–194	Michael Goodman
196	Michael Goodman (*top*), Ben Rose (*bottom*)

IMMEDIATE OPENINGS FOR THE OFFICE OF THE FUTURE. APPLY NOW.

Not long ago, if you mentioned to someone you were in the information management field, you were greeted with a polite smile and a vacant stare.

Today, you're liable to get a job offer.

Suddenly major corporations are hiring Directors of Information Management and Managers of Information Services. Who, when they arrive on the job, will find several hundred million pieces of information waiting to be managed.

Fortunately, they'll also find a company who can help them manage it: Xerox.

We offer Xerox copiers and duplicators that not only reproduce information, but also sort, reduce and copy on both sides.

Electronic printers that take information directly from computers—and then print it out with laser beams.

Electronic typing systems that type and store information—so you can readily find and revise it.

In fact, today there are Xerox machines that can perform almost any information management job. Including the most important one:

Helping information managers do their job.

XEROX

LETTERS

Sirs:

Michael Burgett's letter [SCIENTIFIC AMERICAN, June], referring to Jearl Walker's "The Amateur Scientist" for April, points out that the pollen grain of *Beta vulgaris* L. is dimpled like a golf ball and then says: "It is readily apparent that natural selection preceded man by more than a few million years in designing a sphere that would travel long distances."...

The facts are these. As almost any elementary book on fluid mechanics explains, it is only in a certain limited range of Reynolds numbers that roughening of a sphere acts to reduce the drag. The anomalous effect is due to the effect of boundary-layer behavior on flow separation and the formation of a low-pressure wake. At the extremely low Reynolds numbers applicable to the aerodynamics of pollen, however, the flow with respect to the pollen grain is essentially inertia-free, there are no boundary layers, there is no separated wake and indeed the roughening or dimpling of the surface has virtually no effect on drag.

Apart from this aerodynamic fact the irony is that grains of pollen travel long distances because of high drag, not low drag. If the drag force were small with respect to the weight of the pollen grain, the grain would fall to the ground, if not like a lead shot, then perhaps like a Ping-Pong ball. The true relation between aerodynamics and natural selection is that the smaller the pollen grain is, the farther it will be carried by wind, because the ratio of drag force to weight is inversely proportional to the square of the diameter.

Incidentally, Walker's article correctly states that I did experiments on dimpled v. undimpled golf balls [the film *The Fluid Dynamics of Drag,* National Committee for Fluid Mechanics Films, and the book *Shape and Flow: The Fluid Dynamics of Drag,* Doubleday, 1961], but it was from other sources that I reported that dimpled balls travel four times farther than smooth ones.

ASCHER H. SHAPIRO

Institute Professor
Massachusetts Institute of Technology
Cambridge, Mass.

Sirs:

I enjoyed reading "The Physics of Karate," by Michael S. Feld, Ronald E. McNair and Stephen R. Wilk [SCIENTIFIC AMERICAN, April]. It might be pointed out that karate strikes are of interest not only to physicists but also to physicians. In 1880 what is called march hemoglobinuria was first described in German soldiers who had walked long distances. In the early years of this century it was observed that a psychiatric patient in Boston who struck his head constantly with the palm of his hand also had developed hemoglobinuria. The cause of march hemoglobinuria was then understood: it is the breaking of red blood cells by trauma. March hemoglobinuria virtually disappeared when infantrymen were issued shoes with rubber soles and heels.

We still see the condition, however, in marathon runners, in bongo drummers and occasionally in basketball players. While I was working at Nippon University in Tokyo it occurred to me that we might find this condition in karate performers. Many individuals who practiced by regularly breaking bricks and boards, as did the authors of your article, had hemoglobinemia and even hemoglobinuria.

The purpose of this note is to warn people who participate in such activity with any regularity that they will probably break red blood cells. This, however, should not cause any permanent damage if the condition occurs only sporadically. Persistent hemoglobinuria, however, could lead to tubular necrosis of the kidneys. On rare occasions that could be fatal.

HERBERT BERGER, M.D.

Professor of Medicine
New York Medical College
New York

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 40 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

(We need your opinion)

A system with remote for less than the other people charge for just a telephone answering machine

This ad's a test. A kind of consumer survey. With a special Free Offer. To see if the low price of the new Call Jotter remote telephone answering system can turn one of the biggest selling business items into one that's successful with consumers, too.

Ordinarily, information like this comes from a consumer panel. Focus groups. But the manufacturer wasn't about to commit for the enormous sum required based on talk. He wanted facts. And came to us, as one of the largest mail merchandisers, for help. Because orders are facts he could act on.

Quality Features

For our part, we tested the Call Jotter thoroughly. And can tell you it's exceptionally well made. (It has to be to get our guarantee.) With solid state, microprocessor technology and plug-in simplicity. It's F.C.C. approved. And delivers the freedom and convenience you get with systems selling for $299.95—which is the going price, as you know, for remote telephone answering machines.

One thing we did tell the manufacturer: something extra should be given to those who participate in this test. He agreed. So, you'll receive with your order a FREE professionally recorded tape that answers and records 30 messages . . . a FREE blank tape for recording your own messages or for when you're using the machine as a cassette recorder and player . . . plus a FREE adapter for connecting the Call Jotter to your telephone jack.

An Extraordinary Convenience

Now, we ask you, how much would it be worth to you, to your wife, even your teen-age children to never miss or worry about a phone call again? And to get your messages without having to wait until you get home—from any phone, anywhere in the world. Resetting the machine to take 30 more messages by touching a button on the Tele-Key remote control.

Of course, you'll use your Call Jotter to answer the phone when you're working outside and when you're in a part of the house where there's no telephone.

How much would you spend for an answering machine when the phone rings and you're up on a ladder painting the house? It's true—isn't it—the phone always seems to ring at exactly the wrong time. Like the critical moment in your favorite show and whenever you're taking a nap.

Think of the time you've spent just waiting for someone who's promised to call. And what value do you place on your privacy . . . on working without interruption?

Because your Call Jotter has a monitoring system that lets you listen without answering, you can go back to what you were doing the instant you know it's a nuisance call or for someone who's out of the house. Naturally, you can take any call that's important.

Two For The Price of One

For additional value, Call Jotter works with a single cassette, like a pocket recorder or dictating machine, so you'll use it to listen to your favorite tapes and for recording your own tapes, for dictating letters and memos to be transcribed at the office.

Save $130.00!

You'll use your Call Jotter, then, when you're away—whether it's running to the corner store or spending a month in Europe.

Still—we agree, you probably wouldn't want to spend $299.95 for something that isn't business related. With Call Jotter, though, you save $130.00! And that's a different story.

At **$169.95** (plus $4.35 shipping and handling) Call Jotter's the lowest priced remote telephone answering machine you can get. (Without the remote, it's even less, only $99.95 plus $4.35 shipping and handling.)

Now, we invite you to discover the convenience and freedom it brings—especially if you're an active family—without risking one cent.

You can order either model with any national credit card simply by calling the toll free number below at any time. If you prefer, send your check to Douglas Dunhill at the address below. (Illinois residents are required to add the sales tax.)

Call 800-621-5554

Illinois Residents Call 800-972-5858
In operation 24 hours, 7 days a week

Remember, the low, down-to-earth price includes the Tele-Key remote control and the two FREE tapes plus the FREE adapter that fits your telephone jack. (If you don't have a phone on a jack, the telephone company will install one for a modest, one time charge when your system arrives.)

You Must Be Satisfied

Use your Call Jotter for 30 days. If you're not completely satisfied return it to us for a complete refund, no questions asked. Simply use the carton it comes in and follow the simple procedure in the directions we send you.

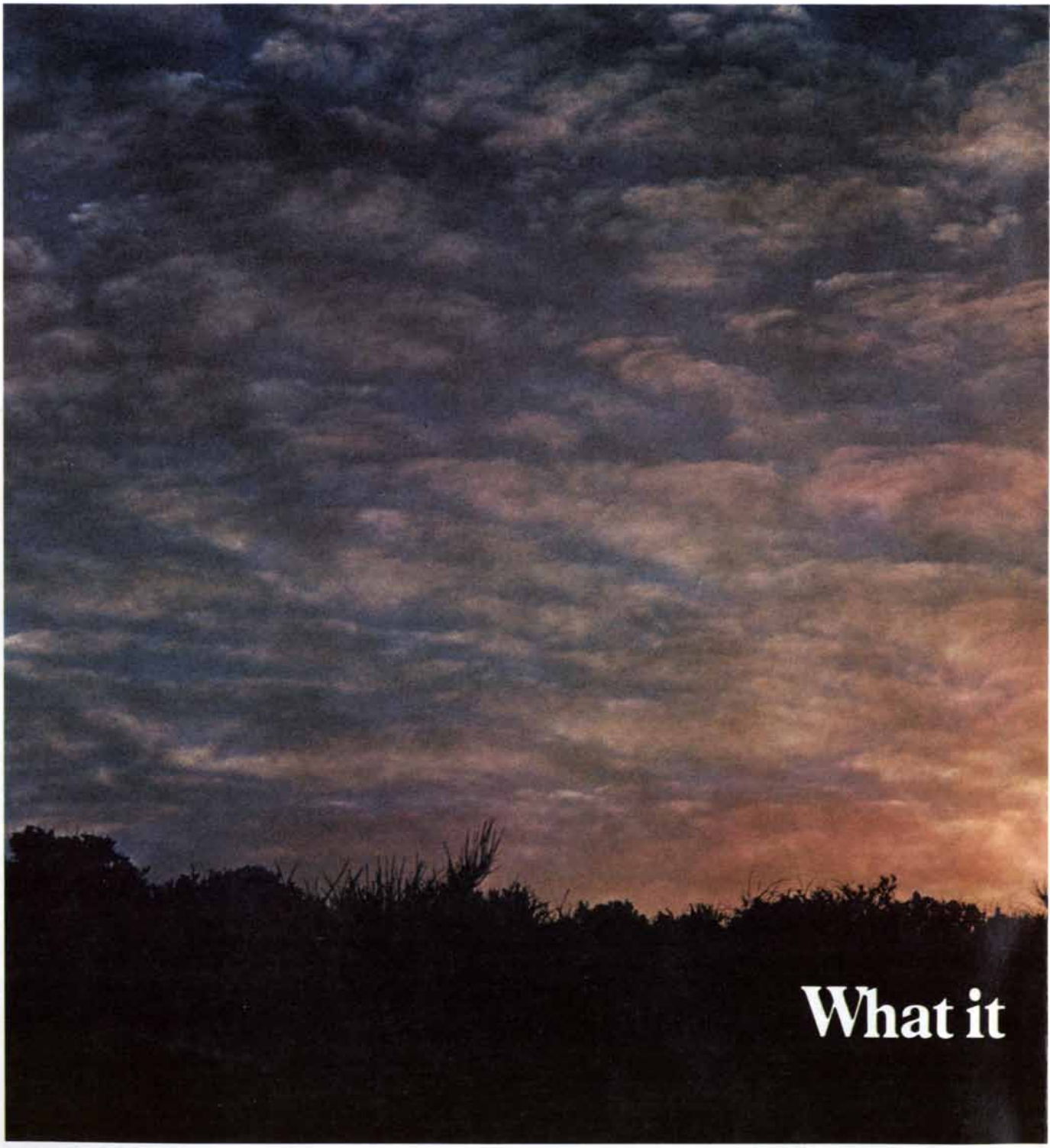
If this test is successful, the manufacturer will go into full production and you'll be seeing the unit in stores everywhere in six to nine months. You'll have played a part in this success—for which we thank you. Meanwhile, we'll be filling orders while we can from the supply on hand.

- Approved for connection in accordance with telephone company filed F.C.C. regulations
- Uses standard 60-minute cassettes
- Plugs into any phone on a jack with adapter supplied free
- Dynamic microphone, full fidelity speaker, push-button tape controls, call light, recorder-player operates on standard A.C. current
- Tele-Key complete with 9V battery for remote control from any phone anywhere in the world (2″ x 3″ x 1″—4 oz.)
- Hi-fi styling. Black and walnut color. Just 9½″ x 10½″ x 2¾″

The convenience and freedom you want . . . At the price you've been waiting for.

Douglas Dunhill
INC. AFFORDABLE QUALITY

Dept. 50-2578
Ten Douglas Dunhill Drive, Oak Forest, IL 60452



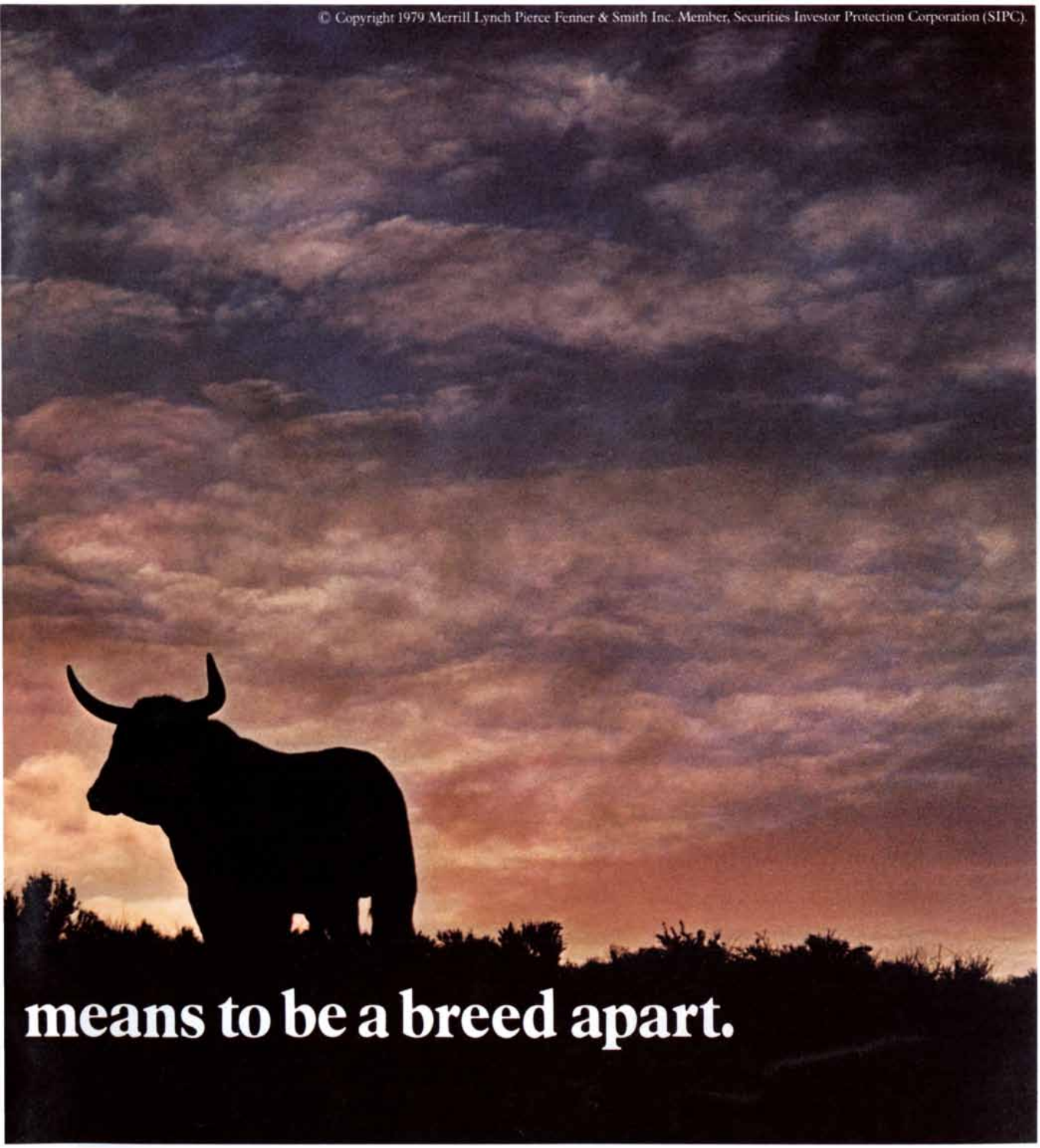
Imagination, instinct, and versatility (not size alone) set Merrill Lynch apart from the herd.

Too many people see only the size of Merrill Lynch; they miss the subtler qualities that truly make us a breed apart.

Imagination. Our investment bankers came up with a creative way to raise $75 million for an oil exploration company.

Traditional ways of raising the capital would have overly diluted earnings or impaired financial flexibility. So we suggested, developed and placed a public offering of convertible preferred stock—all within 40 days.

Instinct. One of our corporate clients wanted to buy back 200,000 shares of its own stock—in a week. Getting the best price was a question of market timing ...of knowing when to walk out of the market, when to watch and wait, when to move in and buy



aggressively. That's instinct.

And because of it, we were able to buy back the whole 200,000 shares within ½ point of the starting price.

Versatility. We can also help you reach your personal financial goals, despite changing markets and tough tax laws.

In 1953, a man walked into our St. Louis office and bought $487 worth of stock.

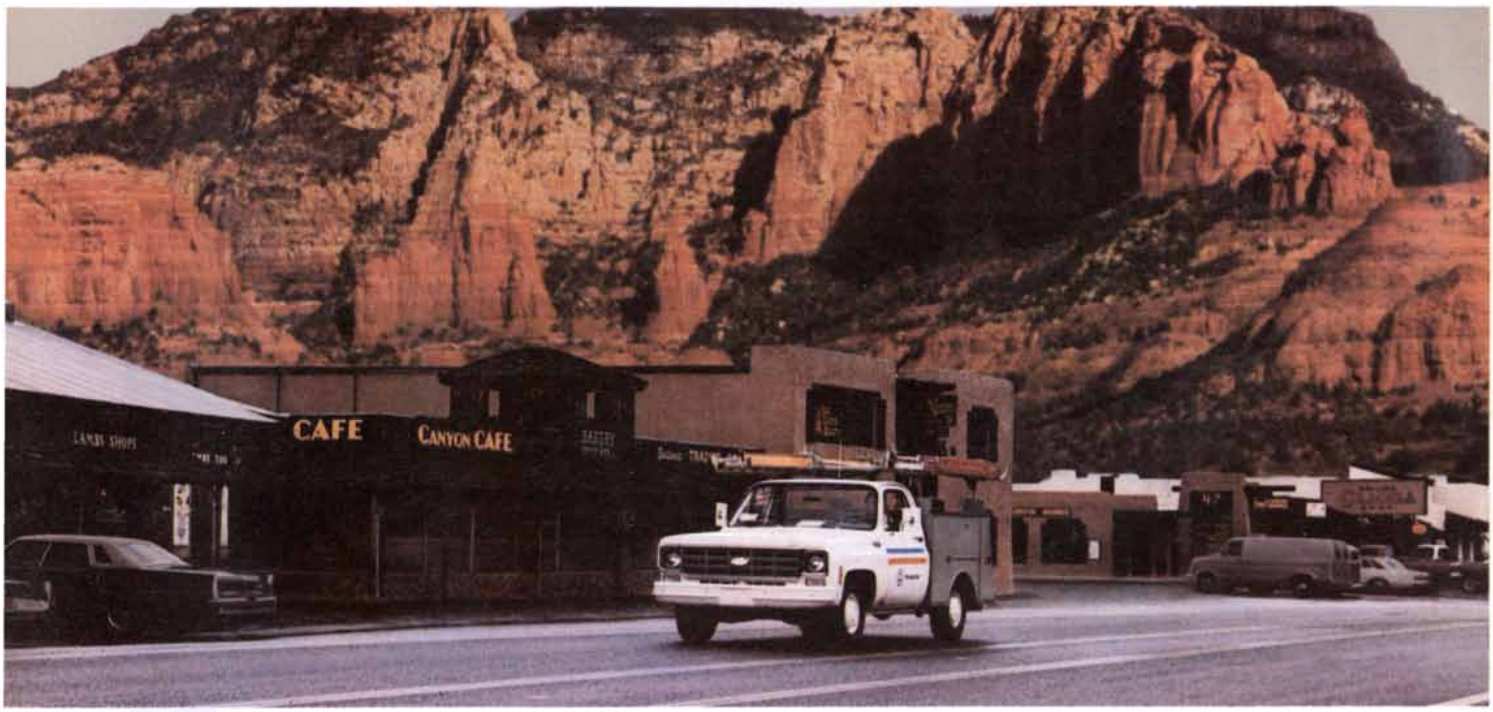
Over the years, he moved up in his company; as his assets grew and his needs changed, his Merrill Lynch Account Executive kept pace—with equities, commodities, real estate partnerships. Now he's buying tax shelters in $10,000 lots.

How can Merrill Lynch's unmatched resources benefit you? Ask us. Find out what it means to deal with a breed apart.

Merrill Lynch
A breed apart.

NEW YORK CITY AND SEDONA, ARIZONA HAVE ONE IMPORTANT THING IN COMMON.

THE LATEST BELL SYSTEM TECHNOLOGY.



One example of Bell System technology working for you is electronic switching.

Bell designed an electronic switching system especially for small towns. So the people in Sedona get the same sophisticated phone service as people in big cities like New York.

Roy Ronan is the Bell System man in charge of Sedona's electronic switching system.

Electronic switching handles your calls more efficiently than other systems. And it makes your telephone a more versatile, useful instrument. For example, if you're already talking, the phone beeps at you to let you know there's another call waiting.

Or you can set your home phone so that your incoming calls "follow" you around town.

If you want to talk to two people at once, three-way calling lets you do just that. And, to reach the people you call frequently, you can program your phone so that you just push two numbers to make the call.

Electronic switching. Bell Labs designed it. Western Electric custom-builds it.

For big cities and small towns alike.

With electronic switching, your calls can follow you wherever you go.

Bell System

Keeping your communications system the best in the world.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1929: "Many have sincerely been misled into the belief that there is a broad cleavage between scientists who accept evolution and those who do not. To them our readers may find it advantageous to show the following statement quoted in part: 'The Council of the American Association for the Advancement of Science has affirmed that so far as the scientific evidences of the evolution and plants and animals and man are concerned there is no ground whatever for the assertion that these evidences constitute a "mere guess." No scientific generalization is more strongly supported by thoroughly tested evidences than is that of organic evolution. The Council of the Association is convinced that any legislation attempting to limit the teaching of any scientific doctrine so well established and so widely accepted by specialists as is the doctrine of evolution would be a profound mistake, which could not fail to injure and retard the advancement of knowledge and of human welfare by denying the freedom of teaching and inquiry that is essential to all progress.'"

"The radial air-cooled aircraft engine possesses many advantages over the water-cooled engine in that it is lighter and avoids the complication of a radiator and piping. It is not so advantageous from the point of view of vision, and it introduces quite a large amount of air resistance, or drag. The National Advisory Committee for Aeronautics has recently conducted exhaustive research into the problem, measuring the drag in the wind tunnel with various forms of cowling. With a cowling developed by the N.A.C.A. 60 per cent of the drag due to the engine disappears. The great advantage of this type of cowling, which covers the entire engine, is that it separates the cooling air from the general flow about the body. With this cowling a cabin airplane equipped with a 200-horsepower engine will have its speed increased from 125 miles per hour to about 133 miles per hour."

"Professor Raymond A. Dart in 1925 startled the world with his announcement of the discovery of the skull of a small man-like ape that he called *Australopithecus.* The skull was found fossilized in a limestone cave at Taungs, near the western border of the Transvaal. Dart, with the intuition of genius, boldly made *Australopithecus* the type of a new family intermediate between the higher apes and man. It was perhaps a little daring on the evidence, and most of his critics have considered that in this he was wrong. William J. Sollas of Oxford University has shown that the skull differs greatly from that of the chimpanzee and concludes that '*Australopithecus* makes a nearer approach to the Hominidae than any existing anthropoid ape.' Robert Broom, late professor of geology and zoology at Victoria College in Stellenbosch, South Africa, states that '*Australopithecus* is a small human-like ape that no one need be ashamed of having as a forefather.'"

SCIENTIFIC AMERICAN

AUGUST, 1879: "The mouth of the Mississippi is now permanently open, and the currents of the river are so controlled that the mighty stream will henceforth be the chief factor in keeping its channel clear of the barriers it naturally tended to build up against the commerce of the world. When the Mississippi valley harbors, as it soon will, a more numerous population than the whole country can boast of now; when its farms and factories are doing, as they soon will, half of the productive work of the world—then it will be possible to form some adequate idea of the industrial and commercial benefit to flow from the unbarring of the outlet of what cannot fail to be the great artery of our national and international trade. It is a grand victory, and Captain James B. Eads, its engineer, may be sure that popular appreciation of its grandeur will grow with the growth of the commerce it makes possible."

"At this time public attention is forcibly drawn to the plague that prevails at Memphis and Havana, and threatens every commercial city of the country. Dr. Alfred Stillé traces the origin of yellow fever to the West Indies. There it was first discovered, and from West Indian ports it has in all instances spread. A high temperature is essential to its propagation; salt water and unsanitary conditions favor it, but the morbid poison must be imported in ships and in porous substances capable of absorbing and retaining the contagious effluvia. A strict quarantine is always efficient in preventing the dissemination of the disease. Its essential cause has never been isolated or defined, but it is assumed to be a specific poison, distinct from all other fever poisons. In the system it acts primarily in two ways, by disintegrating the blood and by inflaming the stomach; secondarily, it tends to impair the eliminating function of the kidneys."

"It is now 77 years since Brahmah proposed to transmit power to great distances by means of water under pressure, and it is half a century since Mann proposed and described the means for transmitting power over still greater distances by means of compressed air, but it cannot be said that either of these means of transmission has been employed to the extent their proposers expected they would be. It may be that electricity will come to the aid of the transmission of power. It seems that at a distance of about a mile the work realizable by electro-dynamic machinery is from 40 to 50 per cent of that done upon the dynamo-electric machine, which would presumably be situated at a waterfall. The electrical apparatus offers the great advantage of facility of transference, and the cable from the falls once laid or carried upon posts might remain in good order and without need of attention for years. There are cases where the electrical transmission of power might be found to be economical even when the electricity had to be generated by steam power."

"It was only last July that the first telephone was put up in Chicago. The American District Telegraph Company now has over 1,500 patrons—that is to say, it has put up that number of telephones with the corresponding number of wires. The carbon telephone constructed by Edison is used. The company has a central office and nine branch offices, which take in nearly every portion of the city."

"Several months ago a new spectroscope, containing compound sulphide of carbon prisms, was constructed by M. Thollon, wherewith he effected a very much greater dispersion of light than had been attained previously. With its aid he has produced a remarkable map of the solar spectrum. The work was done in Italy, as the Italian climate offers great advantages in this respect over that of Paris. Prince Nicholas of Oldenburg, who has taken a lively interest in the subject, provided M. Thollon with a small observatory at San Remo for his operations. The map is no less than 10 meters in length and is composed of about 4,000 lines. The well-known map of Angström contained 1,600 lines in a length of three meters."

"The well-known German professors Gustav Fechner and Wilhelm Weber have added their testimony to that of Johann Zöllner in favor of the 'spiritualistic' mediumship of Slade. In remarking upon Zöllner's curious method of accounting for the phenomena by supposing an unknown fourth dimension of space, Professor Gaston Tissandier thinks it strange that men who occupy some of the most important chairs of the German universities should be so easily deceived."

THE AUTHORS

RAYMOND DEVORET ("Bacterial Tests for Potential Carcinogens") is on the research staff of the Enzymology Laboratory at the Centre National de la Recherche Scientifique (C.N.R.S.) at Gif-sur-Yvette in France. He obtained his medical degree at the University of Paris Medical School and then worked part-time as a physician specializing in occupational radiation hazards. At the same time he took a degree in physics from the University of Paris and studied the effects of alpha rays on the bacterium *Escherichia coli* at the Radium Institute. In 1961 he was appointed to a full-time position in the C.N.R.S. Since then Devoret's research has been devoted to elucidating the mechanism of induction of dormant viruses in bacteria and the mechanism of chemical carcinogenesis.

HAROLD P. FURTH ("Progress toward a Tokamak Fusion Reactor") is professor of astrophysical sciences at Princeton University and associate director of the Princeton Plasma Physics Laboratory. He did graduate work in solid-state and high-energy physics at Harvard University, receiving his Ph.D. in 1960. He then joined the controlled-fusion research effort at the Lawrence Livermore Laboratory, becoming head of a group studying the toroidal magnetic-confinement approach to a fusion reactor. In 1967 he moved to Princeton, where he has initiated a number of major toroidal confinement experiments and done extensive research on the theory of plasma instabilities. His best-known publication is a poem in *The New Yorker* that describes an encounter between matter and antimatter, exemplified by the meeting of Dr. Edward Teller and Dr. Edward Anti-Teller. Furth would like to acknowledge the assistance of Robert Goldston in the preparation of his article.

ANDREW M. T. MOORE ("A Pre-Neolithic Farmers' Village on the Euphrates") is Wainwright Fellow in Near Eastern Archaeology at the University of Oxford. He studied modern history at Oxford as an undergraduate and then was trained in archaeology at the Institute of Archaeology of the University of London. In 1969 he was given a scholarship at the British School of Archaeology in Jerusalem, and for the next two years he lived in the Near East studying museum collections and visiting ancient sites. In 1973 he was awarded a Wainwright fellowship at Oxford; he obtained his Ph.D. last year. Moore has done extensive field work in the Near East: he took part in excavations at Jerusalem directed by the late Dame Kathleen Kenyon in 1966 and those at Knossos in Crete directed by J. D. Evans in 1969. From 1972 to 1973 Moore was in charge of the major excavation at Tell Abu Hureyra in Syria that is described in his article. This fall Moore will be a visiting professor in Old World prehistory at the University of Arizona.

THOMAS F. GOREAU, NORA I. GOREAU and THOMAS J. GOREAU ("Corals and Coral Reefs") have done pioneering research in coral-reef biology and ecology in Jamaica. Thomas F. Goreau died in 1970. The son of the photographers Fritz and Grete Goro, he was born in Germany, which his parents left in the early 1930's for France and soon thereafter the U.S. After his graduation from Clark University he did graduate work in ecology at Yale University under G. Evelyn Hutchinson. Goreau first became interested in coral-reef ecology while serving in the 1947 Bikini Scientific Resurvey, although the first opportunity to pursue this line of research did not arise until 1951, when he moved to Jamaica to join the faculty of physiology in the medical school of the University of the West Indies. On receiving his Ph.D. in 1956 he instituted, under the auspices of the New York Zoological Society, a long-term research project on Jamaican coral reefs that continued until 1967. In March, 1970, shortly before his death, he opened a new marine laboratory at Discovery Bay on the northern coast of Jamaica. Goreau's widow, Nora Goreau, is currently senior research fellow in marine biology at the University of the West Indies. Born in Panama, she was educated at the University of Panama Law School, the University of Iowa and DePaul University. She then did graduate work in neurophysiology at the University of Chicago. Thomas J. Goreau, their eldest son, is a Ph.D. candidate in geology at Harvard University. He did his undergraduate work in physics and astronomy at the Massachusetts Institute of Technology and obtained his master's degree at the California Institute of Technology in 1972. He then studied biology and oceanography at Yale, the Woods Hole Oceanographic Institution and Harvard.

WILLIAM HERBST and GEORGE E. ASSOUSA ("Supernovas and Star Formation") have collaborated on studies of star formation with optical and radio telescopes. Herbst is assistant professor of astronomy at Wesleyan University. He did his undergraduate work in astrophysics at Princeton University and received his Ph.D. in astronomy from the University of Toronto in 1974. After a postdoctoral fellowship at York University he was a fellow for two years of the Carnegie Institution of Washington. Herbst joined the Wesleyan faculty in 1978. Assousa is research professor of astrophysics in the Department of Terrestrial Magnetism of the Carnegie Institution. A Palestinian Arab, he was born in Jerusalem in 1936 and attended the American Friends School at Ramallah in Jordan. He then came to the U.S. to finish his education, obtaining his bachelor's degree at Earlham College, his master's degree at Columbia University and his Ph.D. in experimental nuclear physics from Florida State University in 1968. That year he was appointed a Carnegie fellow; two years later he joined the staff of the Carnegie Institution. Assousa directs the Foundation for Arab-Israeli Reconciliation (FAIR), which seeks to improve communication and cooperation between Arab and Israeli professionals and scholars.

MARTIN E. HELLMAN ("The Mathematics of Public-Key Cryptography") is associate professor of electrical engineering at Stanford University. He received his B.E. at New York University and his Ph.D. from Stanford in 1969. After teaching for two years at the Massachusetts Institute of Technology he joined the Stanford faculty. He is best known for his invention of public-key cryptography in collaboration with his students Whitfield Diffie and Ralph Merkle; he has also done research on information theory, error-control coding and statistics. Hellman wishes to acknowledge the National Science Foundation's support of his work.

KENNETH G. WILSON ("Problems in Physics with Many Scales of Length") is James A. Weeks professor of physical science at Cornell University. He obtained his bachelor's degree at Harvard University and his Ph.D. from the California Institute of Technology in 1961. The renormalization-group theory, his principal contribution to physics, resulted from his attempt to understand quantum field theory. He is currently applying improved computer technology to extend the capabilities of the renormalization-group approach. His pastimes include international folk dancing and detective stories. Wilson also has to his credit a 4 : 17 mile.

DAVID CREWS ("The Hormonal Control of Behavior in a Lizard") is associate professor of biology and psychology at Harvard University and an associate of the Museum of Comparative Zoology. He did his undergraduate work at the University of Maryland and received his Ph.D. in animal behavior from Rutgers University. He then did postdoctoral research at the University of California at Berkeley and at Harvard, where he was appointed to the faculty in 1976. Crews's interest in reptiles and their behavior goes back to his childhood in Florida.

How to project a beautiful image.

To make the most of your slides, nothing does it like a Kodak Carousel projector. For bright, crisp, colorful images on the screen. With eleven models to choose from, you can get just what you want in the way of convenience. And in the way of price.

All Kodak Carousel projectors use dependable gravity to drop each slide in place. All feature a superb, *f*/2.8, curved-field lens for bright, sharp viewing. All are whisper-quiet.

Attention, international travelers. Kodak now offers an automatic multi-voltage Carousel projector for use in most parts of the world.

Kodak Carousel projectors

The dependable ones.

Kodak

MATHEMATICAL GAMES

The imaginableness of the imaginary numbers

by Martin Gardner

"The imaginary numbers are a wonderful flight of God's Spirit; they are almost an amphibian between being and not being."

—LEIBNIZ

In a column on negative numbers in June, 1977, I described how long it took and how painful it was for mathematicians to extend the definition of "number" to include negative numbers. The same process was repeated with even more anguish when mathematicians discovered the enormous usefulness of what unfortunately were named imaginary numbers. It is a strange and beautiful story.

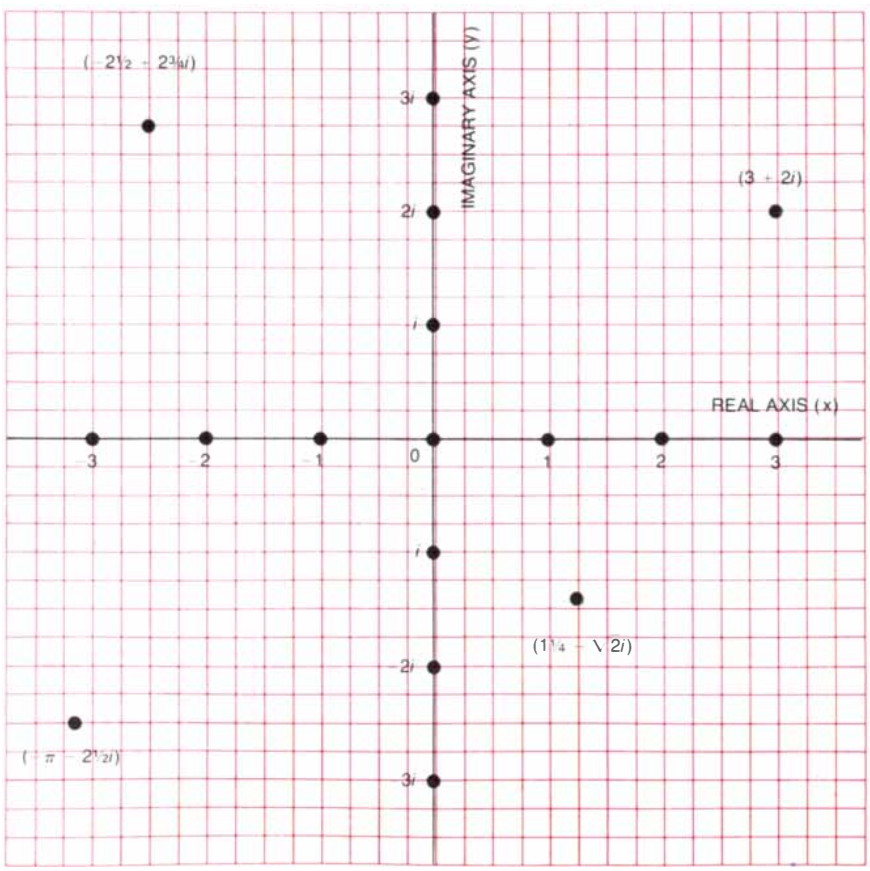
Although there were a few early pronouncements that negative quantities cannot have square roots (because the square of any real number must be positive), the story of imaginary numbers really begins in 16th-century Europe. At that time mathematicians, in particular Rafaello Bombelli of Italy, found that in solving algebraic problems it was often useful to assume that negative numbers did have square roots. In other words, just as the equation $x + 1 = 0$ could be solved only by setting x equal to -1, so $x^2 + 1 = 0$ could be solved only by setting x equal to $\sqrt{-1}$.

The seemingly preposterous assumption that there was a square root of -1 was justified on pragmatic grounds: it simplified certain calculations and so could be used as long as "real" values were obtained at the end. The parallel with the rules for using negative numbers is striking. If you are trying to determine how many cows there are in a field (that is, if you are working in the domain of positive integers), you may find negative numbers useful in the calculation, but of course the final answer must be in terms of positive numbers because there is no such thing as a negative cow.

Throughout the 17th and 18th centuries mathematicians everywhere kept discovering new uses for the square roots of negative numbers. It was Leonhard Euler who in the 18th century introduced the symbol i (the first letter of the Latin word *imaginarius*) for $\sqrt{-1}$. A much-quoted statement attributed to Euler asserts that such roots are not nothing or more than nothing or less than nothing but strictly imaginary or impossible. Mathematicians eventually worked out the algebraic rules for manipulating the "pure imaginaries" (the products of i and real numbers) and what later came to be called complex numbers (the sums of pure imaginaries and real numbers).

A complex number has the form $a + bi$, where a and b can be any real numbers. (In this instance the plus sign is not meant to indicate addition in the familiar sense; it serves mainly to separate the real part a of the complex number from its imaginary part bi.) In other words, if a equals 0 and b does not equal 0, the complex number is a pure imaginary bi. If b equals 0, then bi drops out, leaving the real number a. Therefore the complex numbers include as subsets all the reals and all the pure imaginaries, just as the real numbers include all the integers, fractions and irrationals. In modern terminology the complex numbers form the mathematical structure called a field, whose elements obey all the familiar laws of arithmetic. The complex-number field is also closed with respect to addition, subtraction, multiplication and division, that is, applying those operations to any two complex numbers will always generate another number in the field. There is a sense in which the discovery of the complex field completes traditional algebra because it makes possible the solution of any ordinary algebraic equation whatsoever. The field also turned out to be closed with respect to the operations employed in the calculus, and that discovery gave rise to a vast edifice of mathematics concerning the functions of a complex variable.

Many of the advances of modern



Correspondence of complex numbers with points on the complex plane

Many companies don't call AMP until they need electrical connectors or switches. But why wait till then? If you involve us early enough, for instance, in the design stages, we can help save you money. A lot of money.

Here's a recent example:

The design engineers at a business machine company were upgrading the speed and performance of their equipment. As usual, this meant packing more into less space.

They had been using non-insulated terminals for many of their connections. But they realized that with the new density requirements, a secondary insulation process was necessary to avoid short circuits that could cause the equipment to malfunction.

The cost of this additional process was high, until AMP introduced Ultra-Fast. A connector with pre-fitted insulation and designed for one-step application.

With Ultra-Fast they were able to implement the new designs with confidence, and improve both the performance and reliability of their equipment at the same time.

All at a savings over what they were paying for the old non-insulated terminals.

Can we produce the same kind of results for you?

If your company manufactures products that require connectors, chances are the answer is yes. But it's important that you call us in early.

That's when our experience in developing interconnection products and application equipment for hundreds of industries worldwide can help you the most.

Early involvement.

At AMP, we call this approach to solving our customers' problems "Early Involvement." And it's what makes a 4¢ AMP connector worth so much more than 4¢.

Ask for a copy of our brochure, "AMP Has a Better Way." Call (717) 564-0100, Ext. 8420.

Or write to AMP Incorporated, Harrisburg, PA 17105.

AMP has a better way.

AMP

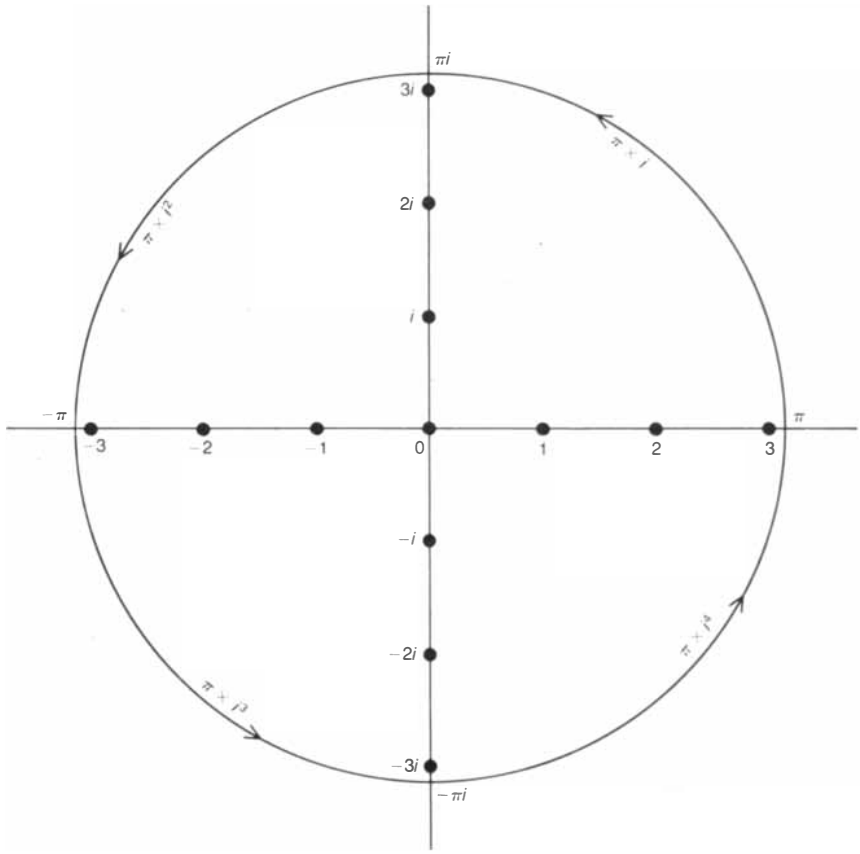
Barcelona • Brussels • Buenos Aires • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • London • Luzern • Mexico City • Montreal
Paris • San Juan • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo

physics could not have been made without the extension of algebra to the complex field. The first major scientific use of complex numbers was made by Charles Proteus Steinmetz, who found them essential for doing efficient calculations on alternating currents. Today no electrical engineer could get along without them, and neither could any physicist working on the area of air or fluid dynamics. The numbers also play a basic role in relativity theory (where space-time is made symmetrical by the stratagem of regarding the three spatial dimensions as real and the time dimension as imaginary), in quantum mechanics and in many other branches of modern physics.

Because there are still misgivings about calling i a number it is not uncommon even today for a physicist, a philosopher or even a mathematician to maintain that i is not really a number but is only a symbol for an operation I shall explain below. No one has disposed of this verbal quibble more effectively than Alfred North Whitehead. In the chapter on imaginary numbers in his *Introduction to Mathematics* he wrote:

"At this point it may be useful to observe that a certain type of intellect is always worrying itself and others by discussion as to the applicability of technical terms. Are the incommensurable numbers properly called numbers? Are the positive and negative numbers really numbers? Are the imaginary numbers imaginary, and are they numbers?—are types of such futile questions. Now, it cannot be too clearly understood that, in science, technical terms are names arbitrarily assigned, like Christian names to children. There can be no question of the names being right or wrong. They may be judicious or injudicious; for they can sometimes be so arranged as to be easy to remember, or so as to suggest relevant and important ideas. But the essential principle involved was quite clearly enunciated in Wonderland to Alice by Humpty Dumpty, when he told her, apropos of his use of words, 'I pay them extra and make them mean what I like.' So we will not bother as to whether imaginary numbers are imaginary, or as to whether they are numbers, but will take the phrase as the arbitrary name of a certain mathematical idea, which we will now endeavour to make plain."

Complex numbers behave so much like ordinary numbers when they are added, subtracted, multiplied and divided (according to the rules of the complex field) that most mathematicians no longer hesitate to call them numbers and regard them as having just as much "reality" as negative numbers. Even the counting numbers are no more than symbols manipulated according to the rules of a deductive system. We think of them as being more "real" than other numbers only because their applications are so close to our practical experience of counting fingers, cows, people and so on. What we forget is that only the fingers, cows and people are real, not the symbols to which we turn to count them. In the realm of pure mathematics i is just as real as 2. If we like, we can think of 2 as nothing more than an operator: a symbol that tells us to double 1.



How pi is multiplied by i, i^2, i^3 and i^4

Most people are so accustomed to working with real numbers, however, that they feel great relief when they discover there is a simple geometrical interpretation of complex numbers. This interpretation, which makes it easy to "see" what the numbers are all about, identifies every complex number with a point on the Cartesian plane. The first person to make this ingenious connection was Caspar Wessel, a self-taught Norwegian surveyor who lectured on it in 1797. A few years later the idea was rediscovered by Jean-Robert Argand, a Swiss bookkeeper (who published a small book about it in 1806), and independently by the great German mathematician Carl Friedrich Gauss.

As is shown in the illustration on page 18, the basic idea is to view the horizontal axis of the Cartesian plane as the real-number line and the vertical axis as the line of points that correspond to the pure imaginary numbers. In other words, one-to-one correspondences are established between the real numbers and the points on the x axis and between the pure imaginary numbers and the points on the y axis. As I have pointed out, both of these sets can be considered subsets of complex numbers, and now the remaining complex numbers can be put in one-to-one correspondence with the remaining points on the plane. To obtain the coordinates of the point associated with a complex number one simply measures the real part on the real axis and the imaginary part on the imaginary axis. The points corresponding to four complex numbers are shown in the illustration.

With this interpretation of the complex numbers it is possible to forget entirely the disturbing notion that i is the square root of -1 (which of course it is not in the usual sense of taking a square root). Now a complex number can be viewed simply as an ordered pair of real numbers: the first number measured on the real axis and the second on the imaginary axis. In other words, by properly defining the arithmetic operations for combining these pairs, it is possible to construct an algebra of ordered pairs of real numbers that is equivalent to the algebra of complex numbers. That opaque phrase "the square root of a negative number" is nowhere encountered

in this new algebra, although the same idea is of course present in a different language and in different notation. If this algebra of ordered pairs had been developed before complex numbers, today perhaps no one would remember imaginary numbers and wonder whether or not they exist.

After the discovery of this geometrical interpretation of complex numbers mathematicians immediately asked whether the basic concept could be generalized to three dimensions, that is, to points in space. The answer, alas, is no—not without a radical modification of the laws of arithmetic. It was the Irish mathematician William Rowan Hamilton who made the first breakthrough into "hypercomplex numbers" when he invented quaternions: four-part numbers that combine a real number with three imaginaries. The key to manipulating them is the fact that they do not obey the commutative law for multiplication: the rule stating that for any two numbers a and b, ab equals ba.

The idea of dropping this law came to Hamilton in 1843 as he was strolling with his wife at dusk along the Royal Canal in Dublin. He was so elated that he stopped to scratch the basic formula on a stone of Brougham (pronounced "broom") Bridge. The original graffito weathered away in Hamilton's time, and now the bridge is known only as the one that crosses Broombridge Street. There is a tablet in the stone commemorating the great event, however, and in 1943, a century after Hamilton's revelation, Ireland honored it with a postage stamp. Quaternions do not form a field (their structure is called a division ring), but the algebra of quaternions is equivalent to an algebra of ordered quadruplets and is often applied today as a part of three-dimensional vector theory. The discovery of the algebra of quaternions marked the beginning of modern abstract algebra, in which all kinds of "numbers" much stranger than the complex numbers can be defined.

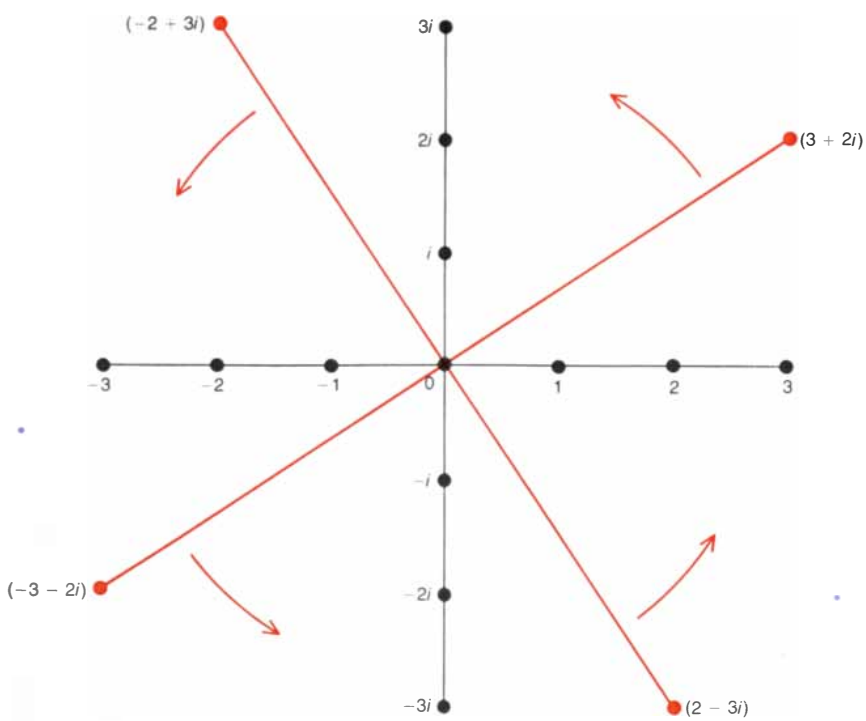
Because of the correspondence between complex numbers and points on the Cartesian plane, when the plane is used in this way, it is called the complex plane. (It is also called the z plane for the unspecified complex number z equals $a + bi$ and sometimes an Argand diagram because for many decades no one knew about Wessel's earlier discovery.) I shall not go into the details here of how complex numbers can be added, subtracted, multiplied and divided by geometrical diagrams on the complex plane. Readers who do not already know the rules governing these operations can find them in any elementary algebra textbook that covers complex numbers. A brief explanation of multiplication by i, however, is necessary to introduce an elegant theorem about the roots of numbers.

To multiply a number on the complex plane by i one takes the radius-vector line to the point corresponding to the number (the line from the origin of the plane to the point) and rotates it 90 degrees counterclockwise; the new end point of the vector corresponds to the product of the number and i. It is in such a sense that i can be viewed as an operator. To understand this idea consider what happens when i is raised to various powers: i raised to the first power of course equals i, and it is easy to see that i^2 equals -1, i^3 equals $-i$ and i^4 equals 1. This four-step cycle repeats endlessly: i^5 equals i, i^6 equals -1, i^7 equals $-i$, i^8 equals 1 and so on. All even powers of i equal 1 or -1 and all odd powers equal i or $-i$.

The illustration on the preceding page shows how these observations apply to the multiplication of a number (in this case pi, or π) by i. The point corresponding to π on the positive side of the x axis is located and moved 90 degrees counterclockwise around the circle of radius π centered at the origin of the plane. An arrow shows how the end point of this operation is the pure imaginary, πi, which lies on the upper part of the y axis. Multiplying π by i^2, then, is equivalent to multiplying it twice by i: the point corresponding to π is moved 180 degrees around the circle and ends up at the point $-\pi$ on the x axis, or real-number line. Similarly, multiplying π by i^3 effects a turn of 270 degrees, ending at the point $-\pi i$ on the lower part of the y axis; multiplying π by i^4 is the same as multiplying π by 1, and so we are back to π. We can continue in the same way with all the higher powers of i. Each next-highest power takes us another quarter turn counterclockwise around the circle.

The inverse operation of multiplication by i is division by i: moving clockwise 90 degrees around the center origin of the plane. In other words, for any complex number draw a radius-vector line from the origin to the point that represents the number. Then to multiply the number by i rotate the vector 90 degrees counterclockwise [*see illustration at left*], and to divide it by i rotate the vector 90 degrees the other way. (As a joke a friend of mine once suggested that i times infinity equals 8 because multiplying by i turns the infinity sign upright.)

With this interpretation of multiplication it turns out that if complex roots are counted, every nonzero number (real or complex) has exactly n nth roots. In other words, every number has two square roots, three cube roots, four fourth roots, five fifth roots and so on. It follows that every cubic equation has three solutions, every fourth-degree equation has four solutions and so on, and when we diagram the roots of individual numbers on the complex plane, an unexpected and delightful property is revealed. The n points corresponding to the nth roots all lie, separated by equal dis-



Multiplying the complex number $3 + 2i$ *by* i, i^2, i^3 *and* i^4

POLICE RADAR MEETS ITS MATCH.

A good driving record makes for lower insurance rates and prevents possible loss of your driver's license. GUL can help protect yours. It may even pay for itself within the first year of use, depending upon your driving habits.

With GUL's Micro Scan, you can own a high-technology radar detector for only $79.

Your radar detector should do two things well: give you as much advance warning of police radar as possible, and screen out false signals. And few competitive units are able to perform both these functions well. However, the GUL G-73, developed by B.E.L. Laboratories represents a breakthrough in both sensitivity and false signal rejection in a compact, well-designed case.

Road Superiority.

The 10-year researched GUL has been tested numerous times against such units as the Fuzzbuster XK, the Whistler XK and the XK Snooper. In separate tests conducted by the Canadian Tire Corporation and Motor Consumer Report, the GUL demonstrated a detection distance significantly greater on both X and K bands, than all three competitors. Its nondirectional elliptical horn antenna receives these signals from front, back, and sides, and is even sensitive enough to detect the new short pulse KR-11 radar. Signals of less than .01 millionths of a volt (ranging from 9.445-24.445 GHz) are sufficient to trigger the alert signal and provide valuable warnings when approaching hills and curves. An adjustable gain control knob lets you optimize GUL's sensitivity for both city and country conditions.

Doesn't Cry "Wolf."

The GUL takes full advantage of modern integrated circuit technology to provide a capability of more than 100 transistors. Many competitive units try to get by with as few as 20 transistors and they just don't work as well. GUL's sophisticated circuitry is completely shielded by a diecast and extruded aluminum case. This not only provides a better wearing case than the often used plastic ones, but also protects the input terminals from stray signals, the most common cause of false alarms. In addition, the GUL uses an internal voltage regulator, commutating filter, and synchronous detector to reduce power consumption to a minimum and provide a more stable gain setting—features usually found only in units costing twice as much.

This is illegal in Michigan and Virginia.

Two Warnings You Can't Miss.

The GUL uses a double warning system. Both a tone and light come on in the presence of radar. And both remain on until radar is gone. You're never lulled into a false sense of security by a tone that sounds once and then goes off. And you can control the tone with a front-mounted on/off switch.

Quick Installation.

The GUL detector is one of the most compact units available, measuring only 4"x3"x4½", and weighing just 20 oz. Its aluminum case is finished in glare-free matte black and comes with its bracket in place. To mount, simply take the adhesive Velcro pad, peel off the back, and adhere it to the top of your dashboard. The bottom of the bracket has a matching Velcro foot which anchors your unit securely on top of the dash. It plugs into your cigarette lighter. That's all there is to it. The same Velcro mounting also allows for easy removal in seconds to prevent theft. The small size GUL will fit into any glove compartment.

Try It Yourself.

Order the new GUL detector today. Better yet, road test it yourself. See how comforting it is to observe police radar when it's observing you. If you don't find it to be the best electronic counter-intelligence you can have for $79, simply mail it back. Your GUL comes with The Sharper Image two-week return privilege and a manufacturer's 120-day warranty.

ORDER NOW TOLL-FREE

Call the toll-free ordering number below and charge it to your credit card, or send check for $79 plus $2.50 delivery (add $4.74 tax in CA).

(800) 227-3436

In California (800) 622-0733

THE SHARPER IMAGE™

260 California St., Dept. GL—071
San Francisco, CA 94111
(415) 788-4747

tances, on a circle whose origin is at the center of the plane. In other words, the points mark the corners of a regular n-sided polygon. For example, the illustration below shows the locations of the six sixth roots of 729. If, as in this instance, the number is real and has an even number of roots, two corners of the polygon will lie on the real axis. If the number is real and has an odd number of roots, only one corner of the polygon will lie on the real axis.

Besides being essential in modern physics, the complex-number field provides pure mathematics with a multitude of brain-boggling theorems. It is worth keeping in mind that complex numbers, although they include the reals as a subset, differ from real numbers in startling ways. One cannot, for example, speak of a complex number as being either positive or negative: those properties apply only to the reals and the pure imaginaries. It is equally meaningless to say that one complex number is larger or smaller than another.

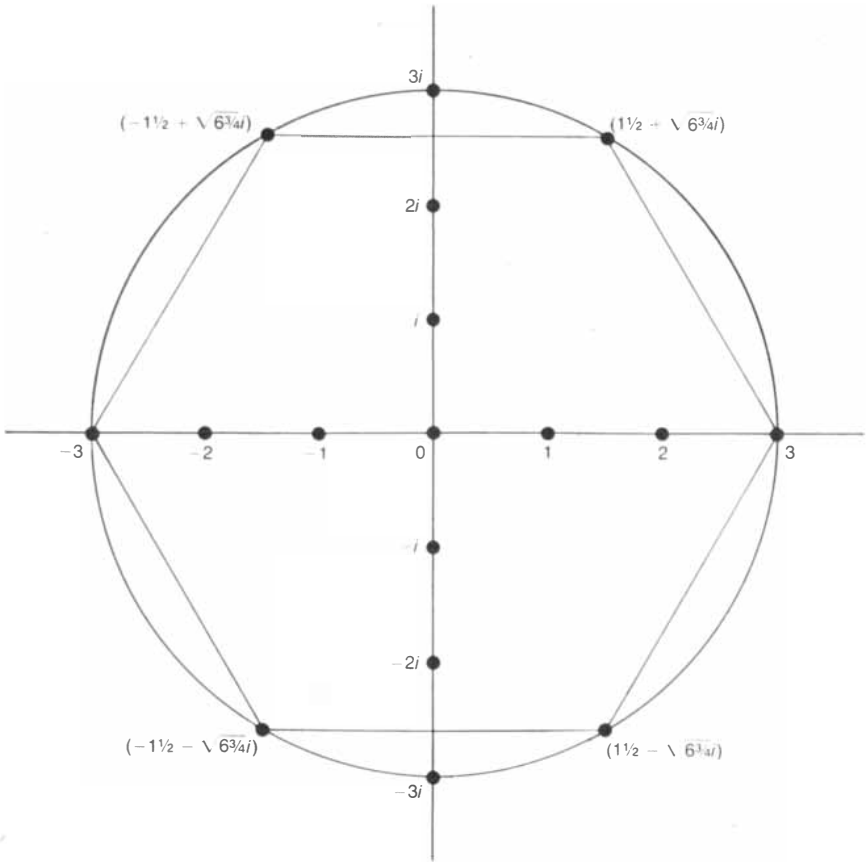
It had been known before Euler that the product of any two pure imaginaries is a real number, but it was Euler who first showed that i^i is also real. It is equal to $e^{-\pi/2}$, an irrational number with the decimal expansion of .2078795763.... Actually this number is only one of an infinity of values, all real, of i^i. They are given by the formula $e^{-\pi/2 \pm 2k\pi}$, where k is any integer, so that when k equals 0, the principal value given above is obtained. The ith root of i is also a real number, the principal value of which is $e^{\pi/2}$, or 4.8104773809....

There are many other formulas in which i is related to the two best-known transcendental irrationals, e (the natural basis of logarithms) and π. The most famous formula, developed by Euler but based on an earlier discovery, is $e^{i\pi} + 1 = 0$, which Edward Kasner and James Newman call in their book *Mathematics and the Imagination* "elegant, concise and full of meaning." They also quote a remark by Benjamin Peirce, the Harvard mathematician who was the father of Charles Sanders Peirce, about the formula: "Gentlemen," he said, after writing the formula on a blackboard, "that is surely true, [but] it is absolutely paradoxical; we cannot understand it, and we don't know what it means, but we have proved it, and therefore we know it must be truth."

Well, the formula is not quite without meaning. Rewritten as $e^{i\pi} = -1$, it can be diagrammed on the z plane as the limit of the infinite sequence: $1 + \pi i + (\pi i)^2/2! + (\pi i)^3/3! + (\pi i)^4/4!....$ (The exclamation point is the factorial sign: $n!$ equals $1 \times 2 \times 3 \times \ldots \times n$.) The terms of this sequence are diagrammed as an infinite set of points on a counterclockwise spiral of straight lines that strangles the -1 point on the real axis.

George Gamow, seeking to dispel the mystery of complex numbers, once devised this puzzle. An old parchment, describing the location of buried pirate treasure on a desert island, gave the following instructions. On the island there are only two trees, *A* and *B*, and the remains of a gallows. Start at the gallows and count the steps required to walk in a straight line to tree *A*. At the tree turn 90 degrees to the left and then walk forward the same number of steps. At the point where you stop drive a spike into the ground. Now return to the gallows and walk in a straight line, counting your steps, to tree *B*. When you reach the tree, turn 90 degrees to the right and take the same number of steps forward, placing another spike at the point where you stop. Dig at the point exactly halfway between the spikes and you will find the treasure.

A young adventurer who found the parchment with these instructions chartered a ship and sailed to the island. He had no difficulty finding the two trees, but to his dismay the gallows was gone and time had abolished all traces of where it had stood. Not knowing the location of the gallows, he could see no way of finding the treasure and so returned empty-handed. Gamow points out that if the young man had been familiar with the technique of manipulating numbers on the complex plane, he could have found the treasure with ease. Readers who know the basic rules for diagramming complex numbers should be able to solve this problem. I shall give the solution next month.



The six sixth roots of 729

The first problem last month was to find the smallest positive integer that can be expressed as the sum of two squares in two different ways. The number is 50, which equals $5^2 + 5^2$ or $1^2 + 7^2$. If zero squares are allowed, however, the number is 25, which equals $5^2 + 0^2$ or $3^2 + 4^2$. If the two squares must be nonzero and different, the solution is 65, which equals $8^2 + 1^2$ or $7^2 + 4^2$.

The second problem was to determine whether or not *MU* is a theorem in the *M, I* and *U* formal system. A simple proof of why *MU* is not a theorem can be found on pages 260 and 261 of the book I reviewed last month: *Gödel, Escher, Bach: an Eternal Golden Braid*, by Douglas R. Hofstadter.

Many readers generalized the rotating-table problem given in February. Lyle Ramshaw, Richard A. Litherland and the team of Stephen Willard and Ted Lewis were the first to show that given a table with n wells and a player with k hands, the bell can be rung in a finite number of steps if and only if k is equal to or greater than $n(p-1)/p$, where p is the largest prime factor of n. Finding a nonrecursive formula for the minimum number of steps required remains unsolved.

How much energy can the nation afford to put off limits?

One third of all land in the United States, and most of the undersea continental shelf, is owned by the federal government.

Millions of acres of this land, including vast areas that have great potential for energy development, have been placed off limits by the government for environmental reasons.

It is ironic that, at a time when our country is dangerously dependent on foreign oil, there are efforts to put even more federal land off limits.

It is ironic—and it's unwise. The record shows that energy development and environmental protection are compatible.

For example, more fish are being caught in the Gulf of Mexico today than before the Gulf became the most drilled body of water in the world. And at the Aransas National Wildlife Refuge in Texas, where Conoco has been producing oil and natural gas for 35 years, there are more rare whooping cranes than when we started.

America's land can be put to a variety of compatible uses, including the development of energy the nation badly needs.

conoco

Doing more with energy.

Conoco, Stamford, CT 06904. © 1979

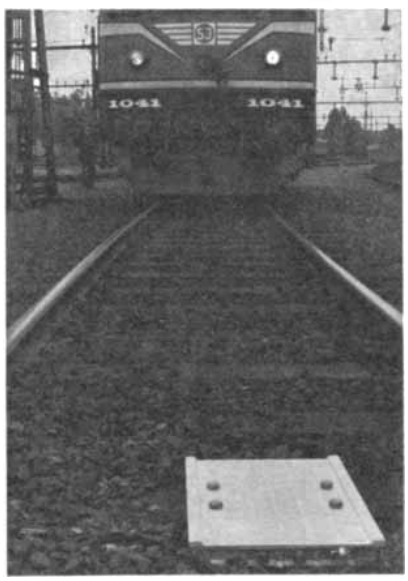
Today's Bunker Ramo is helping the Swedish railways make its curves as safe as its straightaways.

If all the signals, speed restriction boards and grade markers on the Swedish railways were suddenly removed, no one would miss them. Because a new on-board computer system is now capable of transmitting exactly the same information.

What's more, the system assures the observance of speed limits in critical situations. And that means that when the engineer hits a curve that calls for 40 kilometers an hour, there's no way he can do even 41.

Critical to the success of this system are some very special Amphenol® connectors designed by Bunker Ramo engineers in England and Scandinavia. And what makes them special is the fact that each connector plug assembly includes one of 15 different printed circuits which are coded to allow for the various speed controls.

Today, through our well-known Amphenol connectors, Bunker Ramo has a leading position in the aerospace and aircraft industries, a strong foothold in telecommunications, and a growing role in fiber optics. In addition, we are a major producer of the data terminals and minicomputers that comprise the information systems widely used by commercial banks, thrift institutions, insurance companies and brokerage houses.



Most important, we are a company with a reservoir of high technology, a solid base of innovative products, and a strong management team.

Which is why today's Bunker Ramo is a company that's on the move. For our latest financial reports, write Bunker Ramo Corporation, Dept. FR, 900 Commerce Drive, Oak Brook, Illinois 60521.

BUNKER RAMO

BOOKS

What made human populations start to grow, and modern referents of time and frequency

by Philip Morrison

THE MODERN RISE OF POPULATION, by Thomas McKeown. Academic Press ($16). The human population has increased with increasing rapidity since about 1750. A subset much better measured, the population of England and Wales, had risen threefold in the seven centuries from the time of the Domesday Book to 1700. It tripled again by 1851 and has very nearly tripled once more since then. The phenomenon is unique in its scale, duration and continuity. We need to know why, for the sake of our future, if for no other reason. The author, a British demographer of medical background, seeks the answer in this small book of tight argument, first published in 1976. His method is explicitly unhistorical. He proposes to study the problem in the light of present knowledge; what we need to know is not what the physicians of the past thought they were doing but how effective their treatments actually were. Gauging that does not call for the historian's leap into the concepts of the past.

National registers of birth and death supply the first firm data; no effort to repair that lack for earlier times by local records or clever indirect accounts can be relied on. Concordant statistics are provided by Sweden from 1749, France from 1800, England and Wales from 1838 and Ireland from 1871. Birthrates have been falling pretty steadily; allowing for net migration, the population has grown in spite of a lower birthrate. The direct cause of the rise in population is the decline in mortality. For this we have significant detail: the cause of death. The data are again decisive: most of the decline has been due to the reduction of the effects of infectious disease. The decline in deaths from infectious diseases amounted to three-fourths of the total decrease in mortality between 1850 and 1971 (with all rates standardized to the age and sex distribution of 1901). Leading the decline, indeed contributing more than half of it, were airborne infections, including tuberculosis, pneumonia, influenza, scarlet fever and smallpox.

Until about 1900 there was no decrease in infant mortality or in mortality among those over 45; the improvement came through the steady reduction at just those ages where mortality was naturally lowest: the years from two to 45. Tuberculosis was the largest single cause of death in Victorian times; its decline is steady since the first years of registration, a generation before the tubercle bacillus was even identified. Effective treatment began only after World War II, first with streptomycin and then with the BCG vaccine. The detailed consideration of the other important diseases broadly confirms the conclusion: neither treatment nor prophylaxis was the cause of lowered mortality.

Was it reduced exposure? For waterborne diseases the answer is plain; the safety of drinking water, milk and food and the control of sewage made a real difference. For the airborne diseases, however, the reduction of exposure was at most a secondary cause, unimportant before 1900. Even as late as 1950 all adolescents were shown by the tuberculin reaction to have been exposed to tuberculosis; today only a minority bear those immune signs. Was it perhaps a spontaneous reduction in the virulence of the disease? For scarlet fever and influenza such changes are fairly plain, although they are cyclical. Again the uniformity of the effect over a wide variety of disease agents that, like tuberculosis and measles, remain fully infectious today seems to exclude that cause or its counterpart: a selective change in the specific resistance of the human population.

What remains? Argument by exclusion is dangerous, but there are positive grounds for accepting the remaining factor: general health has improved, mainly because of the improvement of the nutritional state of the people. New crops, new seeds, new tools, better farming methods and better transport all resulted in the people's being better fed. Britain fed herself during a trebling of population without substantial imports of food. The improvement in the life expectancy of the aristocrats, documented along with the overall changes, does not contradict this conclusion. Their minority case can be explained by their reduced exposure, not offset by the bad conditions of the cities. The lesson of this analysis is easy to infer. Yet the Birmingham of 1850 is not the Bengal of

On Photographing a Small Step

It is unearthly. The landscape underfoot. It is soft and powdery.

A silver beast stands there. His head is a shiny globe. His huge frankfurter fingers hold a small box that talks.

Click. Whir. Click. Whir.

A man is taking the first pictures on the moon. He takes them with a camera that gives him startling detail. It is a Hasselblad.

It is the first of twelve taken to the moon on six landings. And it is still there today.

Photograph by Hasselblad courtesy of NASA

Hasselblads have taken tens of thousands of photographs during millions of miles of travel in space.

In all, 60 Hasselblads on 27 flights with 67 astronauts have completed 3566 orbits of earth and 363 of the moon.

Every Hasselblad proved reliable on every flight under every condition.

Today, 10 years after, Hasselblad celebrates mankind's giant leap with a commemorative issue of a limited edition (earth version) of the 500EL/M motor drive camera that took those historic photographs.

The only medium format camera on earth or in space that can equal it, is another Hasselblad.

For even in an age of space, much of a year goes into building each Hasselblad; much of it crafted by hand.

100 Hasselblads from one

No Hasselblad can be any better than any other Hasselblad; only more suited to your purposes.

Hasselblad is a prodigious system of 4 cameras, 22 lenses, 8 viewfinders, 9 film magazines, and over 300 other accessories. And there is no obsolescence.

Simply by choosing the right pieces, you can *custom build* over 100 distinct Hasselblad cameras for almost any photographic application.

From a man's first steps on the moon to a child's first steps on earth.

An incredible motor drive

The 500 EL/M is the only hand held medium format camera with an integrated motor unit.

So unlike cameras with optional motor drive attachments, it is built to withstand the punishing demands of motorized operation.

And the possibilities are infinite.

There is a dial set in the body that will order the camera to fire in any of five different ways.

From single shot to speeded up continuous firing at the rate of one shot every 8/10th of a second.

That means you can walk away from the camera, devoting all your concentration to your subject, while firing with a remote cable.

Plug in the intervalometer, for instance, and you can trigger the camera automatically. At any moment you want, from 2 seconds to 14 minutes.

Or trigger up to four cameras simultaneously with a command unit.

Or, fitted for remote radio control, fire from a half mile away.

And the standard film frame on a Hasselblad is 2¼ inches square, *almost 4 times the area* of a 35mm frame. So the ability to hold sharpness and contrast is phenomenal. (For actual size of film frame, see box at lower right.)

The eyes of a Hasselblad

A Hasselblad can see with any one of 22 lenses; all multi-coated (except for two special purpose lenses) in focal lengths from 30mm to 500mm with angular fields from 180 degrees to 9 degrees.

There are lenses for sports, aerial, architectural, and fashion photography. Portrait, landscape, medical, underwater and news photography. Wildlife, laboratory, industrial and child photography.

A lavish brochure on the 500 EL/M and other Hasselblad cameras is available free if you write: Braun North America, Dept.SA8H, 55 Cambridge Parkway, Cambridge, MA 02142—a division of The Gillette Company and exclusive marketer of Hasselblad in the U.S.

There is even the 105mm Zeiss UV-Sonnar f4.3 which can photograph the ultraviolet portion of the electromagnetic spectrum with costly quartz elements that can detect radiations unseeable by the human eye.

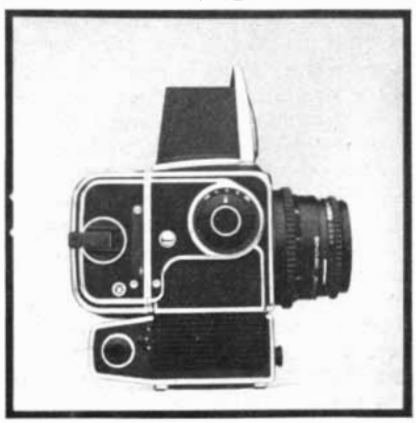
There are also lenses that adjust the aperture fully automatically as the light changes.

Even if you are not yet into these more exotic applications but simply recording travel and everyday happenings, it's nice to know that your camera produces the very best images in the medium format.

So, in effect, there is nothing you can't see with a Hasselblad. And photograph. From what is invisible to the naked eye to all that is visible in the universe.

Is it any wonder then that NASA came to Hasselblad for the camera which was to photograph space? And that top professional and serious amateur photographers still rely on Hasselblad.

HASSELBLAD®



The Motor-Drive Hasselblad 500 EL/M, made in Sweden. The standard lens is a multi-coated Zeiss Planar 80mm f/2.8T.*

1980; the lesson must not be read too narrowly.

McKeown frequently draws indirect support from animal-population studies and from what is known of birth and death among human hunting populations living today. In these matters he has not always found the best and most recent data, nor is anyone quite sure of the generalizations. At the worst, however, these questions lie in the background; his central points are strong, if not yet compelling because of the lack of quantitative knowledge. It is tempting to draw the conclusions over the span of the entire human epoch. Hunting peoples achieved a balance with their environment at a low number, by methods that are not quite clear. Agriculture induced a strong rise in the number of head of human beings and their livestock. The microbial parasites of the domesticated mammals found a new host in the inhabitants of the populous new farm villages; infectious diseases set limits on population once human numbers grew enough to make the food supply again marginal.

Then in western Europe at the beginning of modern times the chain was broken: a new agriculture brought new food supplies, disease receded, population grew apace. This time the growth was limited not by a rise in deaths but by a permanent drop in the birthrate brought about by the conditions of the modern nation state. Better water and better food, preventive medicine, contraception (which reduced the incidence of unreported infanticide) and finally better therapy all now take part in the balance. Infectious diseases are no longer the captains of the men of death; even our fears seem different now.

FREQUENCY AND TIME, by P. Kartaschoff. Academic Press ($24.50). The bodily transport of high-precision clocks is by no means the most expensive way to compare times: "besides its basic simplicity, the method has its special excitements, e.g., running for the nearest power line outlet because the batteries are running low... or trying to convince a customs officer that an atomic clock is not an A-bomb." The winds of experience, although rarely in such familiar domain, blow through this interesting volume by a research-and-development engineer of the Swiss Post Office, where they take time checks seriously. Comparison at the microsecond level, important for radio-navigation systems, is well served by flying the clock. The power lines serve for comparisons only to seconds, and the telephone networks generally provide accuracies good to a few milliseconds at the sound of the tone.

The radio links are varied. WWV in Fort Collins and its score of counterparts, from ZLFS in New Zealand to RKM in Irkutsk, provide for both hemispheres more or less elaborate coded transmissions, fixing frequency and clock time by sky-wave signals on so many frequencies that an experienced operator anywhere can "almost at any time of the day find a useful signal." In Europe the proliferating time signals even begin to interfere. A good receiver and a triggered oscilloscope enable the time to be read from a simple tick to half a millisecond. The chief source of error is the path delay, not hard to estimate for single-hop signals but troublesome once the signal is heard by multiple ionosphere bounces.

The stabler ground-wave propagation at low frequencies has found wide favor, with time comparisons under good conditions reaching a fraction of a millisecond with very simple receivers. With phase-lock reception results are repeatable to within a microsecond per day. The use of very-low-frequency radio—below 30 kilohertz—has been steadily growing. A high-power station can furnish such a signal worldwide, with phase variation from day to day of only a few microseconds "during sunlight on the propagation path." Eight stations exploit this band, including half a dozen that are used mainly for low-speed communication with U.S. Navy ships.

The navigational system called Loran-C provides intricate fast-pulsed signals from chains of stations (one master station and several slave stations, all at 100 kilohertz), which allow a rather complex phase-sensitive receiver to locate its position by ground wave alone to 100 meters or even better. The slave pulses are delayed by known times, given nominally to hundredths of a microsecond, and are corrected after the fact in atomic reference timetables published by the U.S. Naval Observatory. This system is an obvious aid to those who would check their own precision clocks. The author offers a detailed example of his own exploitation of the Sylt slave station "of the Norwegian Sea chain" to check a commercial cesium standard in Bern. The precision of his daily checks was better than a tenth of a microsecond, and the clock, no "well-kept high-performance clock" but "an average apparatus operated without special precautions" and with no external temperature control, was well within the one part in 10^{11} its manufacturers warrant.

Better results require many clocks and tricky weightings for the average. Seven Loran-C chains cover the main waters sailed by the U.S. Navy; a new system, Omega, with eight stations worldwide planned to cover the globe at very low frequencies, promises phase information good to parts in 10^{13} over a long average. Satellite-borne systems are on the way (planned for 1984 deployment) in which a couple of dozen satellites will offer a three-dimensional positioning service, said to be accurate to within 10 meters eventually. Such a system must of course correct for gravitational red shift and much else. From this entire elegant subject there rises a strong sense of irony: these are the military applications of general relativity. The very-low-frequency stations listed include what must be the most powerful radio transmitters in the world, with steady carrier-wave power reaching 2,000 kilowatts.

The book includes the fundamentals for precision work in this field, not at the level of detailed circuitry or design but with concepts, techniques and block diagrams clearly set out. The chapters describe the most frequently applied statistical measures of stability (in the domains of both time and frequency), the various types of standard frequency sources in service (from quartz crystals to cesium, hydrogen, rubidium and methane), the current definition of time scales, the application of electronic counting techniques to the measurement period and frequency, spectrum analysis in this context, phase measurements and the signal-comparison methods sampled above.

The spinning earth is plainly no longer our time giver. The accuracies now offered by atoms or molecules are better by orders of magnitude. The resonant frequencies determined by the quantum states have been made into a true time scale by the wide distribution of the quantum devices. These are now generally equipped with clever nonlinear systems for counting the fast-arriving cycles and performing the four arithmetic operations on the counts. Such apparatus provides accurate registration and display even over long intervals; the expected time between failures for a cesium-beam clock is three years. It is true that the atomic time scale would be lost if all the clocks were destroyed simultaneously, but the author dryly concludes that "such an event might as well leave nobody able to observe the stars" either.

Observatories still maintain the good old earth-spin star time; it is called Universal Time, UT. (Corrected versions are in service, called UT1 and UT2.) Then there is a Keplerian time, defined by the time of the earth's orbit, although it is actually obtained through heavily corrected observations of the moon's motion. It is called Ephemeris Time, ET. Atomic Time, AT, began by agreement at 0h 0m 0s UT2, 1 January 1958. It counts the seconds since (each one defined as 9,192,631,770 periods of the hyperfine radiation of cesium 133 in the ground atomic state). The practical version of AT is reckoned up in Paris after a sophisticated weighting of the readings of many nationally held cesium clocks; it is called the International Atomic Time Scale (TAI). Finally the time signals give us all Coordinated Universal Time (UTC), a compromise time scale generated by the atomic

Now we can replace a lost American Express Card in just one day. And you won't suffer while you wait.

Now in an emergency, most lost or stolen American Express® Cards can be replaced fast—usually by the end of the next business day. That is pretty tough for any other card to beat. But, sometimes, even a day can seem like a long time.

Let's just say you're in a strange country, 6,000 miles from home, and you've lost your wallet, Card, money, travelers cheques, checkbook, luggage, plane tickets. Well, surprisingly, all is not lost. Simply go to the nearest American Express Travel Service Office℠. There are more than 750 of them all around the world.

Here's what they'll do for you:

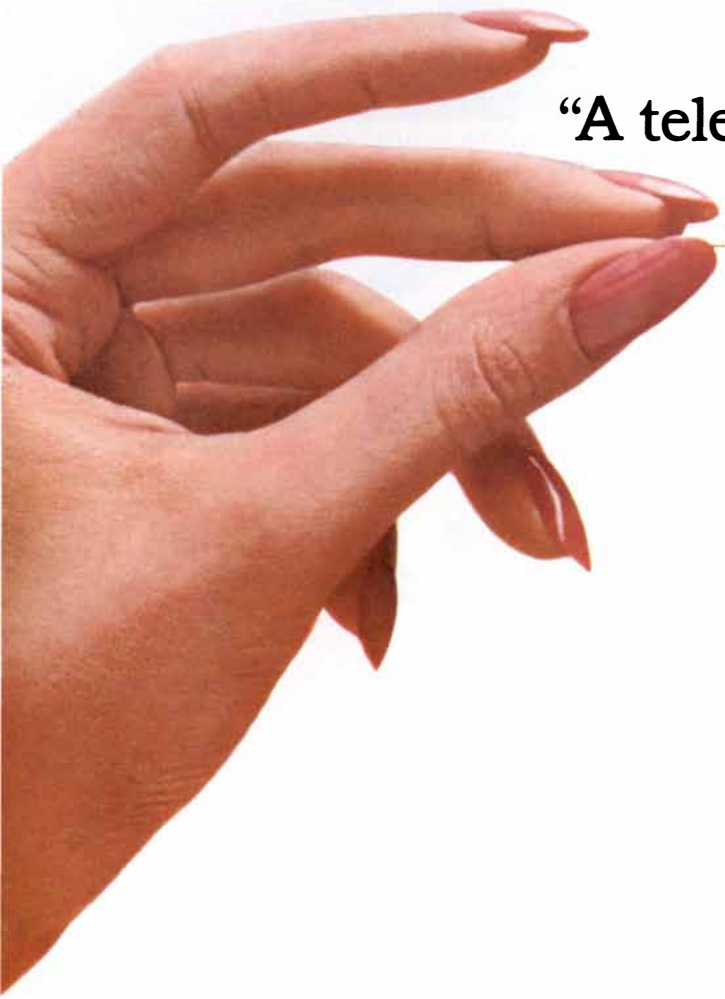
After the Travel Service Office establishes your identity as a Cardmember, most Cardmembers will find they can get a new Card usually by the close of the next business day. Or, in some locations, right on the spot.

If you need lodging until the replacement Card arrives, they'll simply make a phone call to assure you of a room for the night. And so you won't go hungry, they can even arrange for meals.

Lost plane tickets are also a problem they can help solve. They'll confirm a new reservation, and issue a new ticket. And advise you about ticket refunding. And if, for example, you lose the Card in Singapore and are due in Tokyo the next day, they can arrange things so you can pick up the new Card in Tokyo. Or in many other cities all around the world where you'll find American Express Travel Service Offices.

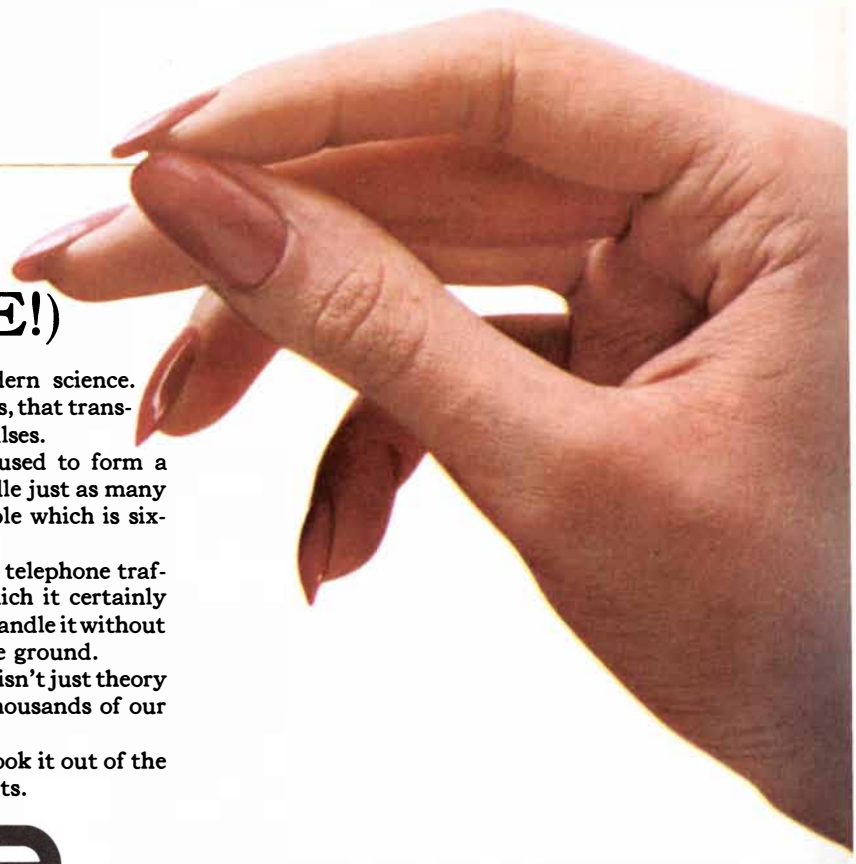
The American Express Card. Don't leave home without it.℠

One Stamford Forum, Stamford, Conn. 06904



"A telephone cable fiber as thin as

a human hair? GEE!"



(No, GTE!)

It's truly a miracle of modern science. A hair-thin fiber, made out of glass, that transmits your voice with light impulses.

Just six of these fibers are used to form a telephone cable. And it can handle just as many calls as conventional copper cable which is sixteen times the size.

The big benefit being that as telephone traffic gets greater and greater (which it certainly will), these cables will be able to handle it without taking up much space under the ground.

"Fiber Optics" (as it's called) isn't just theory anymore. It's actually serving thousands of our customers in California.

Because we at GTE finally took it out of the labs. And put it under the streets.

GTE

Communications/Electronics/Lighting/Precision Materials

THIS FAMILY OF VOLVO OWNERS HAS ONE MILLION MILES ON IT.

Some years ago, Barry Vandenberg, an insurance agent from Oxnard, California, decided on some life insurance for his family.

He wanted his whole family to drive Volvos, because he felt Volvos were the safest cars on the road.

Sure enough, as the Vandenberg children reached driving age, the Volvo population exploded. Today the Vandenbergs own seven Volvos, with a total of one million miles on them.

"These Volvos are still running beautifully, even after all those hard miles," says Mr. Vandenberg. "More importantly, Volvo's safety features help me sleep better at night. My kids are too old to coddle. But they'll never be too old to protect."

Statistics show that 9 out of 10 people who have bought new Volvos are happy, especially with Volvo's safety features.

Building roomy, dependable, safer family cars has always been a Volvo tradition.

It's nice to see that for some families, buying them is a tradition, too.

VOLVO

A car you can believe in.

clocks through TAI but kept within a second of the slowly sliding earth-spin time UT1 by a system of integer leap seconds, plus or minus. These are determined by the Bureau International de l'Heure in Paris, to be entered to taste after due notice on the last second of the month, but particularly in December and June.

THE BLUE REEF: A REPORT FROM BENEATH THE SEA. Told by Alan Anderson, Jr. Alfred A. Knopf ($12.95). "The adventures and observations of Walter Starck . . . at Enewetak Atoll" are rich. Enewetak (the spelling differs from the earlier "Eniwetok" because it was chosen by the people of the atoll in 1972) is a place set apart. Here in 1952 the first thermonuclear device of the unlimited type was exploded; in place of one of the islets of the atoll necklace there appeared "an irregular crater about a mile in length [and] as deep as the lagoon bottom." In 1971 the marine biologist Walter Starck came to the lagoon in his trim white diesel craft, with wife and children, a pet otter, a couple of assistants and a little yellow submarine. His chief aim was offbeat: to approach the gray reef shark close enough to provoke an attack. Six months earlier two even braver shark students from Long Beach, Calif., had 10 times approached such sharks in Enewetak lagoon to elicit the strange display of that species when they are in the presence of a diver. They observed the arched back and wagging, rolling motion of the display, and the case for the validity of their observations seemed strong. Starck was going to clinch the matter under the protection of his submarine, a five-horsepower, free-flooding battery-operated vehicle with canopy openings closed by Plexiglas windows made on the spot. Three months of work and mishap, thoughtful reflection and sharp observation of men and the sea are the stuff of the book.

The gray-shark story is still uncertain, although it is plausibly supported. Many members of this species specialize in taking wounded fish. A surface swimmer is therefore bait for them, although they do not attack a steadily swimming scuba diver. Their ritual display before charging is a signal to other gray sharks, a substitute for combat. Other shark species do not display, apparently because they feed without interspecific competition and on much commoner forms of prey. As the gray shark twists and turns iridescent colors flash along its sides, "rosy and greenish and other hues . . . flashing in the sunlight." Starck's observation is new; it is rare for human beings to describe excited sharks from such close range. The hues may be a diffraction effect, with the fine tips of the dermal denticles acting as diffraction centers.

Another wonder of the lagoon is a certain blenny, a couple of inches long, that mimics the famous cleaning wrasse. The wrasse swims into the mouth of large fishes, or nuzzles their fins and gills, earning its living by cleaning parasites from the big fellows. Just like it in motion and appearance, but with four hidden fanglike teeth, is the sinister blenny, which moves up to the expectant client and takes from some soft fleshy spot a little chunk of meat instead of a parasite! Too many of the blennies and the large fish will no longer tolerate the cleaning wrasse. The diver is also prey to the blenny, but the mouthful the fish takes is tiny, and it does "very little damage" to the diver.

Starck is an autonomous worker of wide and impressive competence. He grew up in the Florida Keys, and he learned there as youths will the subtle arts of the boat handler and the fisherman. Marine scientists who begin to follow the sea only as graduate students lack such hard-won skills, and they draw Starck's cogent criticism for their waste of equipment and time. Much of the time of a little expedition is spent trying to keep tools in working order; these inorganic, mainly metallic, artifacts are under unremitting attack from that "powerful and persistent solvent," ionic seawater. One observation is telling: a cable seemed to have cut its way through inches of a steel beam by its small motions as the seas heaved a moored buoy. It was an illusion, but a common one. The chafing cable had merely removed the corrosion-limiting rust, and in a few years the sea had eaten away the steel under the cable.

Enewetak and its people, exiled and reexiled by our weapons tests, form a background to the tale. Starck is impressed by the normality of the much-irradiated lagoon; there is visible no gross effect of the radioactive load. The debris of human occupation—the huts, cables, towers, wrecks and lost boreholes—are much more obtrusive than direct signs of the scores of weapons tests. The blast-bared islets recover. Starck is of course not able to judge the inner atomic load on the chain of life, which in 1978 was again held to make the atoll unsafe for the Micronesians who once lived there in fear only of typhoons.

This is a book full of fascinating matter, but its manner belies its value. The prose cannot rise above its source, the daily tape recordings of a man whose conversational style clanks with the verbal clichés of our time. ("The sun was pretty brutal in the skiff, and the children were pretty hot by then.") Alan Anderson, an experienced journalist, is unable to supply a style worthy of the depth and novelty of thought. More diagrams would have improved the book; the map supplied and the good photographs of Starck's ship "the *El Territo*" (sic, very often) are more interesting than most of the familiar color pictures of the bright reef fish. Starck has mastered every art of the field scientist save good writing.

AN ILLUSTRATED GUIDE TO POLLEN ANALYSIS, by P. D. Moore and J. A. Webb. John Wiley & Sons ($19.95). Microscopic grains of pollen are coated with a complex polymer of carotenoids and their derivatives, not so very unlike rubber. That coat is only slowly biodegradable, long outlasting other plant tissues; in waterlogged or acid conditions where bacterial action is low the grains remain intact over geologic time. Pollens blow abundantly downwind (a single flower of the common juniper can shed 400,000 pollen grains) and may drift for substantial distances, even thousands of miles seaward. A small grain, say that of flax, measures 20 micrometers across; a big one, say that of the common mallow, measures 100.

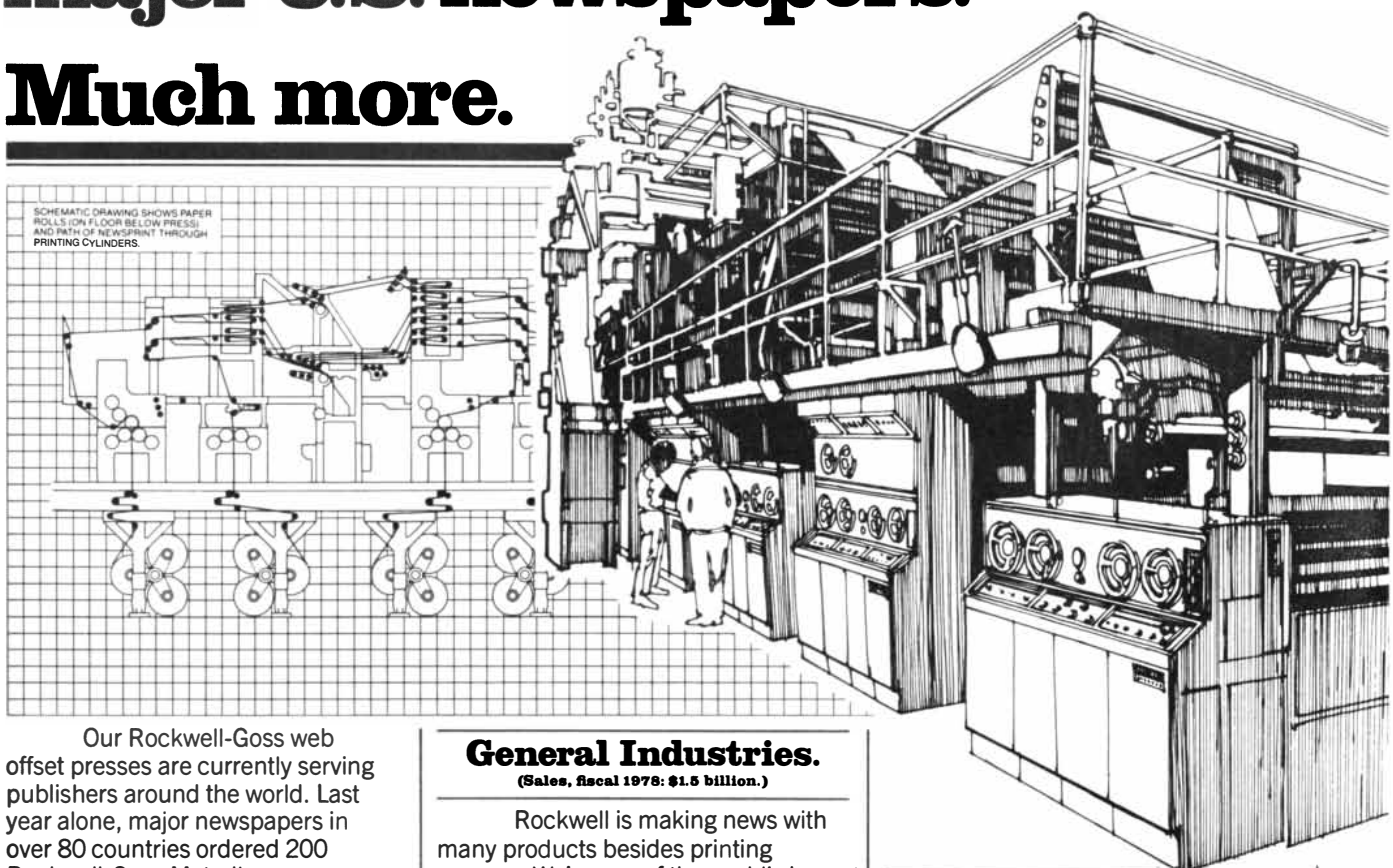
The durable surface is intricately formed, so that under the light microscope the pollen grain of each genus, often that of each species, has a characteristic appearance: ridged, spiny, reticulated and so on. Nowadays the scanning electron microscope reveals some grains of pollen in the round as curious fruit or even as sculptured artifacts, microjewels of the woodland, and a few such pictures are in this book. Mostly, however, the 50 plates, each consisting of some 10 photomicrographs, show under the ordinary light microscope the pollen of hundreds of British plants: sedum and plum tree, aster and mint, oak and elm, ragweed.

The book is at once a text, a laboratory manual and a field guide. The images, made at magnifications of about 1,000, are arranged to provide a workable key to the pollen. That first key relies on the tendency of some pollen to clump in groups, two or four or more, and then on the apertures visible in the grain surface. This may be enough to take the searcher to a single plate of photographs ("with one three-slit aperture in the shape of a Y"), or he may need next a count of the pores or the furrows or both. All together there are 14 classes of grains so keyed, and then specific sculptural detail leads the classifier to just one of the photographs. It is rather harder than sorting wild flowers, at least to the unskilled reader's eye. The authors, two experienced London botanists, nonetheless believe that with their text undergraduates should be able to identify "ninety percent of the pollen found in post-glacial sediments" of Britain.

This taxonomic study is not promoted for its own sake. Over the past 60 years the technique has flowered into a specialized science with its own learned journals, a powerful tool for the study of past climates, vegetation and agriculture, and a valuable means of dating (through the matching of distributions of pollen types layer by layer). The

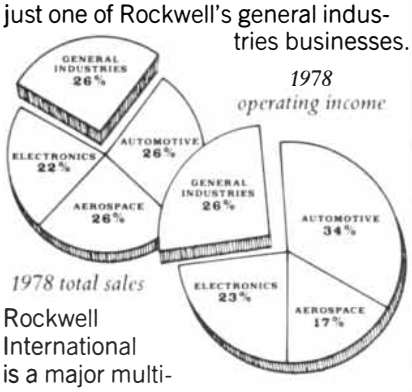
Rockwell is more than a builder of printing presses for most major U.S. newspapers.

Much more.



Our Rockwell-Goss web offset presses are currently serving publishers around the world. Last year alone, major newspapers in over 80 countries ordered 200 Rockwell-Goss Metroliner press units. We manufacture a full line of web offset printing presses, and auxiliary equipment and systems, including some for smaller newspapers and commercial printers.

Printing presses, however, are just one of Rockwell's general industries businesses.



Rockwell International is a major multi-industry company, applying advanced technology to a wide range of products — in general industries, aerospace, automotive and electronics. Following are some examples of our balanced diversification.

General Industries.

(Sales, fiscal 1978: $1.5 billion.)

Rockwell is making news with many products besides printing presses. We're one of the world's largest suppliers of high-technology valves for the energy market and for general industry. We also make textile equipment, industrial sewing machines, power tools and products for utilities.

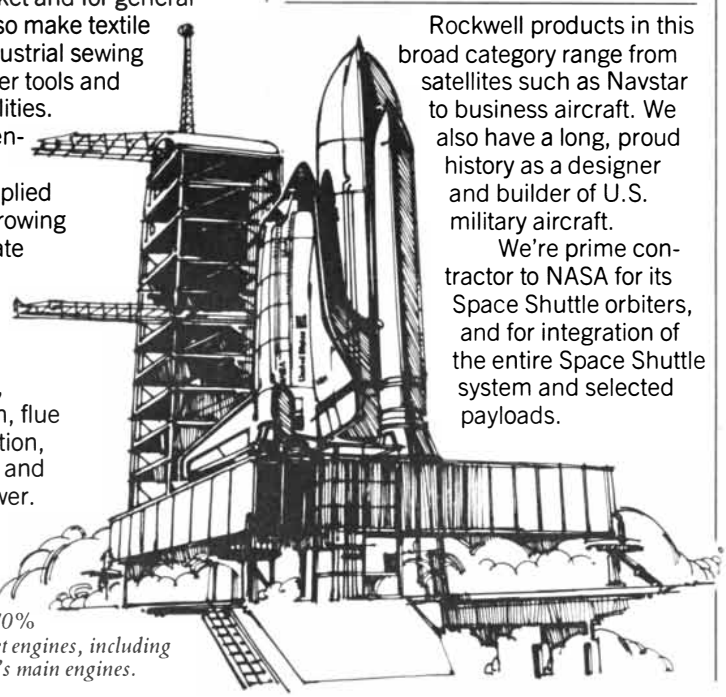
Our extensive technology is also being applied to the world's growing need for alternate sources of energy. We're involved in projects for nuclear energy, coal gasification, flue gas desulfurization, and solar, wind and geothermal power.

Aerospace.

(Sales, fiscal 1978: $1.4 billion.)

Rockwell products in this broad category range from satellites such as Navstar to business aircraft. We also have a long, proud history as a designer and builder of U.S. military aircraft.

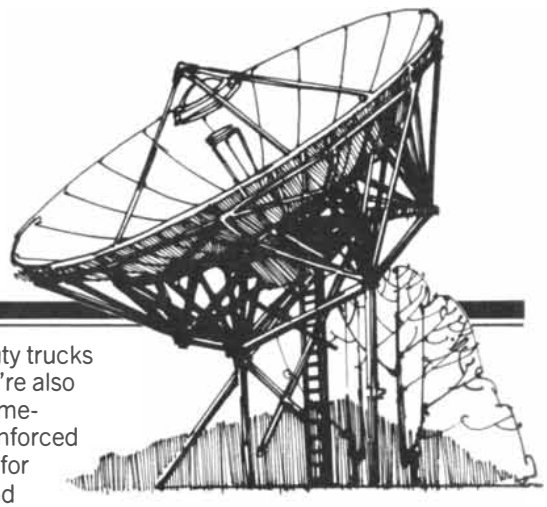
We're prime contractor to NASA for its Space Shuttle orbiters, and for integration of the entire Space Shuttle system and selected payloads.



We've built over 70% of America's rocket engines, including the Space Shuttle's main engines.

Offset printing is now feasible on large presses: the Rockwell-Goss Metroliner offers major daily newspapers improved speed, economy and reproduction.

Rockwell was the first to design and build a complete satellite communications earth station.

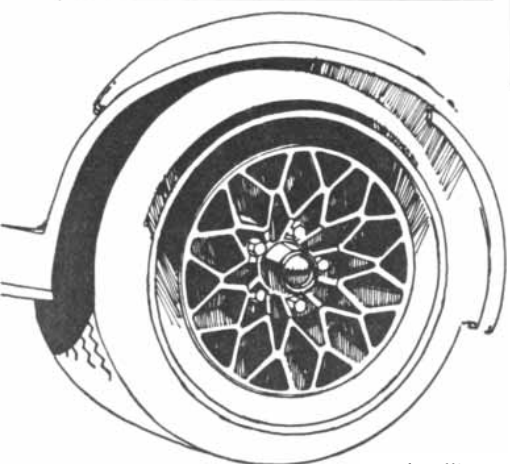


Rockwell's rocket engine technology has played a vital role in America's space program — before and during all of the Apollo moon missions, and into the present age of the NASA Space Shuttle.

The Space Shuttle Main Engine (SSME) we are now developing for NASA is the most advanced project in the history of rocketry. Each SSME will develop nearly one-half million pounds of thrust at Shuttle lift-off, and maintain full power for no fewer than eight minutes. And each SSME will repeat its performance on fifty-five separate launches, or a total of 7½ hours of operation. It's the world's first reusable rocket engine, for the world's first reusable spaceship.

Automotive.

(Sales, fiscal 1978: $1.5 billion.)



Rockwell's styled aluminum wheels can lessen a car's weight up to 34 lbs.

One-half of the highway tandem tractors in North America are equipped with Rockwell axles — and more than half of the heavy-duty trucks stop with Rockwell brakes. We're also a major supplier of drivelines, mechanical devices, castings, reinforced plastic and other components for trucks, trailers, buses, vans and passenger cars.

Producing styled wheels for cars, vans and light trucks is another of our strong automotive capabilities. We're not only a major supplier of styled aluminum wheels to the automotive aftermarket, but to a growing number of America's car manufacturers as well. We also manufacture steel wheels and wheel covers.

Electronics.

(Sales, fiscal 1978: $1.3 billion.)

We're one of the world's leading suppliers of avionics — communications, navigation and flight control equipment for air transport, general aviation and government aircraft. We also make microelectronic systems and devices, telephone call distribution systems, and guidance and control systems. And we manufacture and install telecommunications systems for businesses and governments worldwide.

In addition, we're playing an important role in bringing you television entertainment from the Public Broadcasting Service. Many public television programs are beamed into space and bounced off a satellite for broadcast to homes all over America. And 162 earth stations, constructed by our Collins Transmission Systems Division, do the receiving.

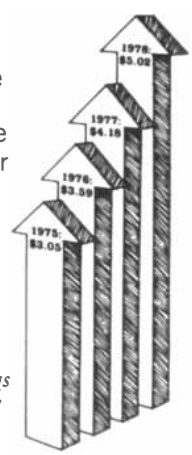
Together, they form the world's largest satellite communications network, and provide public television stations with added economy and program flexibility. Another Rockwell-built network of 192 earth stations is now being completed for the National Public Radio Service.

Over 14,300 scientists and engineers.

Of our 114,000 employees, one in eight is either a scientist or an engineer. They constitute about one percent of America's total scientific-engineering community. This technological base positions us for leadership in each of our product areas. It also makes our corporate slogan, ". . . where science gets down to business," a fact.

Business is good.

Rockwell International's total sales for 1978 were $5.67 billion. Fiscal 1978 was the most profitable year in Rockwell's history. Net income, up 23% over fiscal 1977, was $176.6 million, generating a record $5.02 per share.



Earnings per share

For more of the Rockwell story, please write us: Rockwell International, Dept. 815S, 600 Grant Street, Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

samples of choice are peats, lake sediments and some acid soils. They are taken with simple boreholes or coring tools, and their pollen content is separated and mounted on microscope slides for counting and identification. In easy cases boiling with lye (KOH), sieving, washing, centrifuging and finally staining (a bright red) will yield a purified pollen suspension for the slides. If plant cellulose abounds in the peat, an acetic acid treatment will remove it by acid hydrolysis. For almost all organic peats these two techniques suffice; inorganic impurities call for more difficult procedures.

Then comes the careful path through the key and the long counts. Pollen data are generally plotted as percentages by plant type, stratum by stratum a few meters down. Rather intuitive means of characterization and pattern zonation work well; recent studies with the full sophistication of numerical taxonomy have generally confirmed the judgments of earlier workers. Simple analysis can still tell much; this book is an open invitation for "unskilled practitioners" to take up this "robust and flexible tool." The scanning microscopes, the extraction by hydrofluoric acid, the subtle computer programs are not indispensable. It is true that it may not be so easy to make the decisions of form from the key, nor is it plain that American bog samples will be so well represented by the British forms. The volume nonetheless illuminates the hidden powers of palynology; there can be few frontier research techniques that are so accessible to students.

PHOTO-ATLAS OF THE UNITED STATES, produced by Photo-Geographic International. Crown Publishers, Inc. ($7.95). A volume of colorful Landsat photographs, each plate showing a piece of the earth 100 miles or so on an edge, was reviewed two years ago in these columns. Beautiful and instructive, that big book sampled our globe's wonders, displaying atoll and massif, delta and dune field, fault rift and forest, generally rendered in the bright false-color palette of the three-band spectrum scanner. The pages of this modest photographic atlas are only duotone (halftones in black and blue-green, mainly from pictures made in the red spectral band), optimistically billed as "the actual color of the landscape on a clear, spring day." They do, however, present the entire face of the country: "the first complete photographic atlas of the U.S. using satellite photography."

Each spread, about 15 inches by 12, shows a piece of our land at a uniform scale of 18 miles per inch. Overlaid in white on the Landsat pictures are the state boundaries, and in addition important physical features and an orienting selection of cities and towns are clearly marked with their names. Although it is less visually striking than the three-color atlas, this is a utilitarian prize, a comprehensive physical map of the country offering a clear impression of every part, at a scale rather like that of the unaided view from the window of a long-range jet airplane. These halftones do not bear magnification; resolution is lost in the bargain reproduction of the original mosaic photographs, since the dots run about six to the mile, apart from the overprinting of the two inks. Niagara Falls cannot be made out, nor the Hoover Dam, but the watercourses they cross are clear. The great open-pit mine at Bingham Canyon in Utah is visible, as are the launch pads and roads at Cape Canaveral and many highways.

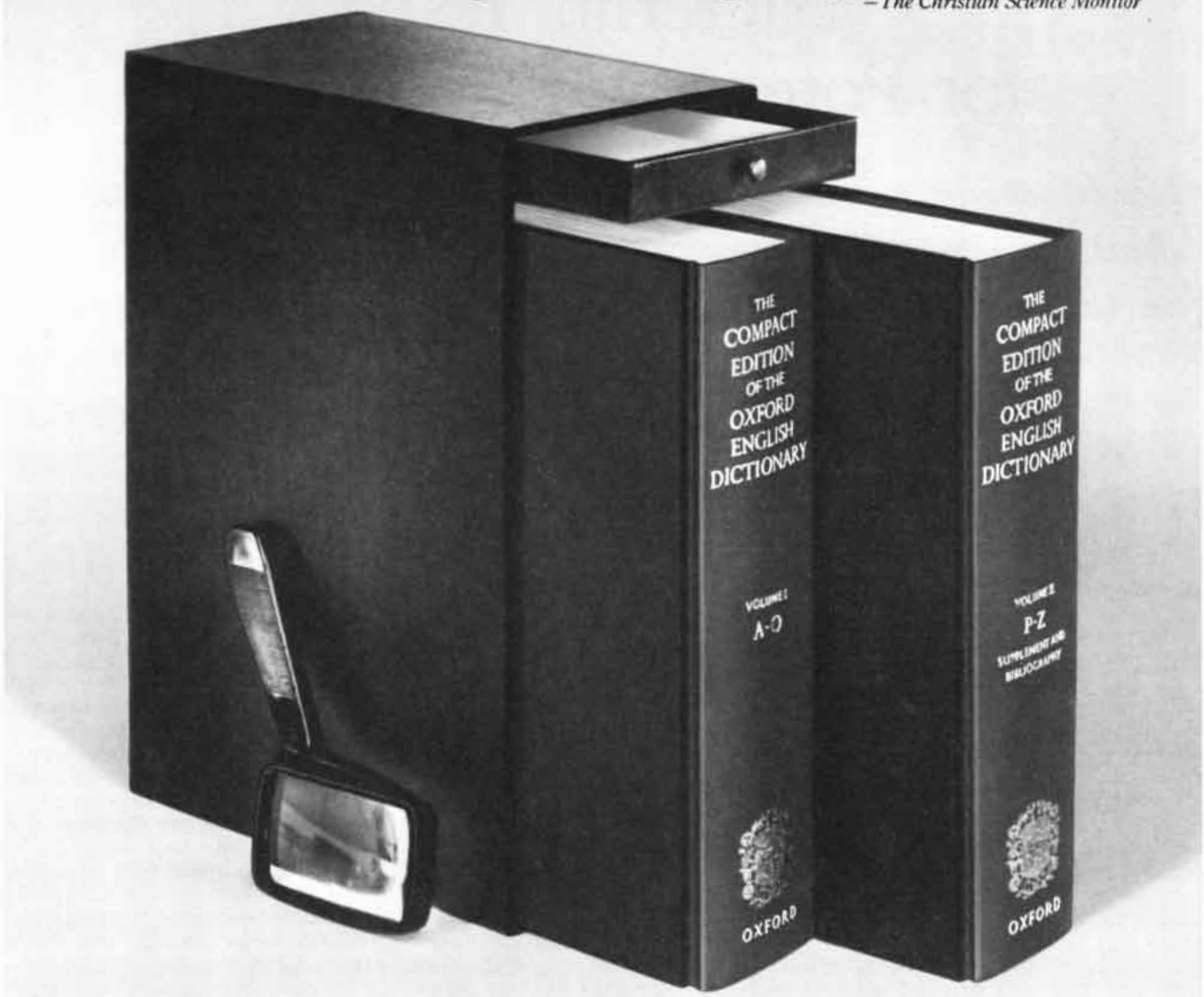
Coasts and shores are sharp. Land-use patterns are generally distinct, from city squares to the section lines of prairie farms. Clouds are a minor annoyance. Across the Turtle Mountains of North Dakota runs the U.S.-Canada border. A section of white line is deliberately omitted over the image of those mountains (as it is in a few other places), yet the knife-straight border remains sharply marked, thick forest on the Canadian side parted from clear-cut logging and road grids to the south. The narrow glacial ridge that divides the source of the Red River, which drains north through Winnipeg to Hudson Bay, from the source of the Minnesota River, whose waters end up in the warm Gulf of Mexico, is clearer here than in any other atlas. Alaska and Hawaii are included, but at much reduced scale.

For lagniappe the producers—a cartographic team from Mountain View, Calif.—have included 10 aerial photographs in color at various scales, showing 10 big cities from the air. The colors are pseudonatural, much touched by human hands. A column of brief helpful commentary runs down the margin of every spread, calling attention to much one can read from the photograph. For example, the waters of the Great Salt Lake hold a mysterious boundary, a perfect straight line for 20 miles, the southern half of the lake much darker in tone than the northern. It is the sign of the embankment that carries the railroad across the lake, "preventing the saltier water to the north from mixing with the water to the south." In the salty part red algae thrive; in the less brackish half the blue-greens abound. The red-favoring detector in orbit records the difference.

Furnished with this physical map, all but alive, augmented by one of the excellent cheap highway atlases for political mapping in detail, any traveler, real or virtual, airborne or even on foot, is soon well located, anywhere in the lower 48. There is a good index map. Meteor Crater is unfortunately caught in the fold of the pages, but this is a chance loss in a useful book, one that would have seemed an incredible marvel only 30 years ago.

"The most complete and most scholarly dictionary of the English language"

—*The Christian Science Monitor*



The Compact Edition of the Oxford English Dictionary. As an introduction to membership in the Book-of-the-Month Club®

YOURS FOR ONLY

$19.95

(PUBLISHER'S LIST PRICE: **$125**)

You simply agree to buy 4 books within a year.

FOR the price of an average dictionary, you can now treat yourself and your family to the world's best dictionary of the English language. The contents of this two-volume edition are identical to those of the original thirteen-volume set, priced at $495.

FEATURES

- Boxed set of two volumes, 9¾" by 13½" each.
- All 16,569 pages of the 13-volume original included in 4134 pages of *The Compact Edition* through a photo-reduction process which permits printing of four pages of original on one page of compact edition.
- Paper is 30-pound Special Dictionary White.
- Binding is library buckram reinforced and gold-stamped.
- Bausch & Lomb magnifying glass included in special drawer of slipcase. 2" by 3⅜" lens scientifically designed to make reduced print easily readable.

Book critics call the OED a "miracle." Join the Book-of-the-Month Club and this "miracle" can be yours now. As a Club member, you'll continue to enjoy benefits on the best and most important books published today.

Bookstore Quality at Book-Club Savings You conveniently shop at home at considerable savings. Whether you're adding up your remarkable savings on the introductory offer, or on books you are offered as a member, these are always true savings...because every Club book is as good as, or better than, those sold in stores. You *don't* settle for the altered or inferior editions that some book clubs send their members.

Book-Dividends. When you remain a Club member after the trial period, every book you buy earns Book-Dividend® credits. These entitle you to choose from a wide variety of significant books at hard-to-believe savings of *at least 70%*.

BOOK-OF-THE-MONTH CLUB®
AMERICA'S BOOKSTORE®

Established 1845 SCIENTIFIC AMERICAN August 1979 Volume 241 Number 2

Bacterial Tests for Potential Carcinogens

New short-term tests can identify environmental agents that cause damage to DNA, the primary event in chemical carcinogenesis. The tests are also valuable for clarifying the mechanism of DNA damage

by Raymond Devoret

The fact that physical and chemical agents in the environment enhance the incidence of cancer has become of great concern. It is clear that an adequate limitation of human exposure to carcinogens would save lives. To identify potential carcinogens in the environment is therefore an urgent task. In the case of chemical carcinogens that is no easy matter. It is estimated that more than 50,000 different man-made chemicals are currently in commercial and industrial use; between 500 and 1,000 new chemicals are put on the market every year. The standard animal tests for potential carcinogenicity take a long time and cost a great deal of money.

Fortunately there is an alternative to the classical animal tests. One can take advantage of the profound unity of living matter and resort to bacteria as the test organisms. A bacterial assay for carcinogenicity takes a few hours or days rather than the two or three years required for an animal test, and it costs far less. An effective bacterial test has been developed by Bruce N. Ames of the University of California at Berkeley, based on the ability of a chemical to cause mutations in bacteria. More recently my colleagues and I at the Centre National de la Recherche Scientifique in Gif-sur-Yvette have devised a group of tests based on a chemical's ability to induce the development of a dormant virus in bacteria. The Ames test and our tests not only provide means of identifying dangerous chemicals but also are powerful new tools for learning to understand the primary events of the carcinogenic process initiated by chemicals.

Cancer is a disease of highly evolved multicellular organisms such as human beings, whereas bacteria are minute single cells at the opposite end of the evolutionary scale. It may therefore seem paradoxical that bacteria can serve to identify substances that cause cancer. Actually, however, there is no paradox.

One tends to think of a cancer as a tumor that can spread through the body to form multiple tumor colonies (metastases). That is a clinical, macroscopic view of cancer at a multicellular stage, since just one gram of malignant tumor already contains a million cancer cells. Cancer begins at the level of the single cell. A cell in an adult tissue evolves in such a way that it departs from conformity with the strict physiological rules governing the set of identical cells that constitute a tissue; it becomes a unique and distinguishable defect in an otherwise monotonous structure. The cell begins to divide and a tumor grows. Some of the daughter cells may break the tissue barrier, invading adjacent tissues and usually metastasizing to distant sites. Cancer cells have a great selective advantage, since they escape the programmed fate of most normal cells: to age and die. For cancer cells the entire body is a culture medium in which they thrive, ultimately to die with the body they kill.

Physical and chemical agents in the environment cause cancer by damaging DNA, the cell's hereditary material. DNA damage initiates a complex cellular process that in mammalian cells can eventually lead to transformation into a cancerous state. Agents that damage DNA are therefore potential carcinogens. DNA is the hereditary material of all living cells, and both DNA lesions and the cellular processes that repair them are remarkably similar in bacteria and in human cells; what is detrimental to bacterial DNA is likely to harm human DNA. That is the theoretical justification for substituting bacteria for mammalian cells in tests to detect damage to DNA.

The theoretical justification is supported by experimental and practical results: bacterial tests distinguish with more than 90 percent reliability between known carcinogens and known noncarcinogens, and they have identified as potential carcinogens new chemicals that have subsequently been shown to be carcinogenic in animal tests. Of course, the manifestations of DNA damage are very different in bacteria from the transformation to the cancerous state that is observed in mammalian cells. In compensation the bacterial tests are so much faster and less expensive as to finally make comprehensive screening for potential carcinogens a feasible objective.

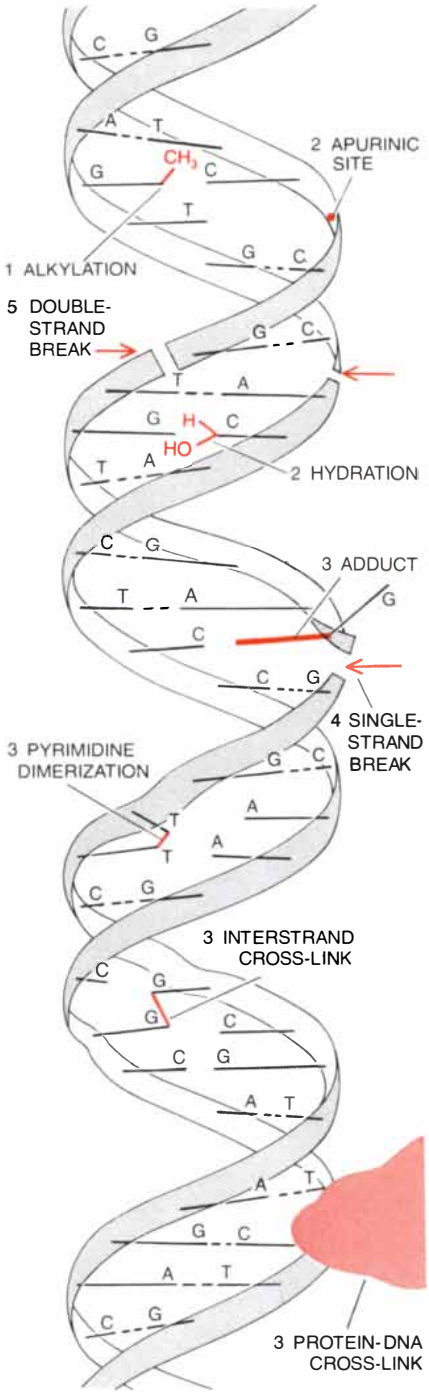
The carcinogenic potential of physical agents, and notably of ionizing radiations, is much better understood than that of chemical agents. Broad popular awareness that radiations can cause cancer has come only in the era of nuclear weapons and reactors, but the danger actually became apparent soon after the discovery of X rays by Wilhelm Konrad Röntgen in 1895. Only four years later it was reported that a technician who checked newly manufactured X-ray tubes by fluoroscoping his own hand was afflicted with a skin cancer; he eventually died of it. The warning was ignored, and most of the first generation of radiation therapists died of cancer. The carcinogenic effect of the ultravio-

let radiation in sunlight has also been known for some time. As long ago as 1905 a French physician named Dubreuilh observed that skin cancers of the back of the neck were particularly prevalent among workers who were exposed to the sun as they tended the vineyards and harvested grapes in the Bordeaux region.

Although X rays, gamma rays and other radiations hit all the components of cells in a random manner, it was recognized quite early that the genetic material must be the most radiation-sensitive cellular target. DNA constitutes only a minor fraction of the chemical components of a cell, but direct or indirect damage to it has great impact on the cell's future, whereas damage to proteins and other cellular components has much less effect. That is because DNA is the memory of the cell. DNA replicates to beget DNA, and so errors in DNA are transmitted from cell generation to cell generation.

The double helix of DNA consists of two chains of sugar (deoxyribose) and

THREE BACTERIAL TESTS for potential carcinogens reveal DNA damage: the Ames test (*top*), the inductest (*middle*) and the lambda mutatest (*bottom*). In each case the culture plate at the left is an untreated control; tester bacteria on the center plate were treated with a moderate dose and those at the right with a higher dose. The Ames test shows the extent of reverse mutations in histidine-deficient *Salmonella typhimurium* that enable the revertant bacteria to proliferate. A background of spontaneous mutant colonies (*red stain*) is seen at the left. Many more colonies grow when tester bacteria are exposed to 250 nanograms (*center*) and 750 nanograms (*right*) of the potent mutagen (and carcinogen) nitrosoguanidine. In the inductest (*middle*) DNA damage is revealed by the induction of a prophage, a dormant bacterial virus integrated in the DNA of "lysogenic" *Escherichia coli;* mature phage particles burst out of the tester bacteria and create plaques on a lawn of indicator bacteria of strain *A* (*red stain*). Here the DNA damage was caused by the antitumor drug mitomycin C, 10 nanograms of it on the center plate, 200 nanograms at the right. In the mutatest (*bottom*) a modified form of the prophage makes plaques on a lawn of *E. coli* of strain *B* (on which nonmutated phage cannot form plaques) when it undergoes mutation in its "operator" regions and can no longer remain dormant. Again the treatment was 10 nanograms of mitomycin C (*center*) and 200 nanograms (*right*).



STRUCTURAL ALTERATIONS are imposed by radiations or by chemical agents on the double helix of DNA, two chains of sugar and phosphate groups (*helical bands*) linked by paired bases: either adenine (*A*) and thymine (*T*), or guanine (*G*) and cytosine (*C*). The alterations can be classified in five categories, examples of which are illustrated: negligible helix distortions (*1*), as by alkylation of one of the bases; minor distortions (*2*) caused by hydration or the absence of a base; major distortions (*3*) caused by insertion of an "adduct," linking of two bases to form a dimer, or cross-linking between the two strands or between a strand and a protein; breaks in a single strand (*4*) or in both strands (*5*). Any structural alteration affects DNA's function as a template for replication, but some negligible alterations are not sensed by a cell as damage to DNA.

phosphate groups linked, as by the rungs of a twisted ladder, by paired nitrogenous bases: two purines (adenine and guanine) and two pyrimidines (thymine and cytosine). Adenine always pairs with thymine and guanine with cytosine, but the sequence of the bases along a strand of DNA is variable and carries a particular coded message. Any alteration, even a slight one, in the structure of the double helix affects the functions of the DNA, one of which is to serve as a template for its own replication.

Not every DNA alteration is sensed in the cell as DNA damage. Defined precisely, DNA damage is an alteration that constitutes a stumbling block for the replication machinery and hence hampers the replication of DNA, endangering the survival of the cell. Once incurred, DNA damage calls for repair, which is accomplished by the interplay of at least a score of enzymes whose action is governed by as many genes. Repair is never totally efficient, and so many cells die. Some cells may survive, however, even though the lesions are not totally removed from their DNA, because a repair process has bypassed the lesions. Replication then reconstitutes an undamaged double helix, but one bearing a coded message different from the original one. The scars left on the DNA by such a process are mutations.

The correlation between radiation-induced cancer and radiation-induced DNA damage has long been apparent to radiation biologists, but it is only recently that molecular evidence has been found that DNA damage is a direct cause of cancer. The evidence comes from patients suffering from xeroderma pigmentosum, who are extremely sensitive to sunlight and, while they are still very young, develop skin cancers of which they may eventually die. James E. Cleaver of the University of California School of Medicine in San Francisco and Dirk Bootsma of Erasmus University in Rotterdam have demonstrated that xeroderma patients suffer from a well-characterized genetic defect: their cells cannot carry out a particular DNA-repairing process.

At noontime on a sunny day the flow of ultraviolet radiation that reaches the earth is strong enough to generate pyrimidine dimers in the DNA of exposed cells by linking two laterally adjacent thymines or a thymine and a cytosine. Most such DNA lesions in skin cells are repaired in normal people by an excision process strikingly similar to the "cut and patch" repair process in bacteria exposed to the same radiation. In xeroderma patients, however, the lesions go unrepaired, and their accumulation appears to bring on cell transformation: DNA damage breeds cancer.

Appreciation of the carcinogenic role of chemicals in the environment has come slowly, even though instances of cancer caused by occupational exposure have long been observed. As early as 1775 the British physician Sir Percival Pott correlated the incidence of cancer of the scrotum in men who had once been chimney sweeps with the accumulation of soot in their groin area many years before. Experimental findings on chemical carcinogens date back at least to 1918, when two Japanese investigators, K. Yamagiwa and K. Ichikawa, showed that skin cancers could be induced by repeated applications of coal tar to the ear skin of rabbits. One of the chemicals responsible for causing such cancers is benzo[*a*]pyrene, which is also present in soot, cigarette smoke and charred meat.

Progress toward understanding chemical carcinogenesis has been slow for at least three reasons. First of all, most chemical carcinogens are not biologically active in their original form, so that testing them in that form does not reveal their carcinogenic nature. Only about a decade ago did it become clear, as a result in particular of the investigations of James A. Miller and Elizabeth C. Miller of the University of Wisconsin, that normal metabolic processes, which convert food into substances the body can absorb and eliminate and convert harmful compounds into harmless ones, transform environmental chemicals into metabolites capable of inducing cancer. Those metabolites had to be characterized in order to show just how their parent substances are threats.

A second delaying factor was that the actively carcinogenic metabolites react with various cell components, including RNA and proteins as well as DNA. Since there are a lot of proteins, performing important functions, it was often assumed that damage to proteins might play the major role in the cancerous transformation initiated by chemical carcinogens. The key role of damage to DNA in carcinogenesis by chemicals was recognized only recently, and the results of the bacterial tests have provided strong (although indirect) evidence for such a mechanism.

Finally, the understanding of chemical carcinogenesis has been obscured by the fact that DNA damage, although it is the essential event in the initiation of carcinogenesis, does not usually in itself lead to cancerous transformation; additional factors are apparently required to promote the complex chain of cellular events culminating in the transformation. DNA-damaging agents in themselves are therefore only potential carcinogens.

As I pointed out at the beginning of this article, only a small fraction of the flood of chemicals reaching the market every year can be tested accurately by means of the standard animal assays. In order to obtain results with statistical significance a great many animals must

be (or at least should be) exposed to each tested chemical; for that reason alone a comprehensive screening of new chemicals would be impractical (other than the few, such as food additives, drugs and cosmetics, for which testing is mandated). Even when animal tests are feasible, manufacturers need low-cost, fast tests if they are to identify DNA-damaging substances while new products are still under development and alternative ones can still be sought.

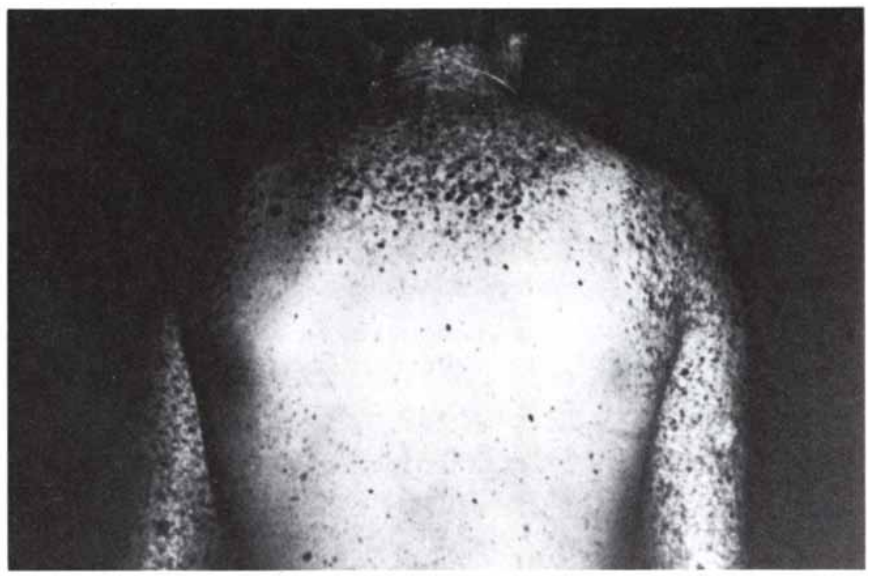
Studies of the epidemiology of human cancers have provided much information on environmental carcinogens, but such studies are of little immediate value in detecting new potential carcinogens because most human cancers appear only some 20 or 30 years after the exposure that gives rise to them. Epidemiological analysis has most frequently been successful in detecting a chemical to which an identifiable subpopulation, such as workers in a particular industry, are exposed and because of which they show a high incidence of a particular type of cancer. For some widely distributed carcinogens there are no clearly identifiable subpopulations for statistical analysis.

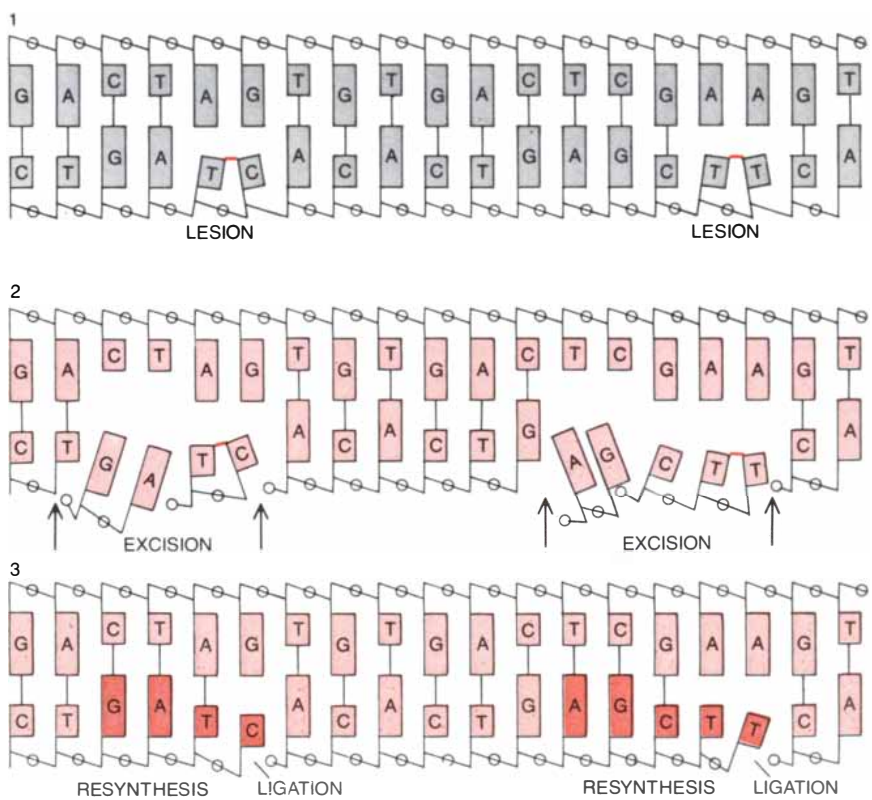
There is no doubt, then, that the present situation calls for simple, fast, inexpensive methods of detecting potential carcinogens, which is to say DNA-damaging agents. One possibility is to measure DNA damage directly in mammalian cells by determining biochemically the incidence of particular forms of DNA damage in cells exposed to a chemical. Such studies are being done by molecular biologists, and their results provide valuable standards of reference. For screening purposes, however, it is cheaper and faster to detect and measure the extent of DNA damage by scoring its manifestations in bacteria.

One of the great advantages of assays done with bacteria is the enormous biological amplification implicit in bacterial manipulations. It is easy to grow as many as a billion (10^9) bacteria per milliliter of culture medium. A mutational event such as a change in a single base pair in the bacterial DNA, which is impossible to detect by standard biochemical methods, will be revealed as a mutant bacterium. That single bacterium can be selected from among 10^9 cells because its daughter cells, and only they, will proliferate and form a colony visible to the unaided eye on an agar nutrient plate. Since a colony consists of about a million (10^6) bacteria, a rare single mutational event with a probability of, say, one in 100 million (a probability of 10^{-8}) would thus be amplified by a factor of 100 trillion (10^{14}).

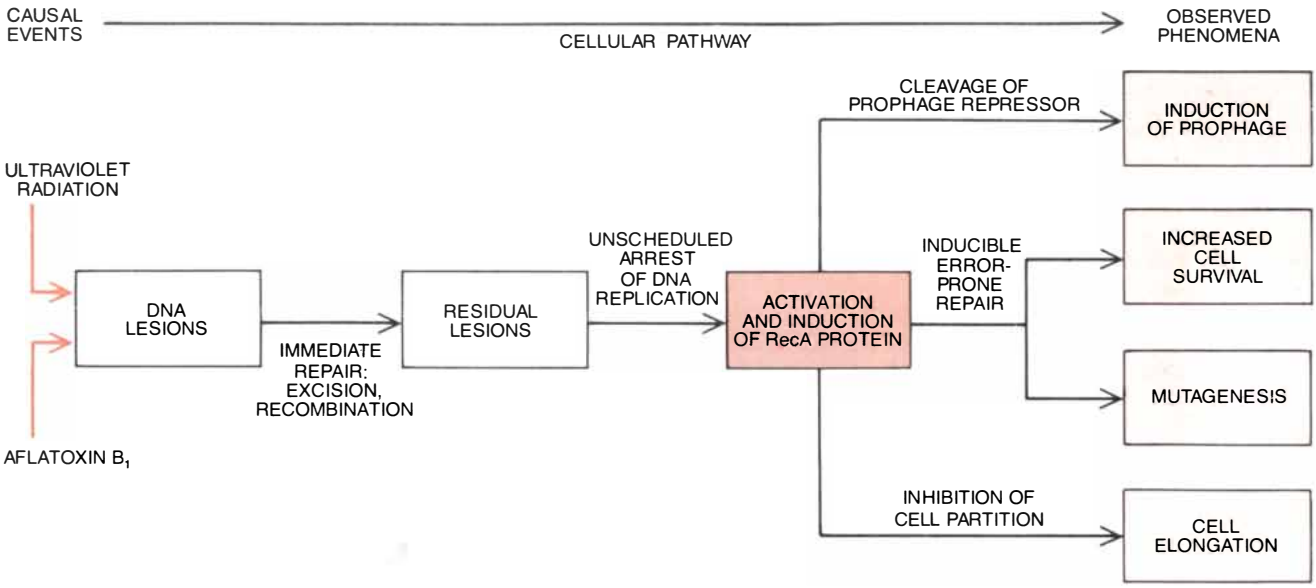
The series of cellular events that leads to mutation, and to several other manifestations of DNA damage in bacteria, has recently been somewhat clarified. The bacterium *Escherichia coli*, which inhabits the colon of a number of mammals including human beings, is genetically programmed to divide and (under optimal conditions for growth) form two daughter bacteria in 30 minutes. When the DNA of a bacterium is damaged, the bacterium may need more than two hours to resume its cycle of division (if it is not killed by the damage). During that time there is a sequence of cellular events.

Immediately after the primary DNA lesions are incurred a first round of repair is effected: an excision repair, in which the damaged segment is cut out of the DNA and replaced by an undamaged DNA sequence. Some residual lesions will remain, however. If they are bulky or clustered, the residual lesions hamper the DNA-replication machinery, and replication stops abruptly. Unless the arrest of replication is transitory

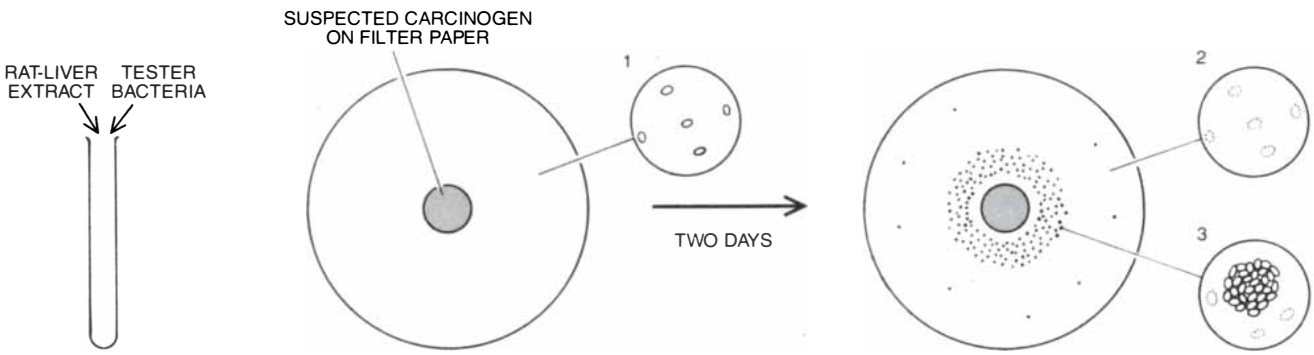




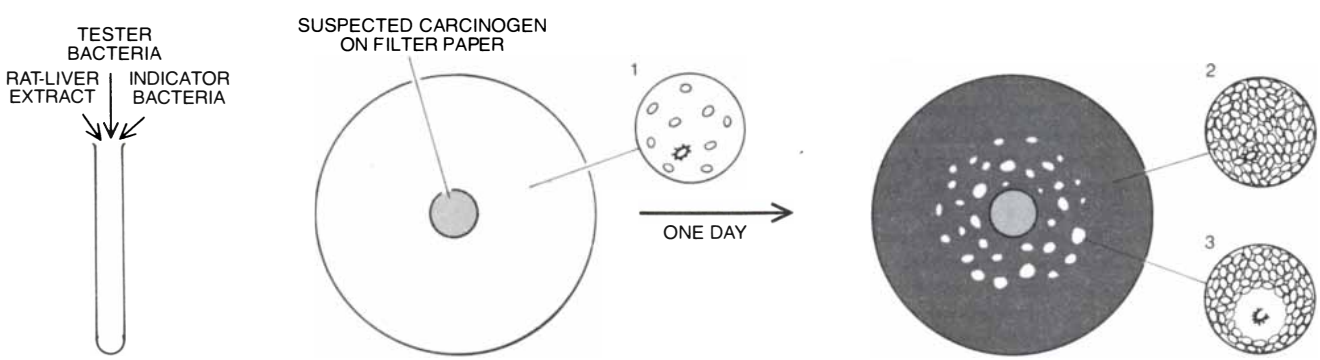
XERODERMA PIGMENTOSUM is a genetic disease whose victims develop skin cancer as a result of normal exposure to sunlight, typically on uncovered parts of the body, as is shown in the photograph at the top. These patients lack a DNA-repair process that in normal individuals removes thymine dimers, the lesions created by the ultraviolet radiation present in sunlight (*1*). In the normal repair process the bases in the lesion areas are first excised (*2*). Then the excised stretches of DNA are resynthesized and the new segments are ligated to undamaged stretches (*3*).



SEQUENCE OF CELLULAR EVENTS follows DNA damage imposed on *E. coli* by ultraviolet radiation or by the carcinogen aflatoxin B_1. Immediate repair processes mend most of the lesions but leave residual damage, causing an unscheduled arrest of DNA replication. This threat to cell survival activates the RecA protein to become a protein-cleaving enzyme and also induces the synthesis of more RecA. Various forms of the protein are involved in three processes that give rise to the four observed phenomena shown at right.



MUTAGENESIS is detected in the Ames test by mixing an extract of rat liver (which supplies mammalian metabolic functions) with tester bacteria (which cannot grow because a mutation makes them unable to manufacture histidine, a necessary nutrient) and plating the mixture on an agar medium so that a thin layer of bacteria covers the medium evenly, as is shown on a microscopic scale (*1*). In this "spot assay" a dose of the chemical to be tested is placed on a disk of filter paper on the tester bacteria. After two days most of the *his*$^-$ bacteria have died for lack of histidine (*2*), but DNA damage caused by the chemical diffusing out from the disk has given rise to mutations, some of which result in reversion of the *his*$^-$ mutation. The histidine-making revertant bacteria proliferate, forming visible colonies (*3*).



INDUCTION of a dormant bacterial virus, prophage lambda, is detected in the inductest. Lysogenic tester bacteria are mixed with rat-liver extract and then with indicator bacteria. The mixture is plated; the medium is covered with a thin layer of indicator bacteria interspersed with a few lysogenic bacteria (*1*). After a day most of the plate is covered by a thick lawn of indicator bacteria (*2*). Where the chemical that is being tested has diffused from the filter-paper disk the DNA damage it causes leads to the induction of mature lambda phage. The phage particles burst out of the lysogenic cells and kill indicator bacteria in the vicinity, making visible plaques on the lawn (*3*).

the cell's survival is at risk. As Miroslav Radman of the University of Brussels and Evelyn M. Witkin of Rutgers University first suggested, the following cellular adaptive mechanisms come into play to cope with the emergency:

1. A second round of repair is induced. This inducible and error-prone repair (nicknamed "SOS repair") tends to restore the structure of the DNA even though there are errors in the coded message; indeed, this adaptive response may be successful partly because it does not "bother" to follow the base-pairing rules of normal DNA replication. At any rate, the price paid for cell survival appears often to be mutagenesis.

2. Cell partition ceases. The elongation of the cell that ordinarily precedes division is protracted, and the cells may form filaments. The elongation may be adaptive in that it facilitates recombination between the two sets of damaged chromosomes in the cell, the intact segments of each combining to yield an intact chromosome.

3. If a prophage, or dormant bacterial virus, is present in the cell, it is induced to develop into a large number of mature progeny phage that burst out of the cell. This adaptive response evolved by the prophage ensures its survival when a host cell appears to be doomed to die: the rats leave the sinking ship.

Each of these adaptive responses is in part promoted by a multifunctional protein called the RecA protein (for "recombination," since defects in the protein impair recombination in general). The RecA protein appears to be required in *E. coli* not only for recombination but also for the responses listed above. The RecA protein is activated and its synthesis dramatically increased when DNA replication is blocked.

Of the various bacterial manifestations of DNA damage, Ames chose mutagenesis as the basis of his pioneering work to develop a test for potential carcinogens. His *Salmonella*–mammalian-liver assay, known generally as the Ames test, is currently the standard test and by far the most widely used. The tester organism is a strain of *Salmonella typhimurium*, another colon bacterium, bearing a mutation (*his*$^-$) that renders it unable to manufacture one of the enzymes required for the synthesis of the amino acid histidine, a necessary component of proteins. As a result of the mutation the bacterium is unable to grow in a mineral nutrient medium unless the medium is supplemented with an external supply of histidine.

On very rare occasions a *his*$^-$ mutation undergoes reversion: a back mutation restores the DNA's normal coding sequence for the needed enzyme and thereby restores the internal supply of histidine. The reversion can be scored because only the revertant bacteria form colonies on a medium that lacks histidine. Obviously the spontaneous rate of reversion, which is ordinarily very low, will be considerably enhanced if the *his*$^-$ bacteria are exposed to a chemical that induces mutations. This is the theoretical basis of the Ames test.

Actually Ames and his colleagues had to introduce three important modifications into the original *his*$^-$ strain to make it a sensitive and versatile tester bacterium. Bacteria such as *E. coli* and *S. typhimurium* have a rather impermeable envelope that reduces or even prevents the penetration of many chemicals into the cell. (This bacterial armor has evolved because the bacteria must usually survive in a hostile environment such as an intestine or a sewer.) Ames and his colleagues overcame the envelope barrier by introducing a mutation that gives rise to defects in the envelope. They went on to make the strain more sensitive to DNA-damaging agents by eliminating its capacity for excision repair, so that most of the primary lesions remain unhealed. And they introduced into the bacterium a plasmid, a foreign genetic element that makes DNA replication more error-prone. By means of these three modifications a strain was constructed in which just a few molecules of a carcinogen are able to create DNA lesions, each of which is likely to engender a mutation; of those mutations, some will be such that the internal supply of histidine is restored.

The real breakthrough, and the one that made the *Salmonella* test truly effective, was Ames's idea of mixing the tester bacteria with an extract of rat liver and thereby subjecting the tested chemical to mammalian metabolic processes. As I pointed out above, it is usually not the original form of a chemical carcinogen that is active but rather one of its metabolites. Because the liver is the body's major metabolic factory the enzymes of a rat-liver extract should convert the chemical being tested into metabolites that react with DNA—if there are any.

In practice the Ames test is usually done by adding the chemical to be examined to *his*$^-$ tester bacteria immersed in a rat-liver extract and plating the mixture on a solid nutrient medium devoid of histidine. (For demonstration purposes the chemical can instead be spotted on a disk of filter paper, which is then placed on a medium on which the bacteria, mixed with the liver preparation, have previously been plated.) After two days of incubation any cells that have undergone the reversion mutation will give rise to revertant colonies. The number of such colonies per mole of the tested chemical provides a quantitative estimate of the mutagenic potency of the chemical.

The simplicity, sensitivity and accuracy of the *Salmonella* test for screening large numbers of environmental sources of potential carcinogens has resulted in its current application in more than 2,000 governmental, industrial and academic laboratories throughout the world; it is estimated that 2,600 chemicals have been subjected to the test. Ames and his collaborator Joyce McCann have themselves validated the procedure by testing more than 300 chemicals that were previously reported, on the basis of animal experiments, to be either carcinogens or noncarcinogens. About 90 percent of the reputed carcinogens turned out to be mutagenic and about the same proportion of the noncarcinogens were negative for mutagenicity. Other mutagenicity tests have since been devised with *E. coli* as the tester bacteria, and their efficiency is about as high as that of the original *Salmonella* test.

Two impressive accomplishments of mutagenicity tests can be mentioned to give an idea of their value. In Japan the chemical furyl furamide, known as AF-2, was added to a broad range of common food products for some years as an antibacterial agent. It had not shown any carcinogenic activity in standard tests on rats in 1962 or on mice in 1971. Then in 1973 T. Sugimura and his colleagues at the National Cancer Centre in Tokyo found that AF-2 was highly mutagenic in bacteria; they could easily demonstrate the mutagenic activity of the additive contained in just one slice of fish sausage! The discovery prompted a new round of more thorough animal tests for carcinogenicity, which showed that AF-2 was indeed a carcinogen. It was withdrawn from the market. If it were not for the bacterial test, AF-2—which had passed two approved animal tests and been declared negative for carcinogenicity—would presumably still be a component of fish sausage and other Japanese food products.

In 1975 Ames and his colleagues reported that 89 percent of the major class of hair dyes sold on the U.S. market contained mutagenic compounds. Since then the cosmetic industry has modified the composition of most hair dyes. It is estimated that several tens of millions of people dye their hair in the U.S. alone, which suggests that the discontinued components had presented a considerable risk.

In spite of all their advantages the mutagenicity tests have some technical and even theoretical limitations. Since mutations are revealed in the Ames assay by a restoration of enzyme activity, any mutation that does not happen to reconstruct the precise DNA sequence that codes for the histidine-making enzyme is not observed. To take one example, the antitumor drug bleomycin, which does its therapeutic work by damaging the DNA of tumor cells (as do about half of all antitumor agents), fails to induce the mutation that is scored in the *Salmonella* test. A false-negative re-

sponse of this kind is a technical problem that can be remedied by substituting other tester bacteria or a complementary short-term test.

False-positive responses are more significant from a theoretical point of view because they may raise some doubts about the validity of mutagenicity tests for identifying potential carcinogens. The point is that some chemical reactions with DNA are highly mutagenic in bacterial and mammalian cells without, as far as one can tell, being carcinogenic. Among these are the incorporation of an analogue of one of the nitrogenous bases and the methylation of certain sites on the bases. Such reactions cause negligible alterations in DNA structure that are not sensed in the cell as DNA damage; DNA replication proceeds on schedule and the new DNA carries a readable—although wrong—coded message. This process is termed direct mutagenesis, and in scoring for mutations it should be clearly distinguished from the more frequent indirect mutagenesis. In the latter process, described above, DNA damage brings about a transient arrest in replication; replication is resumed with the help of the RecA protein on a template that carries lesions, causing mutagenesis of the newly formed strand.

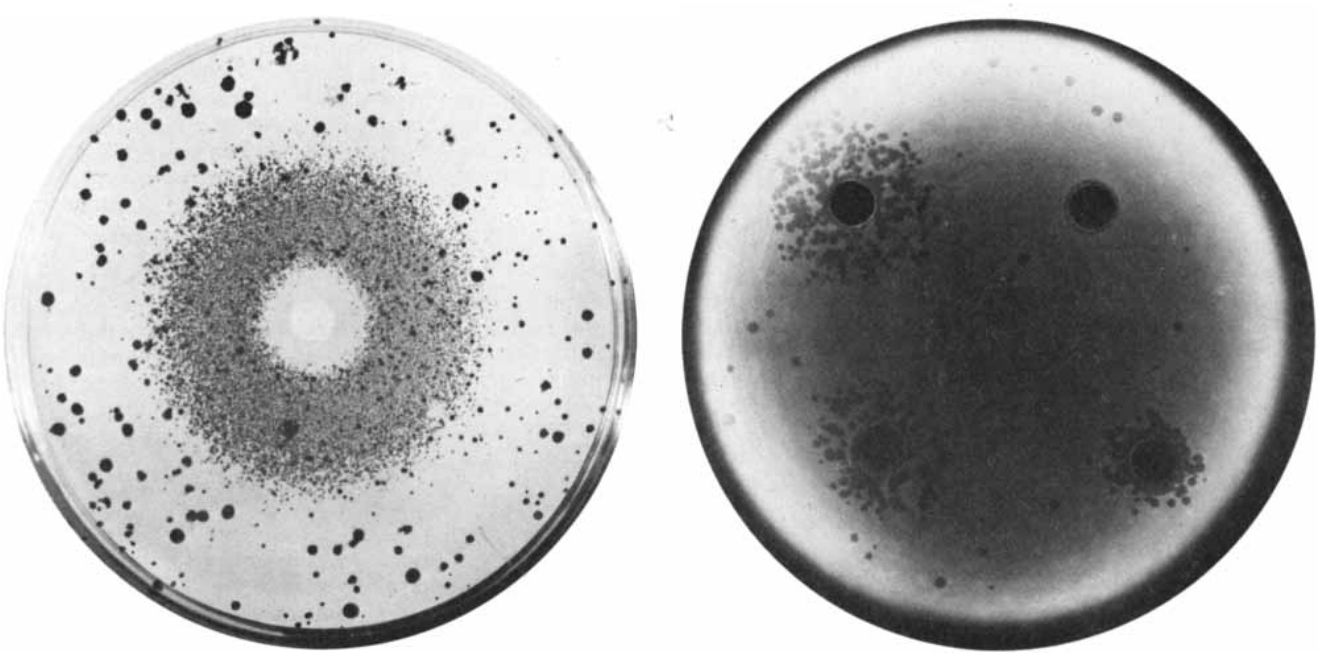
Since it is DNA damage that appears to initiate the cancerous transformation of a mammalian cell, a chemical is a potential carcinogen if it is a DNA-damaging agent, not simply because it causes mutations. A mutagen can have strong genetic effects on a biological population by altering the information encoded in the DNA of individuals' germ cells but nonetheless fail to cause cancer because it does no real DNA damage to the somatic cells: the cells of the rest of the body. Not all genetic toxic substances (chemicals that affect DNA) are potential carcinogens, whereas chemical carcinogens—because they damage DNA—are all indirect mutagens. In other words, mutagenic activity in bacteria is correlated with carcinogenesis in mammals primarily through indirect mutagenesis, which results from DNA damage.

In 1953, before the structure of DNA had even been established, André Lwoff of the Pasteur Institute in Paris anticipated that "inducible lysogenic bacteria might become a good test for carcinogenic and perhaps anticarcinogenic activity." A bacterium is said to be lysogenic when it carries, in a dormant state, the DNA of a "temperate" bacterial virus, which in this dormant state is called a prophage. One such temperate virus is phage lambda, which becomes a prophage when it is integrated into the DNA of certain lysogenic strains of *E. coli*.

When lysogenic *E. coli* bacteria are subjected to any treatment that halts DNA replication, the prophage is induced: its DNA loops out of the bacterial DNA and also directs the synthesis of proteins forming the virus particle, and a progeny of mature phage develops, bursting out of the host cell. The process is called lysogenic induction. Under normal conditions the dormant state of prophage lambda is maintained by a repressor, a protein that lies on the DNA's "operator" regions and blocks the operation of the lambda genes (except the gene that directs the synthesis of repressor to keep the prophage dormant). Jeffrey W. Roberts and Christine W. Roberts of Cornell University discovered that the induction of prophage lambda results from the cleavage of the lambda repressor; together with Nancy Craig they have recently shown that relatively pure repressor can be cleaved when it is mixed with an activated form of the RecA protein.

By triggering the activation of the RecA protein into the form that can cleave a viral repressor, DNA damage results in the induction of a dormant virus. The induction of prophage lambda can therefore serve as a test for DNA damage. Even before the mechanism of prophage development was understood at a molecular level, lysogenic induction was applied in the pharmaceutical industry to identify prospective antibiotics and antitumor drugs. When induction tests were done with ordinary strains of lysogenic *E. coli,* however, they did not give a positive response for such known carcinogens as benzo[*a*]pyrene.

Patrice Moreau, Adriana Bailone and



SPOT ASSAYS visualize the efficacy of bacterial tests qualitatively, although quantitative assays are preferable (*see illustration on page 41*). In the Ames test (*left*) the tested chemical was ethyl methane sulfonate, an alkylating agent and potent mutagen. A dense halo of revertant *S. typhimurium* colonies is seen around the disk from which the mutagen diffused. (Close to the disk there is a zone in which a toxic concentration of the chemical killed all bacteria.) The larger colonies are mutant bacteria that arose spontaneously without exposure to the chemical. In the inductest (*right*) the tested chemical was aflatoxin B_1, a potent carcinogen of the liver. The aflatoxin doses placed on the four test disks were (*clockwise from top right*) 0, 20, 200 and 2,000 nanograms. In the background a few of the lysogenic *E. coli* tester bacteria have spontaneously produced mature phage that have killed the nearby indicator bacteria, producing plaques at random locations.

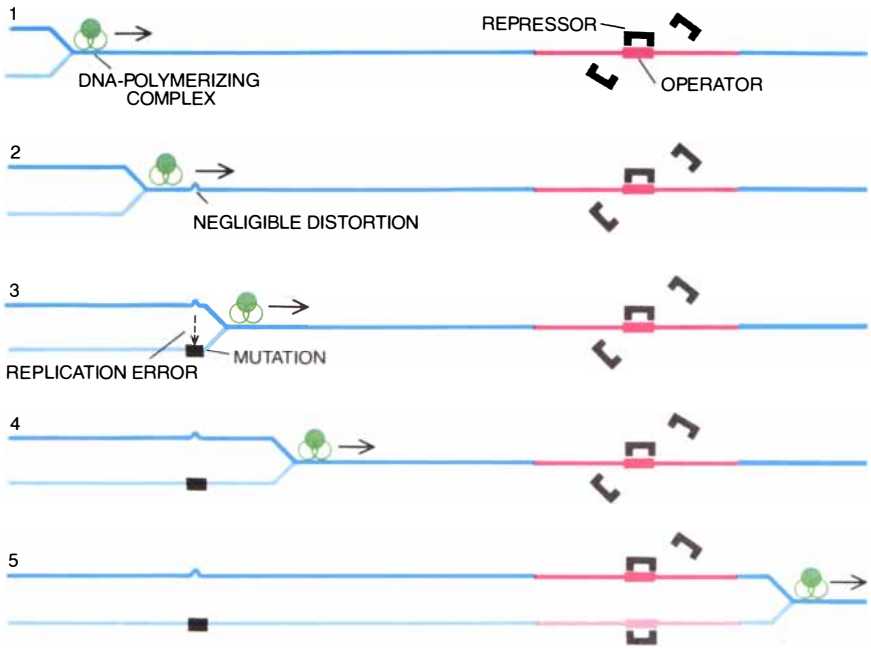
I therefore set out five years ago to renovate the lambda-induction test to make it capable of identifying all kinds of DNA-damaging agents. We reasoned that because the molecular mechanism of lysogenic induction was understood better than that of mutagenesis, a prophage-induction test should provide more insights than mutagenicity tests could into the precise effect of chemical carcinogens on DNA and should provide a supplementary tool for screening as well.

Following Ames's lead, we constructed new tester bacteria that are permeable to chemicals and deficient in excision-repair enzymes, and we assayed potential carcinogens in the presence of a rat-liver metabolizing mixture. The prophage-induction test, or inductest, has turned out to discriminate effectively between carcinogens and noncarcinogens in our laboratory and in several others.

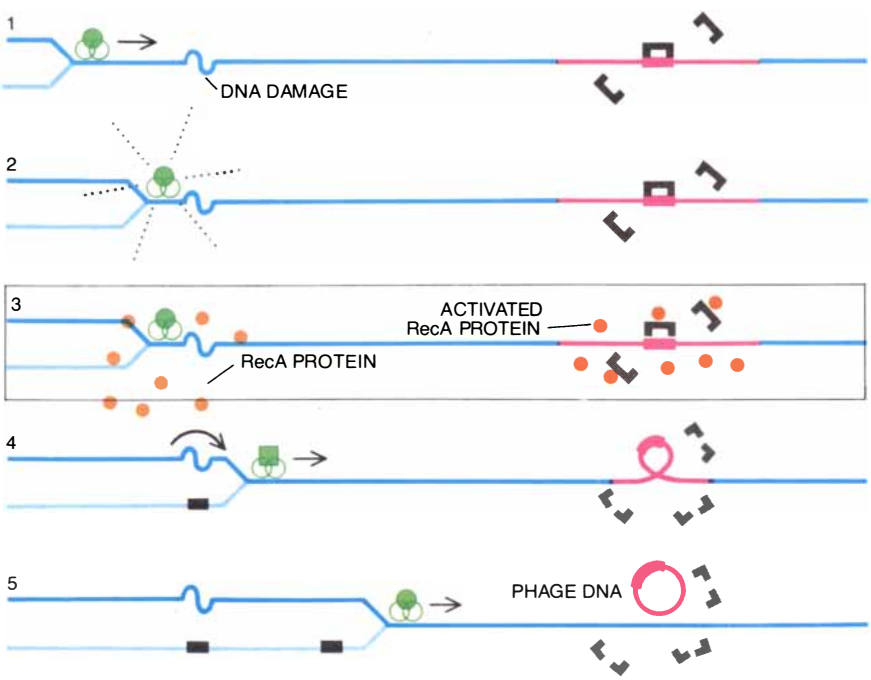
The inductest has certain advantages over mutagenicity tests. Mutations can be scored only in bacteria that survive exposure to a chemical treatment that is often toxic as well as mutagenic, and no more than one bacterium out of 1,000 is likely to be detected as a histidine revertant. In contrast, lysogenic induction can be observed, when it takes place, in the bulk of a cell population. Prophage induction is a mass effect, largely independent of whether or not cells would survive the chemical treatment; a cell that is induced to produce a progeny of phage would die in any case. The inductest can therefore continue to give a positive response for a highly toxic potential carcinogen at dose levels that would kill the tester bacteria in a mutagenicity test.

The fact that in the inductest most cells undergo a dramatic change led us to design a biochemical assay of lysogenic induction. Sankar Adhya, Maxwell E. Gottesman and Asis Das of the National Cancer Institute in the U.S. constructed a bacterial strain in which the gene for the enzyme galactokinase is hooked up to the lambda DNA in such a way that the lambda repressor blocks the synthesis of the enzyme. Alain Levine, Moreau, Steven Sedgwick and I demonstrated that when a DNA-damaging chemical activates the RecA protein in this strain, the repressor is cleaved and galactokinase is synthesized, and the amount of enzyme activity that can be detected reveals the extent of DNA damage. This biochemical assay may be the shortest of the short-term tests; it takes only half a day to test for a potential carcinogen.

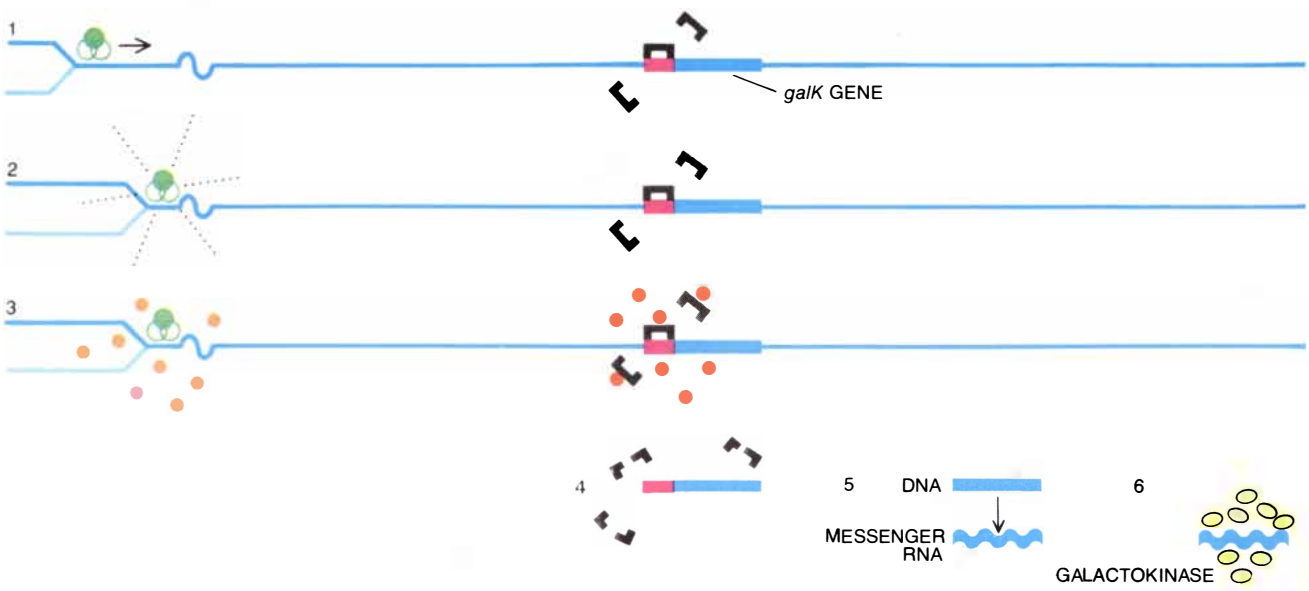
Moreau and I have also constructed strains, containing a new form of prophage lambda, in which one can identify chemicals that induce mutagenesis without damaging DNA. This lambda mu-



DIRECT MUTATION is a result of DNA replication on a minor distortion of a DNA strand. This diagram and the one below show the replication of only one strand of bacterial DNA (*blue*) carrying a piece of viral DNA, prophage lambda (*red*), which is kept in a dormant state by a repressor protein that blocks an operator region. The DNA strand is replicated by a polymerizing complex (*1*). A negligible distortion of the bacterial DNA (*2*) does not obstruct replication, but it does give rise to an error in replication, so that the newly formed DNA strand (*light blue*) carries a mutation (*3*). The negligible distortion is not sensed as DNA damage, RecA protein is not induced, repressor stays in place and prophage is replicated in dormant state (*4, 5*).



INDIRECT MUTATION results from the events that follow DNA damage (*1*), which blocks DNA replication and produces an "SOS" signal (*2*). A large amount of RecA protein (*orange*) is synthesized (*3*). The RecA somehow facilitates DNA replication on a damaged template, apparently by temporarily modifying the polymerizing complex (*4*). This replication is highly mutagenic not only at the lesion site but also elsewhere (*5*). Activated RecA cleaves the repressor (*4, right*), freeing the operator region of the prophage DNA. The prophage is thereby induced to form phage DNA (*5*), which in turn forms virus particles that burst out of bacterium.



ENZYME-INDUCTION TEST is a biochemical counterpart of the inductest. The gene, *galK*, for synthesis of galactokinase is hooked up to prophage-lambda DNA in such a way that enzyme synthesis is blocked by lambda repressor. When DNA is damaged, activated RecA protein cleaves repressor. Liberated *galK* DNA (*4*) is transcribed into messenger RNA (*5*), which is translated into galactokinase (*6*).

tatest, as we call it, reveals whether a chemical gives rise to mutations directly (that is, without causing DNA damage and therefore without inducing prophage) or indirectly (by damaging the DNA and producing active virus).

Although neither the inductest nor the lambda mutatest has yet been validated as extensively as the Ames mutagenicity test, the two appear to be valuable complementary assays, in particular for determining the nature of DNA damage and for establishing a correlation between potency in producing DNA damage in bacteria and potency in causing cancer in mammals. Perhaps more important in the long run, a biological system comprising both a cell and a dormant virus should be a valuable tool for analyzing the pathway of biochemical changes, genetic or nongenetic (cleavage of repressor, for example), that are triggered when cells are exposed to carcinogens.

Given that man is exposed to a flood of chemicals, some of which surely increase the risk of cancer, there is an urgent need to identify potential carcinogens and lower human exposure to them to a level of negligible risk and keep it there. In such an effort bacterial tests can play a major role. In order to ensure broad acceptance of uniform standards, some investigators have strongly favored the adoption of a single, standardized mutagenicity test. Often, however, it is advisable to adapt a screening effort to the particular physical and chemical properties of the substance that is under study. It would therefore seem to be sensible to subject each chemical to a battery of tests. Any test has its foibles, which give rise to false-negative or false-positive responses. Several independent and complementary determinations of a chemical's DNA-damaging capacity can provide a balanced and hence more compelling assessment of carcinogenic risk.

The pronouncement that a particular chemical is a potential carcinogen has obvious social, economic and even political implications. The chemical may present a risk to a very large population, as was the case with the food additive AF-2 in Japan. That situation was dealt with rather easily: AF-2 could be replaced by an innocuous substitute. It was dispensable, and so it was simply banned from the market. Some chemical carcinogens cannot easily be dispensed with and cannot be banned; particular problems are raised by substances as different as cigarette smoke, motor-vehicle exhaust fumes and antitumor drugs. For most indispensable carcinogens it is clearly necessary to reduce human exposure to the safest levels possible. That calls for the adoption of a broadly accepted set of safety standards to be implemented by stringent regulations.

We stand today, with regard to chemical carcinogens, about where we were with regard to ionizing radiations three decades ago. The advent of nuclear power reactors made it necessary to protect workers in the industry from direct sources of radiation and to protect everyone from exposure to radioactive effluents and solid wastes. An international agency sponsored by the United Nations worked out a set of standards and regulations designed to minimize human exposure to various sources of radiation. The regulations were broadly accepted and implemented, and they have stood the test of time. They should provide a valuable reference for evolving standards for chemical carcinogens.

The experts who set the standards for radiation safety have proceeded from two basic facts. One is that mankind is exposed to a natural background of radiations whose biological effect is not perceptible (even though in principle there is no threshold for the effects of ionizing radiations). The other is that biological damage is proportional to the energy released within the living system being considered, and that this linear relation is more or less independent of the particular source of radiation. The experts assumed that for the known sources an absorbed dose only a few times higher than the dose delivered by background radiation, and less than the level that would cause a doubling of the mutation rate in a mammalian population, would constitute a negligible risk; in general such doses were adopted as being permissible for human exposure. In the case of newly devised or newly discovered radioactive materials for which there were no accumulated experimental data, permissible values were calculated on the basis of the linear relation between biological damage and the energy released within the organism.

Since chemical pollution is worldwide, an international committee of experts should propose protective regulations for adoption by individual governments. Some principles should gain broad acceptance rather easily: for example, all unnecessary potential carcinogens should be banned. A list of carcinogens that cannot be dispensed with but to which exposure should be strictly limited will also have to be drawn up. It may be difficult, however, to reach a consensus on such a list and on the permissible environmental concentrations

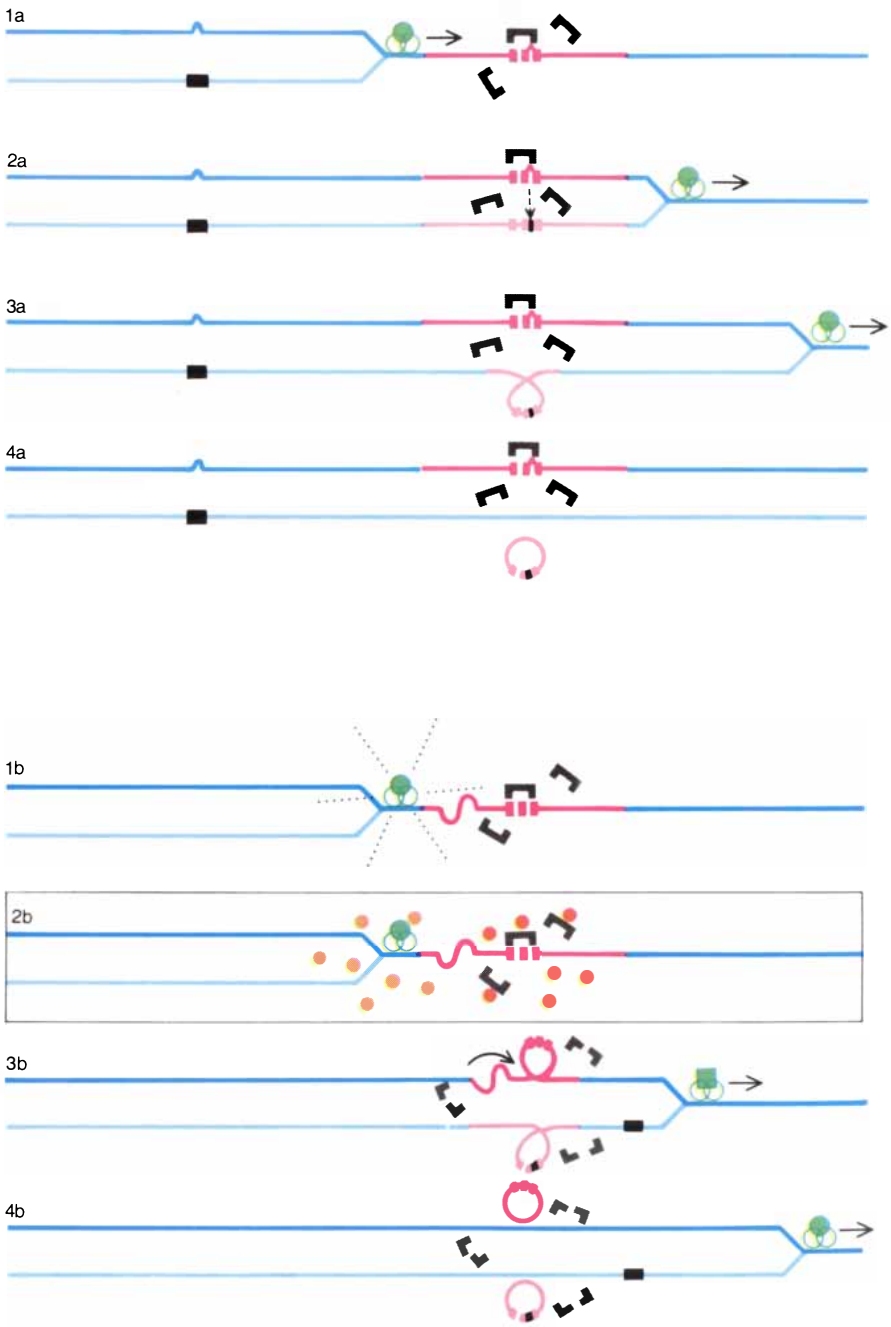
of each chemical. One would like to begin by defining permissible concentrations of indispensable carcinogens in relation to a "natural" background. The trouble is that cultural differences, social habits and other factors have considerably distorted notions of what is a natural background level of negligible risk. (For geographical, economic and religious reasons the carcinogenic background is surely lower in Salt Lake City than it is in New York.)

If a negligible-risk level can be agreed on, the next step would be to estimate, for each indispensable carcinogen, the concentrations that should be allowed for workers in the industry and (at a much lower level) for the general population. Can bacterial tests help in setting such levels? Yes, if there is proportionality between potency in causing DNA damage in bacteria and potency in causing cancer in mammals. Matthew S. Meselson and Kenneth Russell of Harvard University have attempted to demonstrate a direct relation between mutagenic potency in the *Salmonella* test and carcinogenic potency in laboratory animals. For 10 of the 14 known chemical carcinogens they considered there was a correlation.

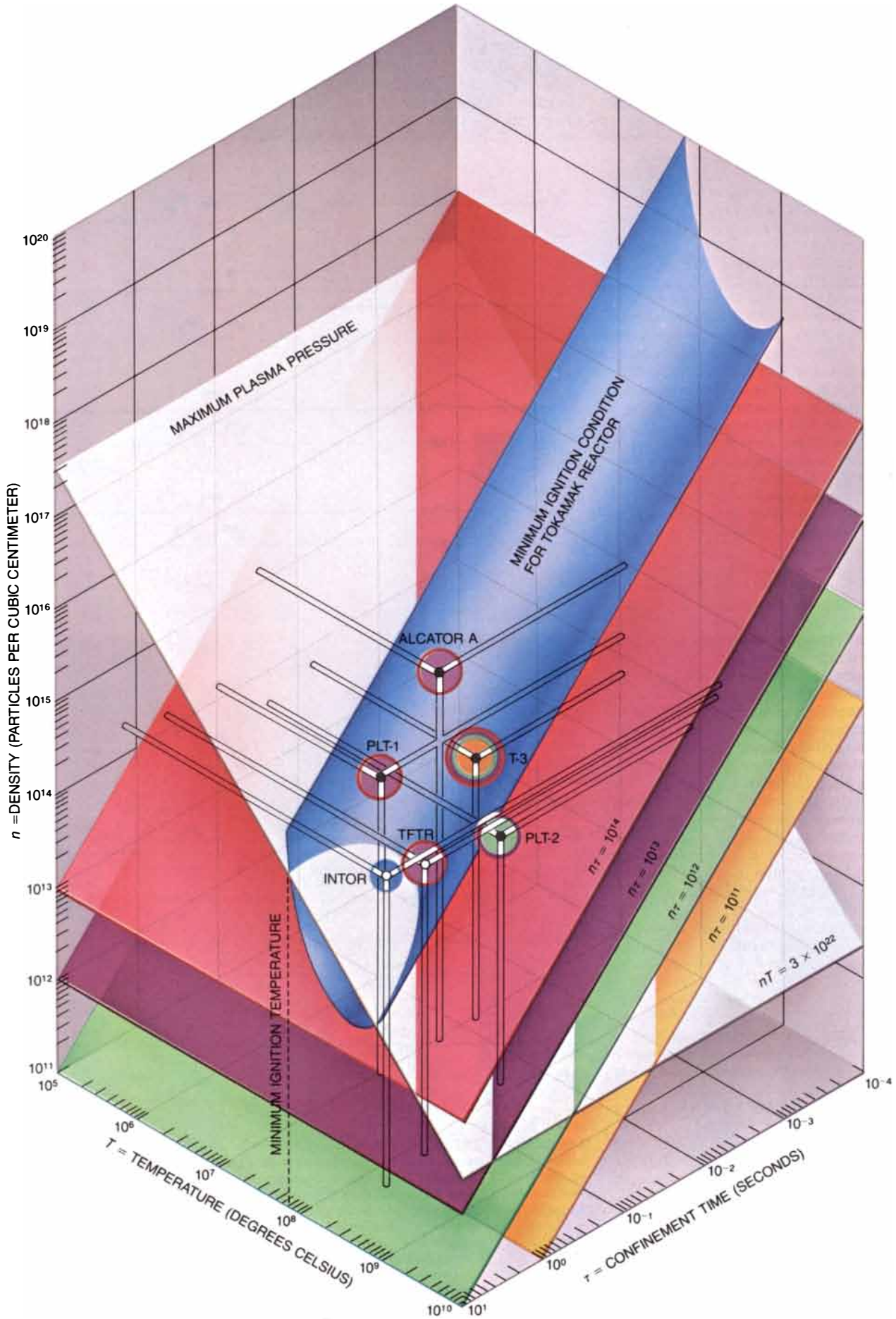
The many exceptions to the rule of correlation were to be expected because in a mammal the extent of DNA damage in a target cell depends on factors (which may be different in bacteria) such as the specific metabolism of carcinogens in the organ involved, the cells' permeability to particular metabolites and tendency to accumulate them and the extent of residual DNA damage after the cellular repair mechanisms have had their effect. Each of these variables must be determined for each individual carcinogen if one is to set valid levels for human exposure. Each chemical—in contrast to individual sources of radiation—requires a thorough specific study to determine its carcinogenic potency. Bacterial tests have good indicative value.

Even though establishing regulations will be much more difficult for chemical carcinogens than it was for radiations, the regulations should be established soon. It is wiser and safer to have safety regulations than not to have them. A large part of the human population contracts cancer; one person in five dies of it. It is important to determine the precise mechanism by which each DNA-damaging agent acts on DNA, not only to prevent cancer but also to develop better ways of curing it. About half of all the drugs administered in an effort to arrest cancer are DNA-damaging agents, and so a significant number of people are being regularly exposed, for therapeutic purposes, to such agents. By establishing, through studies in bacteria, the mechanism whereby antitumor agents affect DNA, investigators may be able to help design more drugs that are effective in curing cancer without also causing cancer.

Bacterial tests for potential carcinogens should help to answer two questions now being asked by industry, government and people at large. What chemicals are DNA-damaging agents? How much of each—in our air, our water and our food, in the products we manufacture and consume—constitutes a biological risk? These two questions will be answered satisfactorily, however, only if a third question is asked: What is the mechanism whereby each chemical affects DNA? Only basic knowledge can breed safety.



DIRECT AND INDIRECT MUTAGENESIS can be distinguished from each other by the mutatest, based on mechanisms diagrammed here. The repressor binds to a slightly modified operator (*1a*, *1b*). A negligible alteration of operator DNA (*1a*) results, on replication, in a mutation (*2a*) that prevents further binding of the repressor (*3a*). The prophage develops into active phage particles (*4a*) that are selectively revealed by strain-*B* indicator bacteria, which are not sensitive to nonmutated phage; phage development independent of RecA-protein activation reveals direct mutagenesis. DNA damage (*1b*), on the other hand, leads on replication (*2b*, *3b*) to RecA-dependent indirect mutagenesis, which is detected with strain-*B* indicator bacteria. The concomitant event, prophage induction (*3b*, *4b*) resulting from activation of RecA, is detected with indicator bacteria of strain *A*, not with strain-*B* bacteria. Mutagenesis can be diagnosed as indirect when it occurs along with prophage induction: both are dependent on RecA.



Progress toward a Tokamak Fusion Reactor

Recent efforts to confine a superhot hydrogen plasma in a toroidal "magnetic bottle" have yielded encouraging results. The feasibility of this scheme for a fusion reactor should be known in a few years

by Harold P. Furth

The vision of a power-generating fusion reactor based on the magnetic confinement of a plasma (a hot gas of charged particles) inside a toroidal vacuum chamber emerged more than a quarter of a century ago. That vision is still very much in view, but it has been colored by a good deal of arduous experience. In retrospect it is clear that although the key theoretical aspects of the problem were fairly well understood at the outset, the experimental difficulties were greatly underestimated. By the mid-1960's, after 15 years of intensive research in several countries, there was still no firm experimental basis for the extrapolation of any magnetic-confinement scheme to the plasma conditions regarded as being necessary for a practical fusion reactor. On the contrary, there were reasonable grounds for concluding that no net output of power from controlled fusion reactions was likely to be achieved in the laboratory during the 20th century.

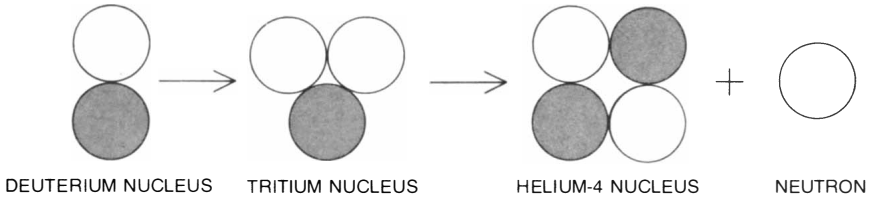
The short-term prospects of fusion research improved dramatically toward the end of the 1960's, when investigators at the I. V. Kurchatov Institute of Atomic Energy in Moscow succeeded in getting impressive results in experiments conducted in a comparatively simple toroidal magnetic-confinement device, which they called a tokamak. Plasma temperatures within an order of magnitude of those required for a toroidal fusion reactor were achieved for the first time in the Russians' Model T-3 tokamak, along with plasma-confinement properties that were somewhat mysterious but were at least mysteriously favorable. On the basis of the improved outlook the world fusion-research effort was stepped up significantly, and a major part of the new work was committed to the exploration of the tokamak scheme [see "The Tokamak Approach in Fusion Research," by Bruno Coppi and Jan Rem; SCIENTIFIC AMERICAN, July, 1972].

HISTORIC MILESTONES on the road to a practical fusion-power reactor based on the tokamak approach can be traced with the aid of the three-dimensional graph on the opposite page. The black dots stand for the best experimental-tokamak performances to date, plotted in terms of the three basic parameters of research on magnetically confined plasmas: temperature (T), density (n) and confinement time (τ). The earliest major achievement represented here, labeled T-3, was recorded in 1969 in an experiment conducted in the Model T-3 tokamak at the I. V. Kurchatov Institute of Atomic Energy in Moscow. The Alcator value, the high point of a series of experiments carried out over the past few years at the Massachusetts Institute of Technology, indicates the best confinement conditions attained so far in a tokamak that depends solely on ohmic, or resistive, heating of the plasma. The best-performance record of the Princeton Large Torus, or PLT, an intermediate-scale tokamak put into operation in 1976 at Princeton University's Plasma Physics Laboratory, is given twice, once for a test involving ohmic heating alone (PLT-1) and again for a test involving ohmic heating plus auxiliary neutral-beam heating (PLT-2). The white dots are estimates of the performance characteristics expected for two future experimental devices, one of which, the Tokamak Fusion Test Reactor (TFTR), is currently under construction at Princeton. The other white dot represents the projected performance of the International Tokamak Reactor (INTOR), an even larger device now in the preliminary design stage. The stack of parallel colored planes denotes successive order-of-magnitude increments of the confinement parameter, a measure of the quality of plasma confinement expressed as the product of the density in particles per square centimeter times the confinement time in seconds (or $n\tau$). The curved blue surface on top of the stack is a plot of the minimum $n\tau$ value that must be surpassed to reach the ignition, or self-heating, regime in a plasma consisting of deuterium and tritium: the two heavy isotopes of hydrogen. The white plane cutting across all the other surfaces is the maximum plasma pressure (nT) that can be maintained in a tokamak reactor. The INTOR device is expected to operate in the space between the white plane and the blue curved surface and so to qualify as the first tokamak ignition reactor.

During the past decade of research on tokamaks of various types in the U.S.S.R., the U.S., Europe, Japan and elsewhere the performance parameters of these experimental devices have advanced steadily, but serious doubts have remained about the confinement properties that could be expected under conditions resembling those of an actual reactor plasma. This issue was put to a crucial test only recently when an experiment in a scaled-up tokamak at Princeton University's Plasma Physics Laboratory, called the Princeton Large Torus (PLT), achieved plasma conditions close to those needed for a fusion reactor. To the surprise and delight of the experimenters no apparent deterioration of confinement was observed.

Plasma confinement, to be sure, is not the only scientific issue that needs to be resolved before an economic tokamak fusion reactor can be built. At this point, however, one can say for the first time in the history of fusion research that there is a substantial experimental basis for the detailed physical description of such a reactor. The actual demonstration that a reactor of this type is scientifically feasible awaits the completion, in the early 1980's, of a new generation of very large tokamak facilities, three of which are currently under construction, one in the U.S., one in Europe and one in Japan. Beyond that critical step technological challenges of the first magnitude must still be overcome before a practical power-generating fusion reactor can be put into operation.

The easiest fusion reaction to achieve in the laboratory brings together the nuclei of the two heavy isotopes of hydrogen: deuterium and tritium. Before these two positively charged nuclei can fuse to become a helium nucleus the electrostatic Coulomb repulsion between them must be overcome; hence an initial investment must be made in the nuclei's relative kinetic energy, or energy of motion. In the case of a deuterium-



FUSION REACTION between a deuterium nucleus and a tritium nucleus yields a helium nucleus (an alpha particle), a neutron and 17.6 million electron volts (17.6 MeV) of energy. In order for the two colliding nuclei to fuse they must be moving with a relative kinetic energy of at least 10,000 electron volts (10 keV). A plasma consisting of charged particles moving with a mean random energy of 15 keV per particle has an equivalent temperature of 10 keV, or approximately 116 million degrees Celsius. The gray balls are protons, the white balls neutrons.

tritium fusion reaction the minimum investment is about 10,000 electron volts (10 keV). From the point of view of the theorist dealing with the overall energetics of the fusion process this investment is quite modest, since the energy released by each fusion event is 17.6 million electron volts (17.6 MeV). From the point of view of the experimenter, however, the requirement is formidable: a mean random energy of 15 keV per nucleus is equivalent to a fuel temperature of 10 keV, or about 116 million degrees Celsius.

The confinement of matter at such high temperatures is facilitated by the dissociation of atoms into their electrically charged constituents—electrons and ions—at temperatures higher than 10,000 degrees C. Because charged particles tend to gyrate along magnetic-field lines the resulting plasma can be confined in a suitably shaped "magnetic bottle." The toroidal magnetic bottle has the special advantage that charged particles cannot escape from it simply by moving along the lines of force in the magnetic field. The principal alternative approach to a fusion reactor based on the magnetic confinement of a hot plasma is the mirror machine, an open-ended magnetic bottle. In addition there are a variety of quite different "inertial confinement" schemes, in which intense beams of laser light or accelerated particles implode a small pellet of fuel. The fuel then expands freely and its heavy hydrogen nuclei fuse.

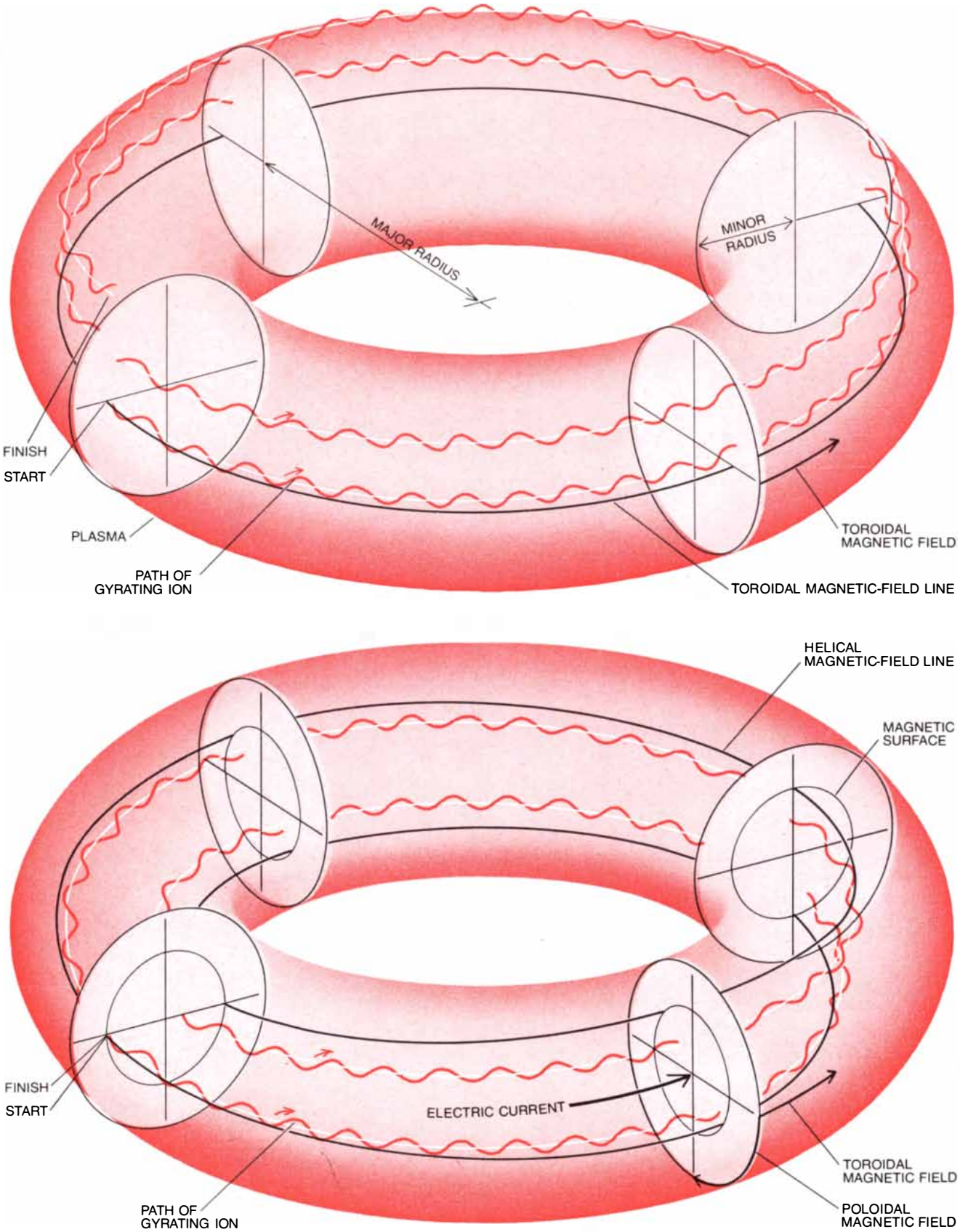
The most important figure of merit for an experimental fusion reactor is the ratio of the output power derived from the fusion reactions to the input power required to heat the plasma. This ratio, called the energy-multiplication factor, or the Q value, depends on the fraction of the hot nuclei that are able to fuse during the time it would take for the plasma to lose its energy. Since fusion is a binary process, the Q value turns out to depend on the product of the energy-confinement time of the plasma (measured in seconds) and the density of the plasma (measured in particles per cubic centimeter). For example, in a 10-keV deuterium-tritium plasma with a normal distribution of particle energies the requirement for "break even" (a Q value equal to 1) means that the product of the confinement time and the density, a measure known as the confinement parameter, should exceed 6×10^{13}. The Q value of a plasma increases rapidly after the confinement parameter exceeds the break-even threshold, because some 20 percent of the energy released in fusion reactions involving deuterium and tritium goes into the production of energetic helium-4 nuclei (alpha particles), which can be slowed down and retained in the bulk plasma, thereby amplifying the input power available for heating. As the confinement parameter approaches 3×10^{14}, the fusion reactions become able to maintain the temperature of the plasma without any input of heating power. At that point the Q value becomes infinite, and the fuel can be said to "ignite" in the conventional sense.

In striving to reach the prescribed range of reactor-quality confinement, fusion investigators have encountered four main energy-loss processes that must be kept under control: (1) macroscopic instabilities, that is, spontaneous magnetohydrodynamic deformations of the confining field that cause the plasma to leak abruptly out of the magnetic bottle; (2) microscopic instabilities, in effect tiny stepwise leaks that allow the plasma particles (and the heat) to diffuse gradually out of the plasma across the magnetic-field lines; (3) Coulomb scattering, which disrupts the orbits of the gyrating particles and gives rise to an irreducible rate of diffusion, and (4) radiative cooling of the plasma, mainly in the form of ultraviolet radiation from impurity ions. Progress in research on toroidal fusion devices has been paced by a gradual improvement in the understanding of these four energy-loss processes and by the development of effective techniques for minimizing them.

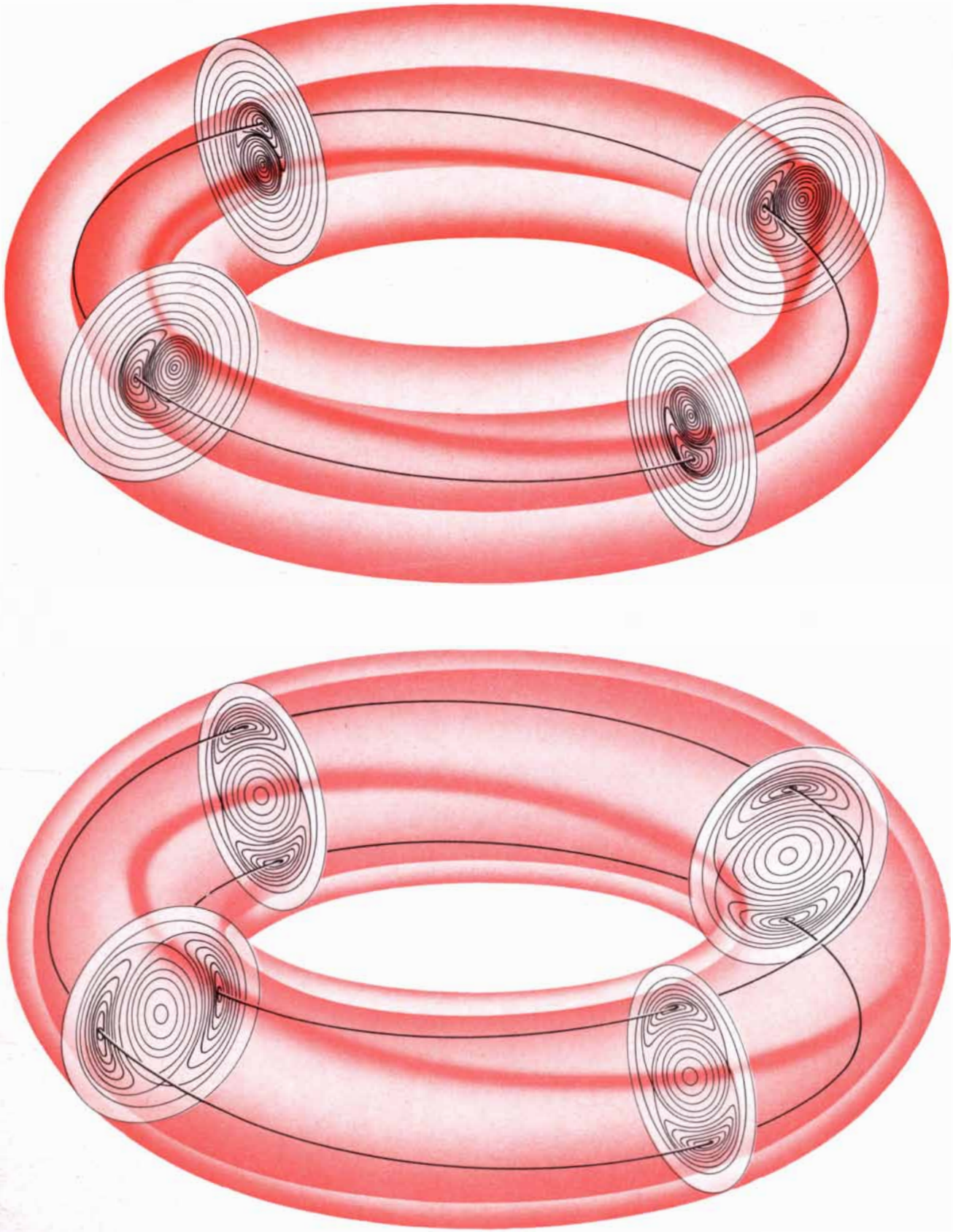
If a plasma is to be confined in the shape of a torus, the magnetic bottle confining it cannot consist only of a toroidal magnetic field; in that case the particle orbits would not close on themselves, and the particles would simply drift out of the bottle across the magnetic-field lines. There are several possible solutions to the problem, including the stellarator approach, which was developed originally by Lyman Spitzer, Jr., of Princeton. By far the simplest solution, however, is provided by the tokamak approach. In this scheme, which was proposed independently in the early 1950's by Igor E. Tamm and Andrei D. Sakharov in Moscow and by Spitzer in Princeton, an electric current is induced along the plasma column, flowing the long way around the torus. The associated poloidal magnetic field (which has lines of force going around the minor cross section of the torus) gives rise to helical field lines, which form a set of nested magnetic surfaces. In this configuration the charged particles have closed orbits that do not depart from a given magnetic surface by more than their radius of gyration in the poloidal field.

This localization of the particle orbits is important in defining the irreducible rate of particle diffusion resulting from Coulomb scattering, which can cause the particles to skip through the width of an orbit from one magnetic surface to another. For ions in the 10-keV range the probability that two nuclei will fuse rather than simply scatter off each other is very small (about one part in 10 million). In order to minimize the importance of this particular heat-diffusion process as an obstacle to the achievement of net fusion power the minor radius of the plasma torus must therefore be many times greater than the orbital diameter of the particles. Because the orbital diameter decreases with increasing poloidal-field strength the criterion for ignition (the achievement of a confinement parameter greater than 3×10^{14}) in a tokamak turns out to be roughly equivalent to the requirement that the toroidal current should exceed four million amperes. As it happens, the same minimum current of about four million amperes is needed to localize the orbits of the energetic alpha particles in the plasma column, a precondition for achieving ignition.

Although in a given tokamak the microscopic confinement of the plasma particles improves at higher plasma currents, the macroscopic stability of the plasma becomes increasingly precarious. The plasma can suddenly deform itself into a number of more or less helical twisted shapes called kink modes, which are a little like the modes of mechanical buckling in the components of a structure under extreme stress. The unstable kinks in the tokamak plasma tend to align themselves approximately with the pitch of the helical magnetic-field lines. As a result there is a simple relation between the "safety factor" (a measure of the helicity of the field lines) and the shape of the plasma's characteristic kink modes. For example, a safety factor of 1 lends itself to a kink that goes once around the torus in the toroidal direction every time it goes once around it in the poloidal direction, corresponding to a simple sideways displacement of the plasma in the minor cross section of



TOROIDAL PLASMA confined in a suitably designed magnetic bottle has the special advantage that the charged particles cannot escape from it simply by moving along the magnetic-field lines. A magnetic configuration of this type, however, cannot consist only of a toroidal magnetic field, since in that case the particle orbits would not close on themselves and the particles would simply drift out of the plasma across the magnetic-field lines (*diagram at top*). The tokamak solution to the problem is to set up a poloidal magnetic field around the minor cross section of the plasma torus by means of an electric current induced along the plasma column. In combination with an externally applied toroidal magnetic field this arrangement results in helical field lines, which form a set of nested magnetic surfaces (*diagram at bottom*). A particle traveling around the torus in this configuration does not deviate significantly from a given magnetic surface.



TWO TROUBLESOME KINKS can appear suddenly in a tokamak plasma at high current densities, threatening the macroscopic stability of the plasma. Since these magnetohydrodynamic deformations are aligned approximately with the pitch of the helical magnetic-field lines, there is a simple relation between the helicity of the field lines and the shape of the plasma's characteristic kink modes. The diagram at the top shows a sideways kink, which goes once around the torus in the toroidal direction every time it goes once around it in the poloidal direction. The diagram at the bottom shows an elliptical kink, which goes twice around the torus in the toroidal direction every time it goes once around it in the poloidal direction. Higher-order kink modes are possible, but the plasma tends to be more resistant to such complex deformations. The sideways kink plays an important role in maintaining size of current channel at center of a tokamak plasma.

the torus (mode No. 1); a safety factor of 2 lends itself to a kink that goes twice around the torus in the toroidal direction every time it goes once around it in the poloidal direction, corresponding to an elliptical deformation (mode No. 2); a safety factor of 3 lends itself to a kink that goes three times around the torus in the toroidal direction every time it goes once around it in the poloidal direction, corresponding to a triangular deformation (mode No. 3), and so on.

For typical tokamak currents one can show theoretically that the higher the safety factor (that is, the shallower the pitch of the helical field lines), the higher the mode number of the possible instabilities and the more resistant the plasma to deformation. It follows that the smaller the current in the plasma, the greater the macroscopic stability of the plasma. The experimental behavior of the plasma column in a variety of tokamaks has been found to be in close agreement with these and more detailed predictions of magnetohydrodynamic theory.

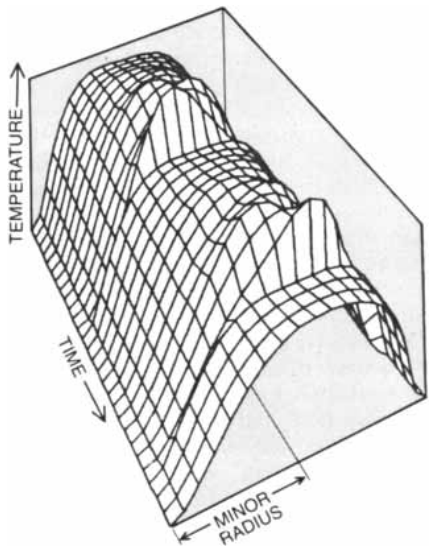
In short, the practical objective in tokamak research is to make the helicity of the field lines as steep as possible, consistent with good macroscopic stability, in order to have as large a plasma current as possible for good particle-orbit confinement. It follows that one must learn to deal with the two kink modes that have the lowest mode numbers: the sideways deformation and the elliptical one. A field configuration that is stable at a suitably low safety factor is not easy to design theoretically. The tokamak plasma, however, arrives automatically at a working solution. Since the hottest part of the current channel is the most conductive, the current density rises slowly near the magnetic axis of the torus until the local value of the safety factor drops below 1; as a result a sideways instability sets in and broadens the current channel again. This periodic readjustment mechanism is localized within the central region of the plasma and need not cause any net loss of hot plasma to the wall of the vacuum chamber. The existence of a central region with a safety factor smaller than 1 has the great advantage that for this type of magnetic-field configuration the more troublesome outward-localized elliptical kink mode cannot grow at all, even though the safety factor equals 2 at larger radii. The result is effectively stable overall plasma confinement. It is precisely this felicitous self-control mechanism of the tokamak plasma that is responsible, in a kind of Darwinian sense, for the emergence of the tokamak as the preeminent toroidal magnetic bottle.

An additional benefit of the current that flows through the tokamak plasma is that it provides a way of giving energy to the plasma electrons by conventional ohmic, or resistive, heating. In this way electron temperatures in the kilovolt range are generated spontaneously in tokamak discharges. The electrons in turn collide with the ions and eventually reach a state of thermal equilibrium with them, to an extent that is determined by the electron temperature and the confinement parameter. In plasmas with a low confinement parameter the temperature of the ions is typically much lower than the temperature of the electrons, but as the confinement parameter approaches the break-even threshold the ion temperature and the electron temperature tend to become roughly equal.

The major tokamak breakthrough of a decade ago was accomplished by means of ohmic heating in the T-3 experiment, which had been evolving at the Kurchatov Institute since 1962 under the leadership of the late Lev A. Artsimovich. The T-3 device had a minor radius of 15 centimeters, a major radius of 100 centimeters, a toroidal-field strength of 35,000 gauss and a typical current of 100,000 amperes. In the process of ohmic heating the temperature at the center of the plasma column rose above 1 keV for the first time in research on toroidal fusion devices, and the ion temperature reached about .5 keV. The energy-confinement time was between five and 10 milliseconds at a central density approaching 10^{14} particles per cubic centimeter. Hence the confinement parameter for the T-3 was close to 10^{12}—still far from the reactor range but much higher than it had been in any previous fusion experiment.

Prior to the T-3 experiment typical confinement parameters for toroidal devices had been on the order of 10^{11} at electron temperatures of about 100 electron volts. Moreover, the observed confinement times had borne no relation to the advantageous type of confinement the toroidal geometry was meant to provide but rather had appeared to follow anomalous scaling laws that were incompatible with the attainment of fusion conditions in a reactor that would not be too big. The suspected reason for the poor plasma confinement, which got worse with rising temperature, was the presence of various microinstabilities, which in some cases could be observed directly. The only comforting element in the situation came from the prevailing microinstability theory, which predicted that such large anomalous plasma-transport rates would not be able to maintain themselves at still higher temperatures.

The new T-3 data vindicated this expectation. For the first time in such experiments gross energy-confinement times turned out to be reasonably close to the values predicted by the Coulomb-scattering theory of diffusion. The plasma ions actually appeared to follow the theoretical prediction quite closely, and only the electrons were found to engage



X-RAY OBSERVATIONS of an actual kink instability at work in a tokamak plasma were used to construct this computer-generated three-dimensional display. The temperature profiles in this case appear to wobble as they build up to a peak before dropping precipitously. The X-ray data were obtained during a test run of the PLT device. The experiment and the computer processing of the results were done by Ned R. Sauthoff of Princeton.

in a residual form of anomalous transport, with scaling laws that remained somewhat uncertain but that appeared to be compatible with a reasonably small fusion reactor.

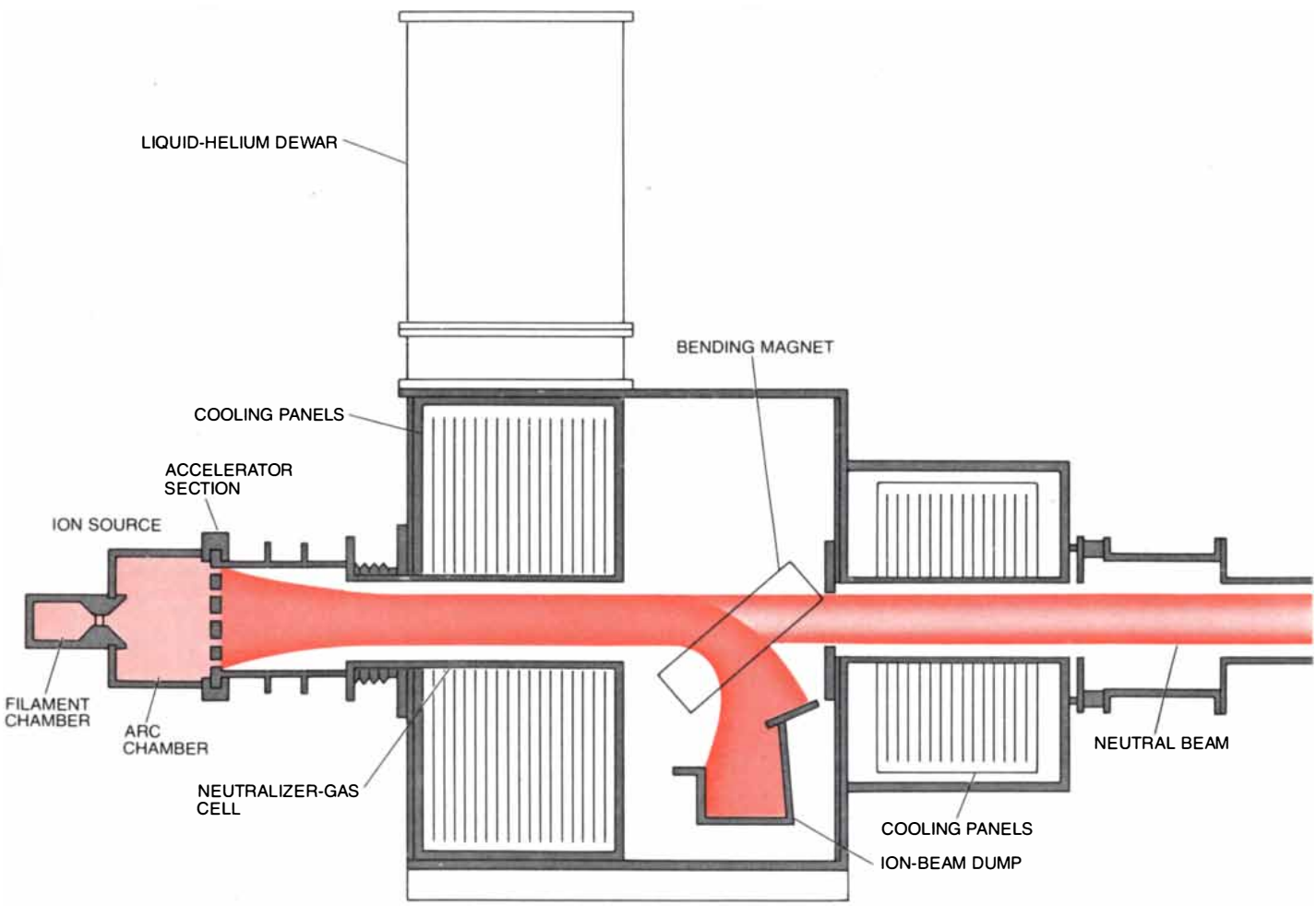
To progress from the T-3 results to actual reactor conditions called for two further improvements. The confinement parameter had to be enhanced by about two more orders of magnitude, and the temperature had to be raised by one order of magnitude. The most obvious way to improve the confinement was to proceed to a larger plasma, since for diffusive heat-loss processes the energy-confinement time scales upward as the square of the minor radius of the plasma. In a tokamak fusion reactor the plasma is typically envisioned as having a minor radius of a meter or two, which is about 10 times larger than the minor radius of the T-3 plasma. It follows that the confinement time in a tokamak reactor would be expected to be 100 times longer than it was in the T-3 experiment, provided the rate of diffusion did not increase at higher temperatures. An alternative way to improve the confinement would be to increase the strength of the magnetic field, but here the options are more limited because the maximum field strength compatible with superconducting magnet coils is probably no greater than 70,000 gauss. As for the missing factor of 10 in plasma temperature, the obvious answer is to raise the power of the plasma-heating supply, either by additional ohmic heating or by some auxiliary method.

For reactor purposes ohmic heating

alone is problematic, since the resistivity of a plasma falls off as an inverse function of the electron temperature, and the current density falls off inversely with increasing plasma size. As reactor conditions are approached the power density delivered by ohmic heating tends to diminish in proportion to competing heat-loss processes (diffusion and radiation). In a deuterium-tritium plasma alpha-particle heating would ultimately come to the rescue, provided ohmic heating could just push the plasma temperature above 4 or 5 keV. This possibility has been examined intensively, but the conclusion has been that there is a real gap here, particularly if the objective is a power-generating tokamak reactor rather than an experiment to achieve and study ignition.

Of course, any plasma experiment that even approaches ignition conditions is of much scientific interest. From the considerations outlined above it follows that the best prospect for advancing toward reactor conditions in a purely ohmic-heated tokamak lies in the direction of using a very small plasma in conjunction with the strongest possible magnetic field. This idea is embodied in the Alcator devices at the Massachusetts Institute of Technology. The plasma in the first of these tokamaks, Alcator *A*, has a minor radius of 10 centimeters and a major radius of 54 centimeters; Alcator *A* has been operated at magnetic-field strengths as high as 85,000 gauss, attained with copper coils of the Bitter design (named after Francis Bitter of M.I.T.). In experiments done in the past few years the ohmic-heating power density in this device has been sufficient to allow plasmas with a central particle density exceeding 10^{15} electrons per cubic centimeter to be heated to temperatures approaching 1 keV (11.6 million degrees C.). The energy-confinement time in these experiments has been found to increase in proportion to the particle density, an anomalous pattern of behavior arising from the dominant contribution of electron diffusion to the heat-conduction process. The ions, on the other hand, have been found to behave in accordance with the prediction of the Coulomb-scattering theory of diffusion. A maximum confinement parameter of 3×10^{13} was reached in Alcator *A* under these conditions, representing a substantial improvement over earlier tokamak experiments. Recently the even more powerful Alcator *C* device has gone into operation.

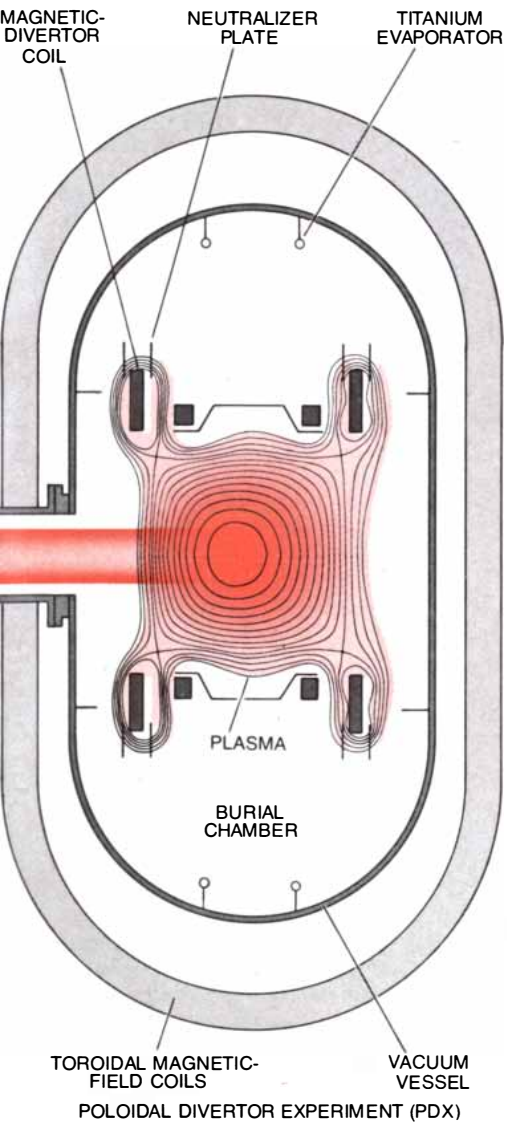
Ohmic-heating experiments in the Alcators and other tokamaks of various sizes and field strengths have provided a wealth of information about the scaling of the energy-confinement



TWO IMPORTANT ADVANCES in the design of tokamak fusion experiments are incorporated in the test system shown in this illustration. The apparatus seen in cross section at the left is a neutral-beam accelerator, an auxiliary plasma-heating source developed primarily at the Oak Ridge National Laboratory and the Lawrence Berkeley Laboratory. The energetic beam of neutral hydrogen atoms produced by the accelerator is injected into the tokamak, passing freely through the magnetic-field lines, whereupon the atoms are stripped of their electrons by collisions with the plasma particles and retained in the plasma as energetic positively charged ions. The tokamak seen in cross section at the right is equipped with a magnetic divertor, an arrangement in which a magnetic-field separatrix is employed to prevent impurity ions from entering the main plasma column, thereby reducing the rate of radiative heat loss. The impurity ions stick to a freshly deposited titanium film applied to the interior surfaces of the two "burial chambers," which accordingly act as a kind of pump. The particular design illustrated is that of the Poloidal Divertor Experiment, or PDX, device at Princeton, which can be operated in several different modes, giving the plasma either the squarish cross section seen here or one of two roughly *D*-shaped cross sections (*D* or *D*).

time characteristic of the tokamak regime. It is not possible in such experiments to vary certain key parameters (such as the poloidal-field strength and the temperature) independently, however, and so the scaling of the energy-confinement time must remain indeterminate to some extent. Ohmic-heating experiments by themselves cannot put the theoretical predictions concerning the tokamak plasma-confinement regime to their ultimate test.

In order to determine exactly how the energy-confinement time depends on these key parameters, some form of auxiliary plasma heating is needed. There is another motive for proceeding in this direction. The plasma particles in a tokamak reactor are expected to have very long mean free paths (or low scattering probabilities) because of their high temperature and comparatively low density. In this virtually collisionless regime, theory predicts that a new family of microinstabilities, called trapped-particle modes, could become the dominant energy-loss mechanism of a tokamak. In exploring plasmas with high particle-collision rates (that is, plasmas with a moderate temperature and a high density) the ohmic-heating experiments were able to achieve fairly high confinement parameters, but they left unanswered the question of how a collisionless reactor regime would behave. The task of exploring the critical high-temperature scaling of energy-confinement time calls for tokamak experiments with intense auxiliary heating sources. A number of such projects were launched in the early 1970's, and they have recently begun to yield significant results.



POLOIDAL DIVERTOR EXPERIMENT (PDX)

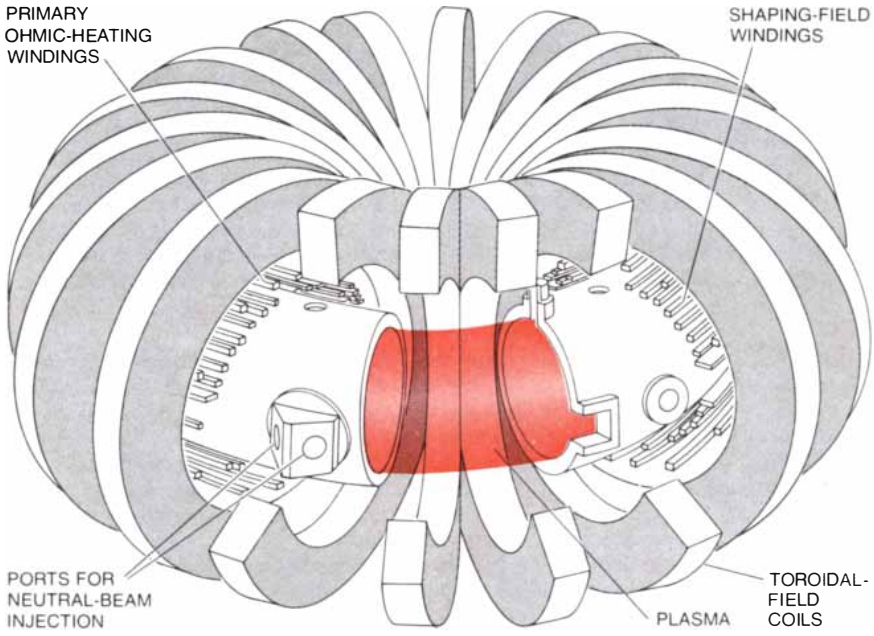
Before proceeding to the vicissitudes of auxiliary tokamak heating it is timely to bring up a question that has substantial impact even in the ohmic-heating phase: With so much heat flowing into the plasma, where does it all go when it flows back out? Heat transport radially across the magnetic surfaces consists of an outward flow of hotter plasma particles and an inward flow of colder ones. At the edge of the plasma, which is defined by the first magnetic surface that intercepts any solid exterior structure, there is a further flow of energetic particles along magnetic-field lines to the intercepting structure. The resulting concentration of particle bombardment tends to cause sputtering or even melting of the solid structure, which results in an influx of impurity ions into the plasma. Traditionally the plasma has therefore been bounded by special "limiters," heavy bars of a heat-resistant metal such as tungsten or molybdenum. Recent experiments have shown, however, that even a very small admixture of tungsten ions (as little as .1 percent) can give rise to a rate of heat loss by radiative cooling that greatly diminishes the energy confinement at the center of the plasma. Overheating of the limiter therefore tends to be self-curing, since the plasma then switches from being cooled by particle transport to being cooled by impurity radiation. This kind of remedy, however, is not conducive to the achievement of a reactor plasma with a high confinement parameter.

In the case of tokamaks with ohmic-heated plasmas the problem of the interaction of the plasma and the limiter can be solved in a straightforward manner, particularly at high plasma densities. The usual tokamak start-up procedure is to produce an initial gas discharge with a moderate density, which is then built up gradually by means of a pulsed input of hydrogen gas. The surface cooling of the plasma due to the influx of comparatively cold hydrogen gas tends to protect the limiter and suppress the influx of metallic impurities. If the plasma has a substantial admixture of light impurity ions, such as oxygen and carbon, then radiation cooling at the surface of the plasma becomes even more effective. Indeed, the surface cooling in poorly cleaned vacuum chambers can easily become all too effective, with the result that the gas-discharge channel shrinks away from the limiter and continues to contract until it becomes grossly unstable; in such a situation only low plasma densities can be reached in stable operation.

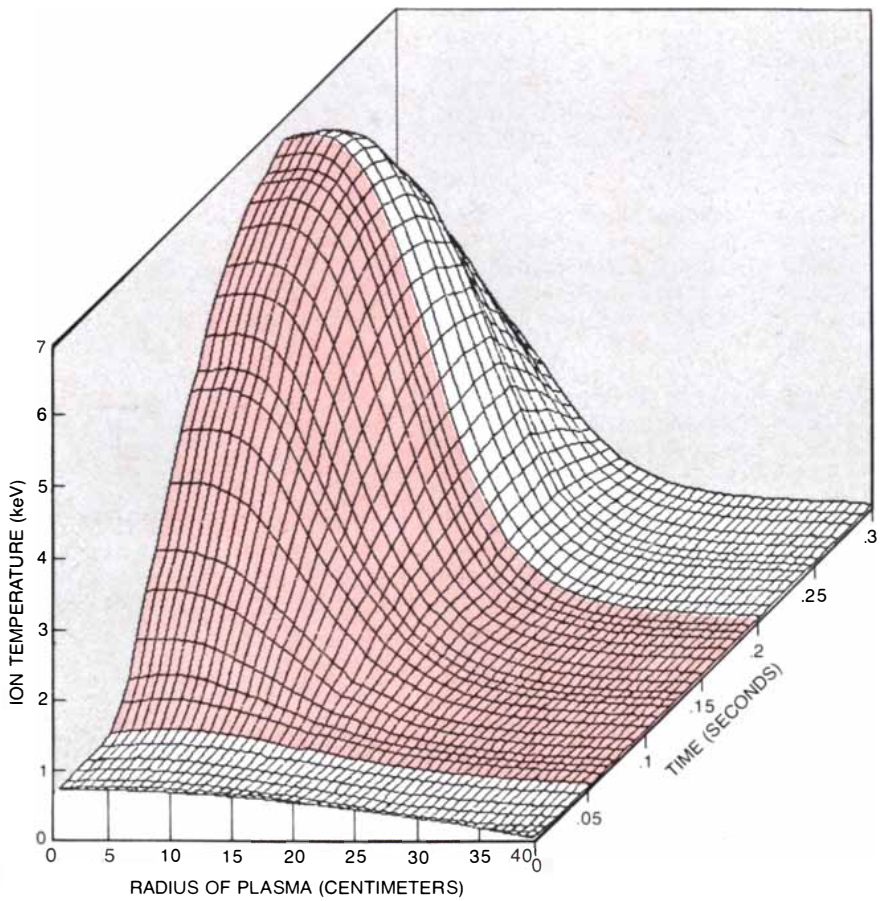
In tokamaks with auxiliary plasma-heating sources the problem of moderating the heat outflow to the limiters becomes much worse, particularly if the objective is to reach reactorlike conditions in a large plasma. Even a planned admixture of light impurity ions could not result in enough radiative cooling to carry off the heat outflow from a reactor plasma without also diluting the hydrogen-ion component to an extent incompatible with the requirements of the fusion-power balance. For reactor purposes it may be necessary to use a magnetic divertor, an arrangement in which a magnetic-field separatrix (a boundary line between two regions of different magnetic-field-line topology) takes over the role of the ordinary mechanical limiter. An alternative approach to the removal of a large input of auxiliary heating power is to create a high level of heat transport in a cool "blanket" region surrounding a hot, well-confined plasma core.

The most effective method of depositing large amounts of heating power in a tokamak plasma has turned out to be the injection of intense beams of neutral hydrogen atoms. In this approach, which was developed during the 1960's for open-ended fusion devices, the neutral atoms freely enter the magnetic bottle and are then stripped of their electrons by collisions and retained in the plasma as energetic ions. The trapped ions slow down gradually, transferring energy to the plasma particles. The application of this heating method to the tokamak became practical in the early 1970's, thanks to the development at the Oak Ridge National Laboratory and the Lawrence Berkeley Laboratory of high-current, multiaperture ion sources capable of generating neutral beams in the 100-kilowatt range at energies of tens of kilovolts.

Early in the 1970's neutral beams were applied in the ATC tokamak at Princeton and the Ormak at Oak Ridge. Increases in ion temperature of a few hundred electron volts were detected and were duly hailed as marking the start of the new era of nonohmic tokamak heating. Unfortunately the plasma electrons were generally cooled during this heating process, apparently because of the injection of impurity ions by the impact of the beam particles on the limiter and on the walls of the vacuum



INTERNAL STRUCTURE OF THE PLT, at present one of the largest tokamaks in operation in the world, is exposed in this schematic cutaway drawing. The plasma torus inside the PLT has a major radius of 130 centimeters and a minor radius of 45 centimeters. Experiments currently being done in the PLT are designed to test various plasma-heating methods, including ohmic heating, neutral-beam heating and microwave heating. A large number of diagnostic devices are normally installed on the tokamak to monitor the properties of the plasma.



PLT RESULTS are presented in this three-dimensional computer plot of ion temperature as a function of time and distance from the center of the plasma. The colored region shows the duration of the neutral-beam pulse. The peak value of approximately 6.5 keV corresponds to an ion temperature of better than 75 million degrees C., a new record in tokamak research.

chamber. By 1977 half-megawatt powers were being injected, and ion temperatures were being raised to the 2-keV level in the Ormak and also in the TFR device at Fontenay-aux-Roses in France. The electrons, however, remained reluctant. Small rises in electron temperature began to be seen, but not distinctly enough to make possible an assessment of the temperature dependence of the energy confinement.

Meanwhile two new tokamaks, each with a minor plasma radius of about 40 centimeters, had made their appearance: the T-10 at the Kurchatov Institute, under the direction of V. S. Strelkov, and the PLT at Princeton, under the direction of Wolfgang Stodiek. In ohmic-heated operation these giant tokamaks suffered from the expected weakening of power density associated with their size, and they have both been limited to maximum ion temperatures of about a kilovolt and plasma densities of a little more than 10^{14} particles per cubic centimeter. The energy-confinement times were found to scale up as the square of the minor radius, as expected, and to reach values of about 100 milliseconds, but the associated confinement parameters have barely risen above 10^{13}. The real advantage of large tokamaks has turned out to be their accessibility for high-powered external heat inputs.

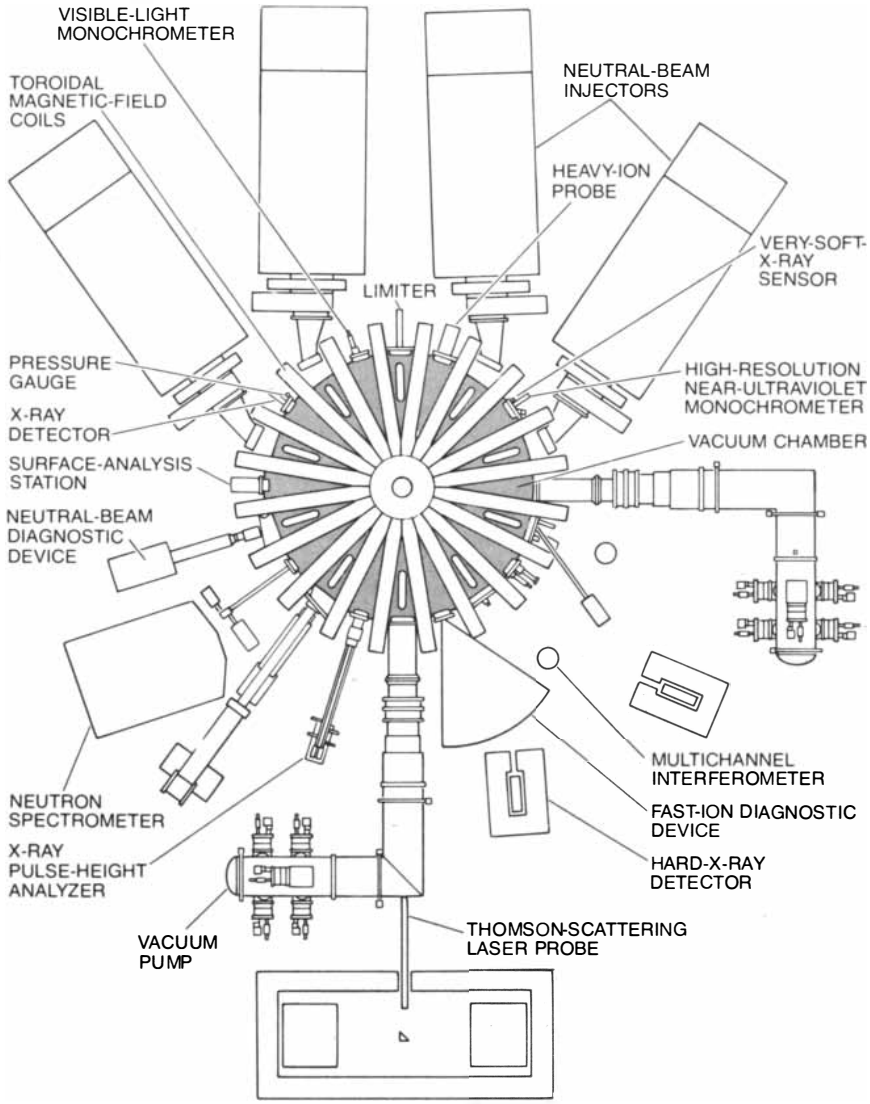
In the fall of 1977 two neutral-beam lines operating at an injection energy of 40 keV were brought into operation on the PLT by a group headed by Harold P. Eubank. These injectors, which represented the first installment of a set of four developed and built at the Oak Ridge facility for the PLT, were capable of injecting a total of 1.1 megawatts of neutral-beam power into the plasma. In December the ion temperature in the PLT rose to 2.2 keV—a new record by a whisker—but the ions were clearly being held back by collisions with cooler electrons resulting from severe radiative losses to the ions of metallic impurities. Under these conditions several important findings were made on the spectroscopy of highly ionized tungsten and iron, but the opportunity for further progress in ion temperature seemed limited, even with the injection of higher-powered neutral beams.

Some exploratory runs were made with graphite as the limiter material, and they turned out to be promising. Carbon sputters more readily than tungsten, but an admixture of carbon ions is far less damaging to the energy balance of the plasma because carbon becomes fully ionized in the plasma's interior and radiates mainly from its cold edge. Early last year the remaining two neutral-beam lines were fitted to the PLT, and at the same time a new set of water-cooled graphite limiters was introduced. In July two megawatts of neutral-beam power was injected into the PLT, and the electron temperature finally responded, ris-

ing to 3.5 keV. Under these conditions the ion temperature was able to climb to 5.5 keV. This peak ion temperature, equivalent to more than 60 million degrees C., was substantially better than had been hoped for and had a special appeal to fusion workers: for the first time a toroidal device had exceeded the minimum ion temperature required for ignition in a deuterium-tritium plasma.

The main scientific excitement, however, centered on the scaling of the plasma confinement. A comparatively low central plasma density of about 5×10^{13} particles per cubic centimeter was used in the PLT heating experiments, since the highest particle temperatures can be achieved when there are fewer particles to share the input heating power. In this regime of high temperature and low density the PLT plasma for the first time reached a degree of collisionlessness resembling that of a tokamak reactor, without encountering any observable deterioration in confinement. The magnitude of the heat transport by ions continued to be roughly consistent with the predictions of the Coulomb-scattering theory of diffusion, rather than showing the marked enhancement that had been predicted by the trapped-particle theory of microinstabilities. As a kind of consolation prize for microinstability theorists the PLT plasma did begin to show a new type of high-frequency density fluctuation under the highest-temperature conditions. Detailed diagnosis of the fluctuation may yet reveal that it is indeed of the predicted type. The new phenomenon, however, has had no measurable impact on the actual plasma confinement; this observation remains true even though the ion temperature has meanwhile been pushed to 6.5 keV (75 million degrees C.), with 2.5 megawatts of injected neutral-beam power.

The marked increase in electron temperature produced by the nonohmic neutral-beam heating in the PLT has also made possible a first glimpse of the true temperature dependence of the heat transport by electrons. To a first approximation there has been no departure of the energy-confinement time for electrons from the value prescribed by the simple empirical scaling law observed in tests on Alcator *A* and other ohmic-heated tokamaks: confinement time is proportional to the density multiplied by the square of the minor radius of the plasma. Closer inspection of the evolution of the electron-temperature profile during neutral-beam heating, however, has revealed the surprising and encouraging result that the heat transport by electrons actually decreases in the hottest part of the plasma core while remaining about the same at the edge. Now that the electron temperature of a tokamak can be varied independently of the poloidal magnetic field and the particle density of the plasma the establishment of an explicit empirical scaling law is in sight. With such a law it will begin to be possible to make a choice among theoretical heat-transport models.



NEXT GENERATION OF TOKAMAKS is represented by this plan view of the TFTR system, which is scheduled to go into operation at the Princeton laboratory in 1981. The vacuum chamber of the TFTR will have a major radius of 265 centimeters and a minor radius of 110 centimeters. The TFTR will be the first magnetic fusion device to use a deuterium-tritium fuel mixture and also the first such system capable of producing fusion energy in any significant quantity. The estimated cost of the entire TFTR facility is on the order of $300 million.

How far is the PLT regime from meeting the requirements of a full-scale power-producing reactor? Since the temperature domain of such a reactor was entered in the PLT at the low-density end of the operating range, the confinement parameter was only about 10^{12}. Higher neutral-beam powers and injected-ion energies will be needed to push up the plasma density and the confinement time. Even within the present PLT regime, however, the release of fusion power is already substantial. When deuterium plasmas and deuterium beams are used, about 4×10^{13} fusion events are produced per pulse. The deuterium-tritium reaction is not used in the current series of plasma-physics experiments, because the radiological safety problems associated with handling tritium would needlessly complicate the experimental procedures. Instead one calculates an equivalent deuterium-tritium yield on the basis of the known fusion-reaction rates, which are roughly 300 times greater for a deuterium-tritium fuel mixture than for pure deuterium. In the PLT the equivalent deuterium-tritium fusion power is about 50 kilowatts, corresponding to a Q value of .02, a record in fusion research to date.

The comparatively high Q value attained in the PLT is attributable to the nature of the heating method. Neutral-beam heating introduces the input power in the most favorable way: it enhances the high-energy "tail" of the ions' ener-

gy distribution, which is responsible for most of the fusion reactivity. Because of this special feature, the average ion energy at the center of the PLT plasma with an ion temperature of 6.5 keV is actually about 13 keV (counting the injected deuterium ions), and the fusion yield is about 100 times greater than it is when a nonreactive hydrogen beam is used for heating.

From the perspective of the 1950's, when plasma temperatures of less than 100 electron volts and confinement parameters of 10^{10} were typical, such nuances of reactor design as the degree of departure from a normal bell-shaped energy-distribution curve did not loom very large. The goal was simply to exceed the magic combination of a 10-keV temperature with a confinement parameter of 6×10^{13}, a threshold commonly known as the Lawson criterion for deuterium-tritium plasmas (after the British physicist J. D. Lawson). Now that actual experiments are about to enter the reactor regime the outlook has changed, and the fine points of the toroidal-reactor objective are beginning to have important consequences in terms of hardware and cost.

The basic plan of all deuterium-tritium reactors is roughly the same. About 80 percent of the fusion power would emerge from the plasma in the form of 14-MeV neutrons and would be captured as heat in a lithium blanket, which would also serve to "breed" the tritium supply for the plasma. The neutron-generated heat, together with the outflow of heat from the plasma itself, would then be converted into electric power by a conventional steam cycle. Another common design feature is that the magnet coils of the reactor would have to be superconducting, since the power consumption of ordinary copper coils would be expected to exceed the capacity of the reactor for generating electric power.

The individual distinctions among the various designs for a deuterium-tritium reactor relate largely to features of the plasma itself. In toroidal reactors there is a hierarchy of possible operating modes, rated in order of ascending Q values. The particles in the most easily achievable regimes have special energy distributions that are comparatively rich in energetic ions. In this case Q values as large as 2 or 3 should be achievable at confinement parameters as low as 10^{13}. Unfortunately there is a great economic disadvantage in having a reactor with such a low Q value: a substantial part of the output of electric power would have to be recirculated to keep the plasma hot, and extremely high system efficiencies would be needed to produce net power. The ultimately attractive solution is clearly a reactor operating with an ignited or nearly ignited regime, where most of the plasma heating would be done by the alpha particles produced in the deuterium-tritium reaction. In this case the plasma particles would tend to be well distributed in energy by Coulomb collisions, and the confinement parameter would have to reach about 3×10^{14}. Since the energy-confinement time is expected to scale up as the square of the minor radius of the plasma, this means that the anticipated minimum size for an ignited tokamak plasma tends to be several times larger than that for a tokamak plasma with a low Q value.

NEXT GENERATION OF PROBLEMS in tokamak research is expected to include a new class of macroscopic instabilities called ballooning interchange modes, one of which is illustrated here. In experiments with a high ratio of kinetic plasma pressure to magnetic-field pressure the plasma may have a tendency to squeeze out between the field lines of the magnetic bottle while deforming them slightly. The result in this case would be a ripple effect around the outside surface of the plasma torus, aligned approximately with the pitch of magnetic-field lines.

The attainment of Q values somewhat greater than 1 is a realistic prospect for the next generation of tokamak experiments. Two of the large tokamaks currently under construction will be capable of demonstrating this type of operation on a short-pulse basis in actual deuterium-tritium plasmas. The devices, the Tokamak Fusion Test Reactor (TFTR) at Princeton, which is scheduled for completion in late 1981, and the Joint European Torus (JET) at Culham in England, scheduled for 1982, will have plasma currents of about three million amperes, neutral-beam heating powers of tens of megawatts and expected fusion-generated output powers of between 10 and 100 megawatts. The minor radius of the plasma in the TFTR and the JET will be respectively two and three times larger than the minor radius of the plasma in the PLT; it is estimated that the cost of each of the new facilities will be on the order of $300 million.

In principle one could seek to progress "straightforwardly" from these next-generation experiments to a high-duty-cycle power-producing ignition reactor in the 100- to 1,000-megawatt range by increasing the size of the plasma and adding a lithium blanket and superconducting magnet coils. The estimated cost of such a step, however, is about $1 billion. Moreover, many substantial scientific and technical problems remain to be solved. In order to address these problems with maximum effectiveness the leading contributors to the world fusion effort—the U.S., the U.S.S.R., Western Europe and Japan—have recently formed a joint study group called INTOR (for International Tokamak Reactor) under the auspices of the International Atomic Energy Agency in Vienna. Within the U.S. fusion program there is an analogous design effort named ETF (Engineering Test Facility), which is centered at Oak Ridge. These studies are expected to generate well-considered preliminary designs of a tokamak ignition reactor to be completed in the early 1990's, as well as programmatic outlines of the research-and-development tasks that remain to be addressed by existing facilities.

Under the latter heading a key topic is the maximizing of the kinetic pressure of the tokamak plasma with respect to the magnetic-field pressure. The attainable pressure ratio, a parameter called the beta value of the plasma, is expected to be limited by a class of macroscopic magnetohydrodynamic instabilities called ballooning interchange modes, wherein the plasma squeezes out between the field lines of the magnetic bottle while deforming them slightly [*see illustration at left*]. The smaller the beta

value is, the less the plasma is able to deform the magnetic field and the more reliable is the plasma confinement. On the other hand, the larger the beta value, the higher the plasma density at a given temperature, the higher the confinement parameter and the greater the projected fusion-power production for a plasma of given size.

At present there is an intensive theoretical effort, using some of the world's largest computers, to design the optimum tokamak geometry and to estimate the maximum attainable beta value. Present theoretical estimates of the beta value range from 3 to 10 percent, which should be generally adequate for a reactor, but much depends on whether the upper or the lower end of the range turns out to be applicable. Current experiments have yet to employ sufficiently intense auxiliary heating for beta values higher than 2 percent to be reached; accordingly no pressure-driven instabilities have yet been encountered. The first significant news on this topic is likely to come during the next two years from the ISX tokamak at Oak Ridge, the PDX at Princeton and the Doublet III at the General Atomic Company laboratory in San Diego, all of which will use high-power neutral-beam injection to determine the beta values of tokamaks with variously shaped minor cross sections. The favorite at present is an elongated *D*-shaped cross section, but the real winner and the magnitude of its beta value are yet to be established.

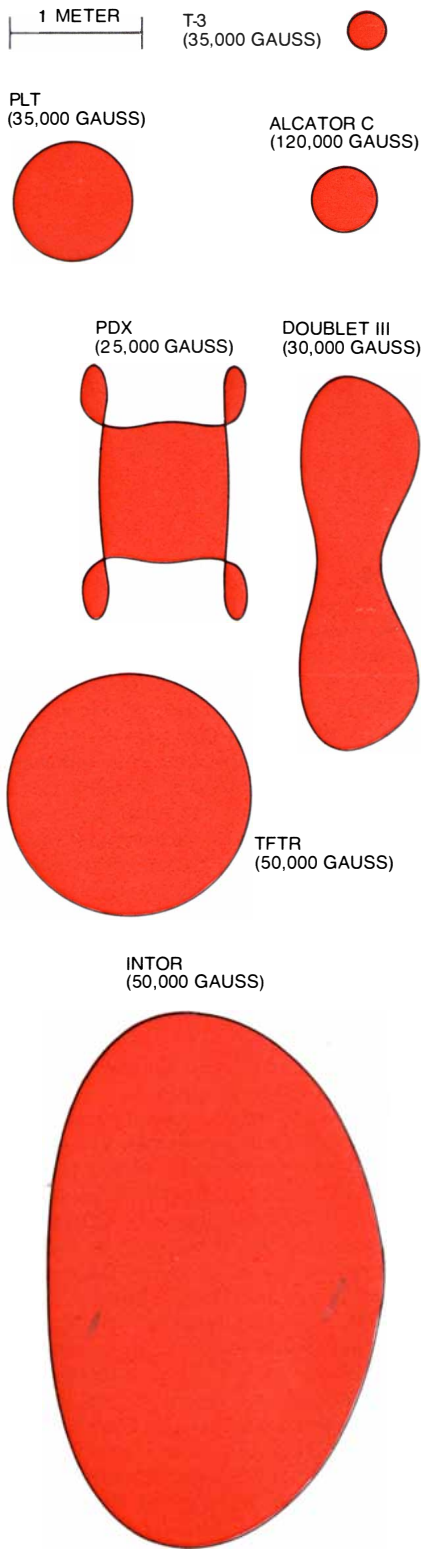
A second question that has far-reaching significance for reactor design concerns the best method of controlling impurities. The magnetic-divertor approach, which is about to be tested in the PDX device, appears to be a good prospect, but it would contribute significantly to the overall size of the reactor. The outlook will be clarified in 1982 by the start of operations on the JT-60, a Japanese tokamak project on the scale of the TFTR and JET devices; the JT-60 will not burn a deuterium-tritium fuel mixture, but unlike the other two giants it will have a magnetic-divertor system.

The problem of maintaining a desirable plasma composition is expected to become particularly severe as tokamak experiments progress from the typical current duty cycle of about a one-second pulse every 10 minutes to pulses lasting for 10 minutes or more separated by much briefer intervals. In order to explore the high-duty-cycle regime a new generation of tokamaks with superconducting magnet coils will be needed. The first important step in this direction is being taken by the Russians with their T-15 device, which will be put into operation at the Kurchatov Institute in 1983. Still another avenue of technical advance that will affect the destiny of tokamak reactors is the development of plasma-heating methods simpler than neutral-beam injection. In particular, microwave plasma heating holds much technological promise and is being explored on a wide front by current tokamak experiments.

The remarkable growth and cumulative success of the tokamak approach to fusion would have gratified Congressional committees that reviewed fusion research in the 1960's. They sought to sharpen the focus of the research so that less promising alternatives could be abandoned. Now that a sharper focus has been achieved a new concern has arisen: Should one particular solution be chosen before others have even had a critical test? A tokamak reactor seems likely enough to work, but its cumbersome size and complex geometry may ultimately limit its economic attractiveness in competition with mirror machines, inertial reactors and an entire spectrum of other approaches that seem to be ordered in terms of increasingly interesting economic potential and diminishing technical evidence.

Although this line of reasoning has a somewhat paradoxical flavor, it makes a basically valid point. The abundance of technological alternatives in fusion research represents a valuable resource, which should be cultivated more vigorously than ever during the period when the first experimental fusion reactors go into actual operation. Although the tokamak is considered the best prospect as the first reactor, other configurations that now seem fanciful may ultimately take over the fusion industry. In the very long run there are many possibilities for further evolution.

The emergence early in the next century of an energy economy based on tokamak fusion reactors burning a deuterium-tritium fuel mixture would allow the nuclear-power approach to the energy problem to realize a large advance in environmental attractiveness and presumably in public acceptance. A deuterium-tritium reactor would be incapable of a runaway reaction, it would have little afterheat from radioactivity and it would produce radioactive wastes with a substantially reduced hazard potential and lifetime. Within the first century or two of commercial fusion power it seems likely, however, that the deuterium-tritium reaction would itself be replaced by fusion processes that are more difficult to achieve but have even more desirable environmental features: first the deuterium-deuterium reaction and finally one of the more exotic fusion reactions, particularly one that releases its heat entirely in the form of charged particles rather than neutrons. The prospect of adapting fusion power perfectly to the needs of society and the environment does not appear to be limited by any narrow physical constraints, only by the incompleteness of scientific knowledge and the limits of technological imagination.



MINOR CROSS SECTIONS of a variety of tokamak plasmas, ranging from the T-3 to INTOR, are drawn to the same scale in this illustration. The strength of the toroidal magnetic field is also indicated for each device. The PDX (shown here in its squarish incarnation) and the Doublet III tokamak at the General Atomic Company are designed to use high-power neutral-beam injection to investigate pressure-driven instabilities in tokamaks with variously shaped minor cross sections. The elongated *D* shape shown here is currently favored for the INTOR minor cross section.

A Pre-Neolithic Farmers' Village on the Euphrates

The Neolithic period is traditionally associated with the emergence of agriculture. At a site in Syria, however, some of the first villagers practiced farming, along with hunting and gathering, even earlier

by Andrew M. T. Moore

The stage in human prehistory when man first adopted an agricultural economy is of fundamental interest to archaeologists because it represents the watershed between the hunter-gatherer existence of our ancestors of the Paleolithic period, or Old Stone Age, and the first civilized way of life. For generations scholars have scoured southwestern Asia in an attempt to find out just where and when agriculture began. The pace of exploration and excavation has greatly accelerated in the past 30 years, but archaeologists are now less concerned with where and when an agricultural way of life developed in the region than with how and why. Recent analysis of work at a site in the Euphrates valley of northern Syria has produced data that suggest answers to these last two questions.

The part of southwestern Asia with the largest number of known and excavated sites of Mesolithic (immediately following the Paleolithic) and Neolithic age is the area traditionally known as the Levant. Today the area is shared by the modern states of Syria, Lebanon, Jordan and Israel. The site of the work I shall describe is in northern Syria. By far the largest early Neolithic site anywhere in the Levant, it is a tell: a large mound consisting of the debris of millenniums of human occupation. The mound, known as Tell Abu Hureyra, was destined to be flooded when the water of the Euphrates was impounded by a new dam being constructed a few kilometers downstream at Tabqa. I first saw the mound in March, 1971, three years before it was scheduled to disappear, and was immediately impressed by its size and potential for investigation.

It was obvious that the thick deposit of debris, measuring up to eight meters in height over an area of 11.5 hectares (some 28 acres), must have been built up over a long period. There was thus every hope that excavations at Abu Hureyra would provide abundant evidence on the beginnings of settled life on this eastern frontier of the Levant, where today only an annual 200 millimeters of rain falls on what is mainly an arid steppe.

The surface of the mound was covered by a dense scatter of flint tools and waste similar in type to material that had been excavated some years earlier at the Neolithic site of Buqras, farther down the Euphrates, by Henri de Contenson of the French National Center for Scientific Research (C.N.R.S.). Buqras was dated to about 6000 B.C. (All the dates I shall give here are based on carbon-14 determinations calculated according to the half-life of that radioactive isotope but without calibration.) This places Buqras toward the middle of Neolithic times in the Levant, where the New Stone Age began about 8500 B.C. and continued in most of the region until as late as 4000 to 3500 B.C. If artifacts like those at Buqras were on the surface at Abu Hureyra, one could expect to find below the surface remains from the archaic, or earliest, Neolithic, and perhaps below them remains from the preceding Mesolithic, which here began some 20,000 years ago and continued until about 8500 B.C.

My colleagues and I began work at Abu Hureyra in 1972 under the sponsorship of the Pitt Rivers Museum of the University of Oxford and with the support of a number of other institutions both in the United Kingdom and in North America. By the end of the first season's excavations our optimistic expectation had been fulfilled. Below the levels of archaic and developed Neolithic occupation, dating from about 7500 B.C. to some time after 6000 B.C., appeared evidence of a Mesolithic occupation at the site that had lasted for several centuries in the late 10th and early ninth millenniums B.C. Another season in 1973 saw completion of work at the site, and the great mound was submerged on schedule early in 1974, leaving our group with an immense stockpile of artifacts and animal and plant remains for later examination. Only now, as analysis of the material progresses, is the full significance of the excavation being realized in terms of the transition from a hunting-gathering way of life to a farming one.

The Mesolithic occupation zone at Abu Hureyra is at the northern end of the mound on the edge of the Euphrates floodplain. It consists of a simple village settlement: a series of pit dwellings and other hollows dug into the original ground surface underlying the tell. Each pit dwelling had a framework of upright wood posts; the roofs, of which no trace remains, were probably made of reeds. The pits and hollows had gradually been buried under occupation debris, which had accumulated to a depth of about a meter. Along with the debris were many of the artifacts the Mesolithic occupants had used.

Prominent among the artifacts were stone pestles, rubbing stones and milling stones. Many of them were stained with red ocher, evidence that they had been used for crushing this decorative pigment into powder. The presence of the milling implements in the pit dwellings and hollows suggests that they were also used to process plant foods.

The artifacts found in the greatest numbers were flint tools of two types. Many are so small that they are termed microliths, but a number are heavier flint objects. Also present are immense quantities of waste flint.

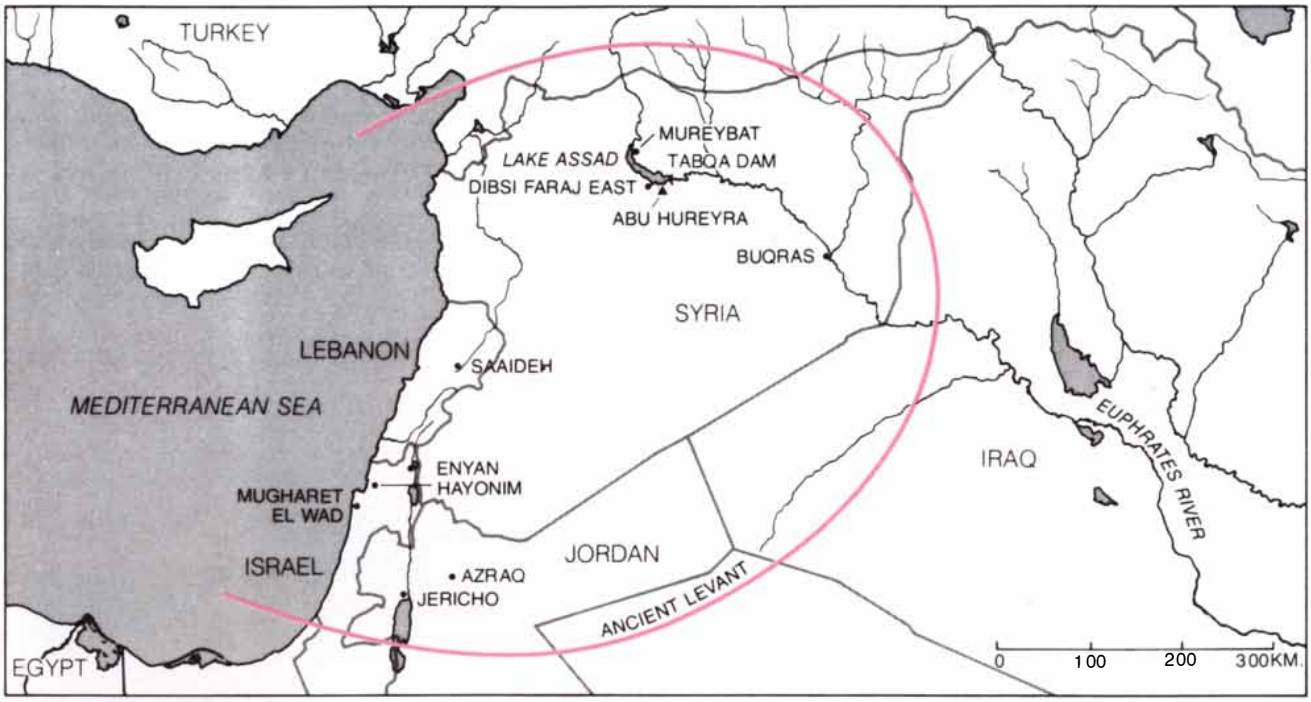
By far the commonest of the microliths are small crescent-shaped blades known as lunates; these bladelets could have been set into wood hafts to serve as barbs for an arrow, the cutting edge of a knife or (as is known to be the case elsewhere in the Levant) teeth for a sickle. Other microlith shapes include bladelets blunted along one edge and bladelets chipped to form tiny points for boring

holes. Most of the larger stone tools are either hammerstones or scrapers made from sizable flakes of flint. The artisans of this Mesolithic phase also laboriously fashioned dishes of stone other than flint and made notches in the sides of pebbles, possibly to fit them for use as weights for fish nets.

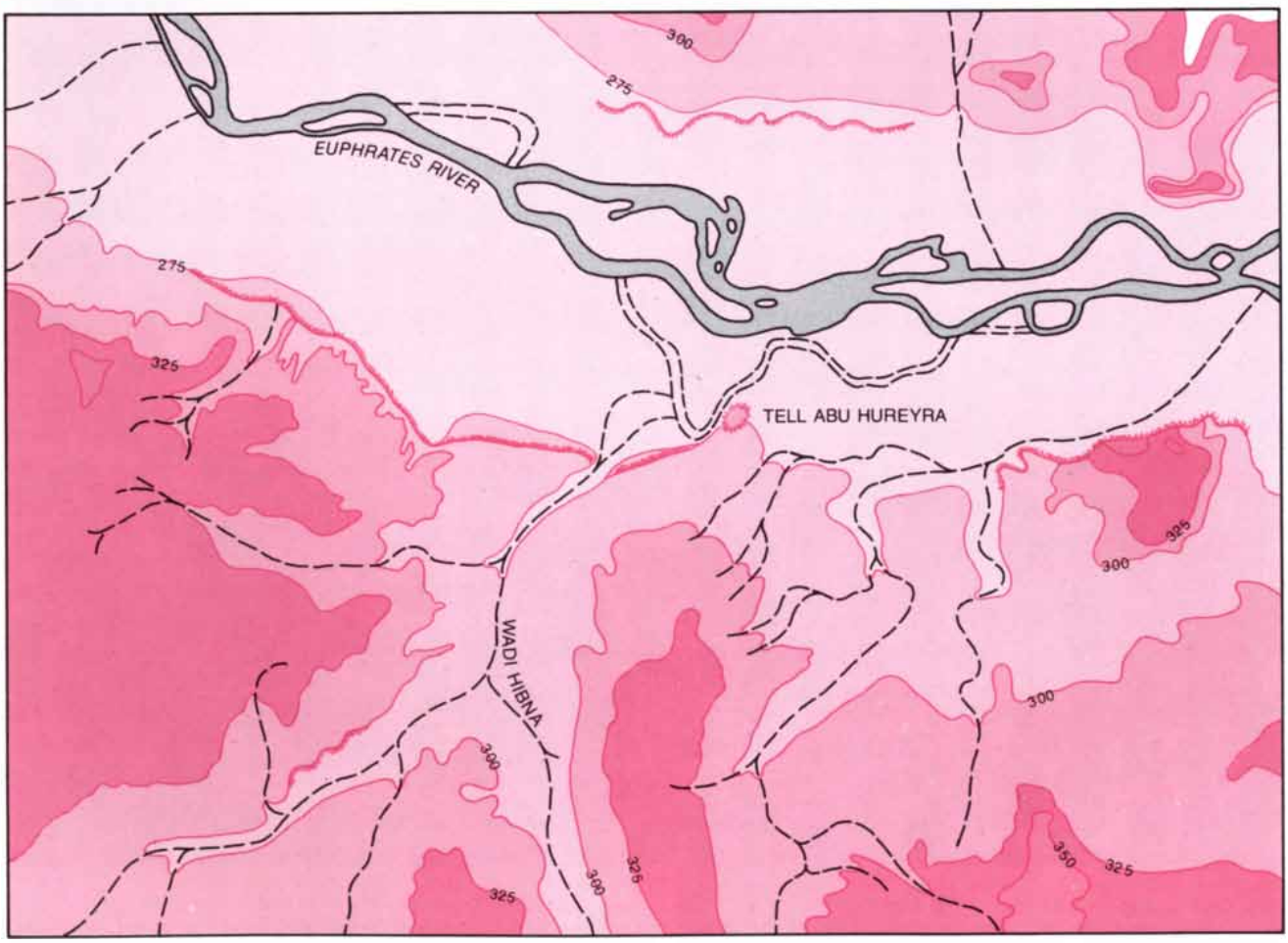
Tools made of bone were also found in abundance. Most of the bone tools are heavy-duty boring points; other bone artifacts include spatulas, needles and double-end points that may have been arrow tips. Among the artifacts of a decorative character were bone beads and pendants, together with beads and pendants made of shell and stone.

The artifacts associated with the Mesolithic village at Abu Hureyra are typical of the later stage of this period elsewhere in the Levant, roughly contemporary with such sites of the Natufian culture in Palestine as Mugharet el Wad, Eynan and Hayonim. During the later Mesolithic period the human population of the Levant was considerably larger than it had been during the preceding several thousand years. The increase in population appears to have been in response to a marked improvement in the environment. Where the earlier stage had been characterized by

TWO OCCUPATION LEVELS at Tell Abu Hureyra appear in this photograph of a test excavation at the northern end of the 28-acre mound. The rectangular mud-brick walls of the dwellings of the archaic Neolithic villagers who settled at the Euphrates River site in about 7500 B.C. are visible at the top. Buried under this Neolithic village until it was exposed by the excavators was a cluster of pit dwellings and work hollows used by sedentary hunter-gatherers of the later Mesolithic period until about 8500 B.C. They evidently raised grain. Postholes around the pits held supports for walls and roofs probably made of reeds. The small holes were made by burrowing rodents.



ANCIENT LEVANT, an area now shared by the modern states of Israel, Lebanon, Syria and Jordan, is the part of southwestern Asia that has the largest number of excavated Mesolithic and Neolithic sites. Tell Abu Hureyra (*triangle*) was flooded by damming in 1974.



VILLAGE MOUND stands on a low terrace south of the floodplain of the Euphrates. Wadi Hibna, now dry, may have been a perennial stream during archaic Neolithic times that could have been tapped for irrigation. The contour numbers indicate meters above sea level.

cool, dry conditions, the later stage and the following Neolithic period saw a gradual rise in temperature and an increase in rainfall.

The consequence for the Levant was an eastward expansion of the zone of Mediterranean forest that had formerly been confined to the coastal mountains of the region. This wooded zone ultimately reached the edge of the interior plateau. The steppe east and south of the enlarged forest area was separated from the forest proper by an intermediate zone of open woods. During the period of increasingly warm and wet weather the steppe itself became richer in plant and animal species. As a result the intermediate zone between full forest and steppe came to offer particularly attractive opportunities that human populations on the threshold of agriculture were free to exploit.

Abu Hureyra, situated on the edge of the steppe, seems to have been one of a number of settlements that came into being as a result of the improvement in the environment. The process is predictable. The greater rainfall would have brought with it a greater abundance of vegetation, which for small groups of Mesolithic hunter-gatherers would have been a factor of prime importance. Today we would say that the more abundant vegetation was able to support a larger biomass; indeed, this is attested to by the demonstrable growth in the human population. Evidently some of the hunting-gathering groups began to coalesce and set up a sedentary, or at least semisedentary, way of life. The concentration of such enlarged groups at sites such as Abu Hureyra in itself created new circumstances, for example richer social relations. It became advantageous under these circumstances to develop new strategies of subsistence in support of the novel sedentary existence.

Such, at least, seemed a plausible hypothesis to account for the Mesolithic village we discovered at Abu Hureyra. To test the hypothesis the excavators took pains to recover as much evidence on the economy of the settlement as possible. This called for excavation techniques rarely used before in the Levant. First, working with fine-mesh screens the excavators sieved all the soil they had removed from the dig. The dry-sieve technique recovered large quantities of animal bone, more than two tons in all. Second, a flotation method was applied to separate fine plant materials from the sieved soil. Two flotation machines, designed at the University of Cambridge, systematically recovered about 1,000 liters of plant remains from the various levels of the excavation.

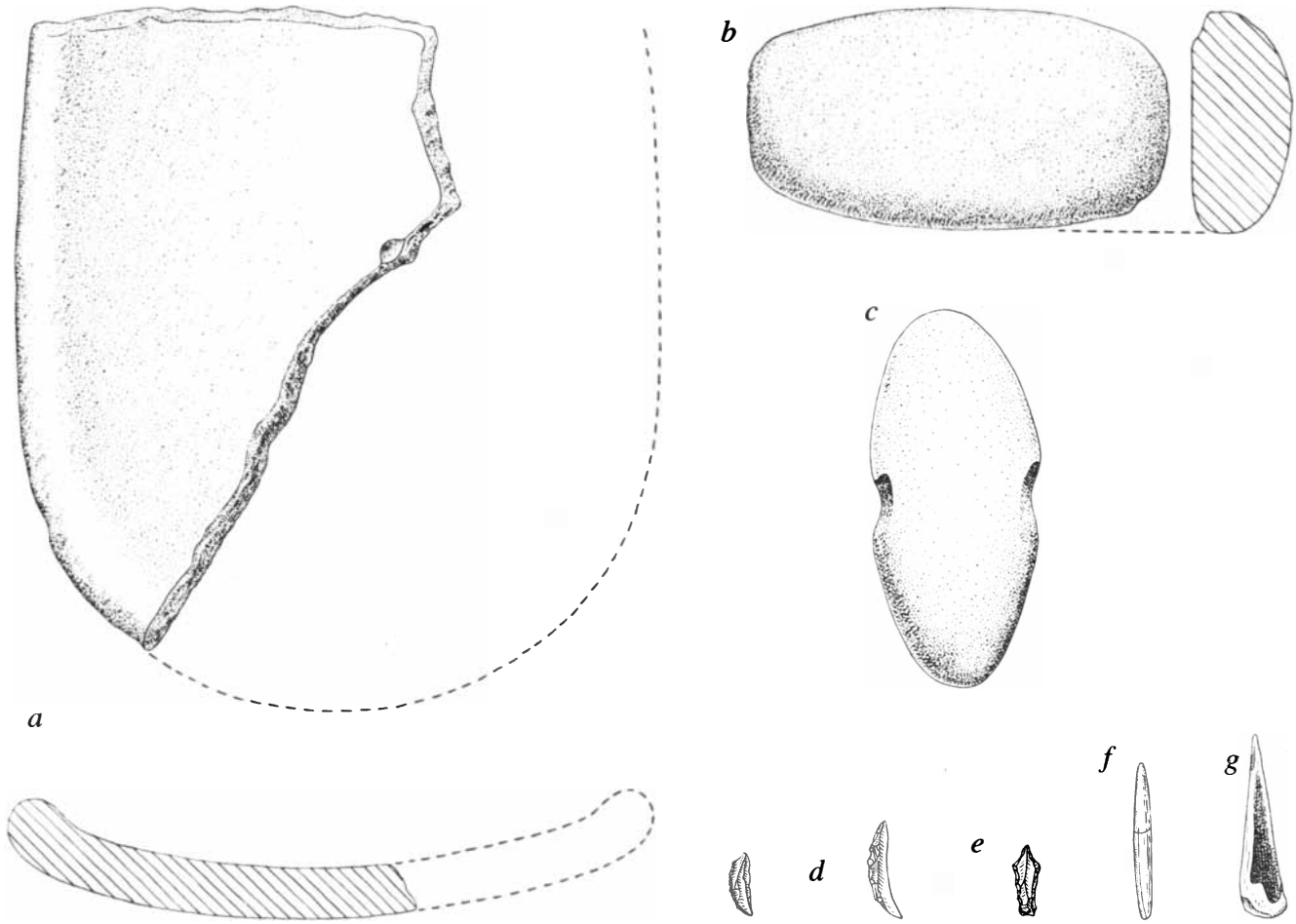
Examination of these plant remains, undertaken by Gordon Hillman of the University of Wales, shows that the Mesolithic inhabitants of Abu Hureyra collected, if they did not actually cultivate, various pulses, including a wild type of lentil. They also gathered tiny nuts from the turpentine tree (a relative of the pistachio), fruits from hackberry shrubs (plants of the elm family), caper berries and the seeds of wild feather grasses. What came as an exciting and significant surprise was Hillman's discovery of cereal grains. The most numerous of them are grains of a primitive wheat: wild einkorn. Two other cereals are present in less abundance: wild rye and wild barley. Because the cultivation of cereal grains is one of the earliest landmarks of the Neolithic revolution the discovery of three kinds of cereal in the Mesolithic levels at Abu Hureyra raised a crucial question. Were any of the three (and perhaps also the pulses) already being cultivated, or were they too only being collected from the wild?

FLOTATION MACHINE designed at the University of Cambridge was one of two machines used by the excavators at Abu Hureyra to recover some 1,000 liters of plant remains from various site levels. The Mesolithic levels held three kinds of cereal grain: wheat, rye and barley.

In considering this question one must first note that neither einkorn wheat nor the other two cereals are found in the wild form anywhere near Abu Hureyra today; the area is too arid. Given the cooler and more humid conditions of the later Mesolithic, however, they might have grown there in the past. This raises another question. Does the fact that the cereal grains recovered from the Mesolithic village at Abu Hureyra are morphologically wild types mean they could not have been cultivated? Not necessarily.

Hillman made use of stands of wild einkorn growing in the Munzur Mountains of eastern Turkey to experiment with various methods of harvesting and with patterns of crop rotation. These experiments were undertaken to determine whether selection pressures favoring mutation from the wild state might operate under such conditions. His conclusion is that as long as the cultivation methods appropriate to simple husbandry were followed the morphology of wild einkorn would remain unaltered even over long periods. Only after unusual cultivation techniques have been applied for many hundreds of years, Hillman asserts, do mutations from the wild to the domestic state appear.

There is evidence suggesting that the



MESOLITHIC ARTIFACTS include basalt milling stones (*a*) and rubbers (*b*) and notched pebbles, possibly net weights (*c*). Among the chipped-flint microliths, lunates (*d*) were the most abundant; microborers (*e*) ranked next. Bone tools include double-end points (*f*), possibly points for arrows, and many borers (*g*). Milling stones were certainly used to grind ocher and perhaps also to process plant foods.

inhabitants of Abu Hureyra practiced just such simple methods. For example, the cultivation of cereal grains, even on a small scale, should have disturbed the natural vegetation in the vicinity. This means that the same weed plants characteristic of cultivated areas near the site today should be present in the flotation samples from the Mesolithic levels at the site. It is possible, of course, that tramping of the ground by the Mesolithic inhabitants, or even by their animals, might have opened the way for these same weeds. Be that as it may, the flotation samples show that three such modern weeds (*Atriplex,* one of the steppe-growing members of the goosefoot family, *Alyssum,* a member of the mustard family that is an ornamental plant today, and *Lithospermum,* a member of the borage family) grew close enough to Abu Hureyra in later Mesolithic times to leave their seeds in the excavated soil.

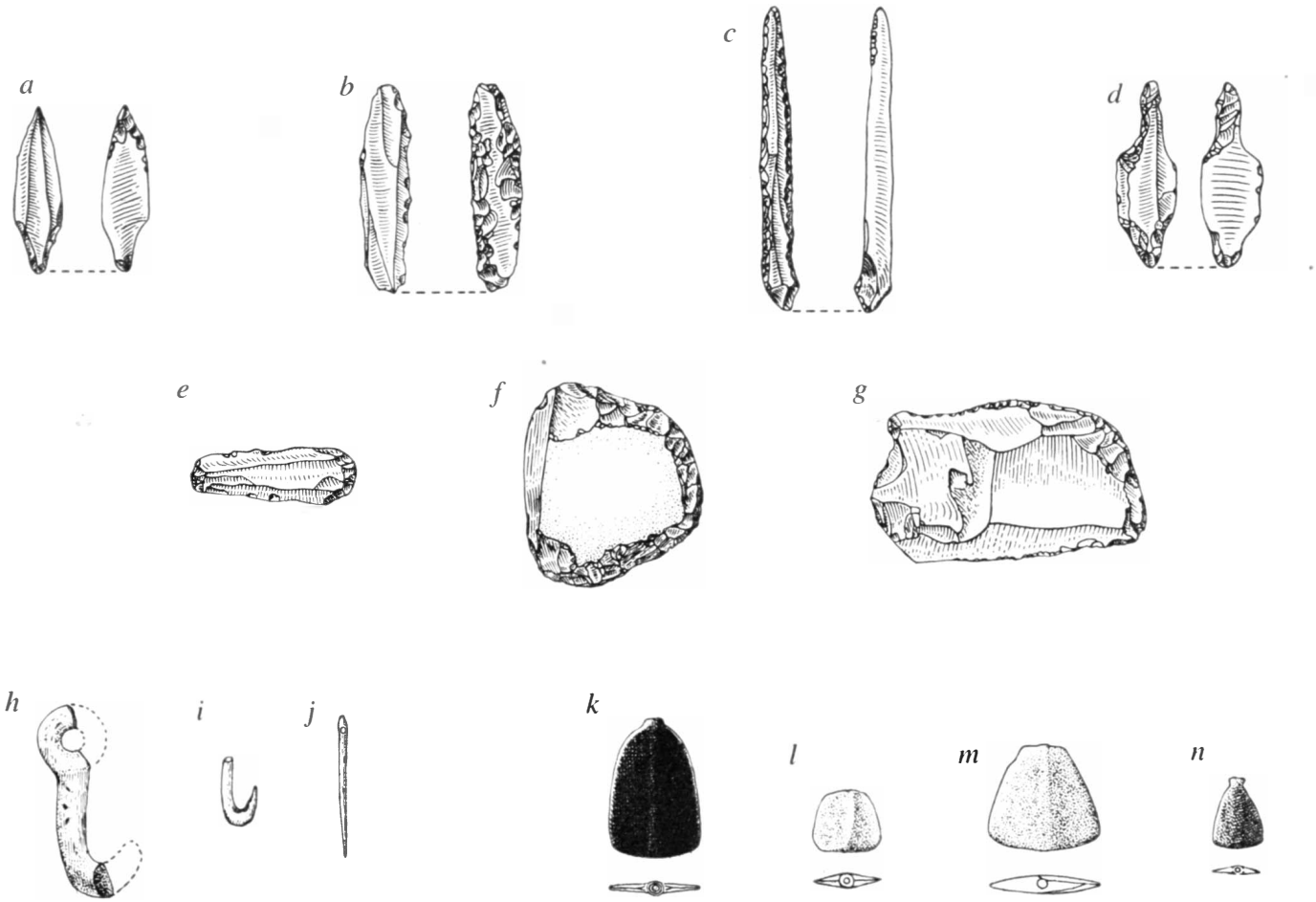
On balance, then, it seems highly likely that the inhabitants of later Mesolithic Abu Hureyra were cultivating three cereal grains (and perhaps also some pulses). This conclusion is strengthened by the similarity between the environment of Abu Hureyra, with much potentially arable land in the vicinity, and the environment of other sites of the later Mesolithic in the Levant: Dibsi Faraj East and Mureybat, sites in the Euphrates valley discovered during the same salvage campaign that saw the excavation of Abu Hureyra, Saaideh in Lebanon and Jericho in Palestine. Sites of the earlier Mesolithic only are not associated with arable land.

A second kind of similarity further strengthens the conclusion that simple husbandry was under way at Abu Hureyra. At least four tells in the Levant that were the sites of flourishing Neolithic agricultural communities have yielded evidence of habitation in later Mesolithic times. Three of them were adjacent to potentially arable land: Jericho, Mureybat and Abu Hureyra.

To turn from the plant remains at Abu Hureyra to the animal remains, Anthony J. Legge of the University of London is in the process of analyzing the two tons of animal bones from the site. His findings show that the inhabitants killed species of game ranging in size from the rabbit to the onager, the wild ass of the Levant. The most popular food animal was the gazelle, but onagers, sheep and goats were also killed regularly. The presence among the bones of freshwater-mussel shells and fish vertebrae shows that the Mesolithic population of Abu Hureyra exploited another source of food: the Euphrates.

In summary, the evidence suggests that in later Mesolithic times the inhabitants of Abu Hureyra were already cultivating cereal grains. (Their contemporaries at some other Levant sites were probably doing so too.) They also consumed many plant and animal foods gathered in the wild. Hence the basic Mesolithic tradition of hunting and gathering had here (and to some extent elsewhere) been combined with simple husbandry. The opportunity for such a practice was present in the form of potentially arable land. The motive may well have been to provide the subsistence necessary for a new kind of life, namely the settlement of an increased population in sizable villages.

Midway through the ninth millennium B.C., after several centuries of occupation, the Mesolithic village at Abu Hureyra was abandoned for reasons unknown. The site remained va-



NEOLITHIC ARTIFACTS include points for arrows (*a*), some sickle blades (*b*), reamers (*c*) and borers (*d*), some made of imported obsidian, and many scrapers (*e–g*). Bone objects include a hook-and-eye hook (*h*), a fishhook (*i*) and needles (*j*). Prominent in the graves are "butterfly" beads (*k–n*), usually made of imported agate or serpentine. Several centimeters wide, the beads are only millimeters thick.

cant for about 1,000 years before a new settlement appeared in archaic Neolithic times. By then agriculture was widely practiced both in the Levant and elsewhere in southwestern Asia and sedentary settlements were no novelty. Nevertheless, once Abu Hureyra was reoccupied (in about 7500 B.C.) it soon came to be the largest of all the archaic Neolithic settlements in the Levant. Thousands of people lived on the great tell.

The implications of so many people occupying a space of less than 30 acres are many. For one, the social relations of the inhabitants of the tell must have been considerably more complex than those in the preceding Mesolithic period; indeed, they must have been more complex than the social relations that have been assumed for early Neolithic societies. For another, regulating the affairs of such a settlement would have called for an effective form of communal organization. Hence one may infer the emergence even at this early date of social and political systems that in a more developed form were to become the essential features of the early city states of southwestern Asia, both in the Levant and in Mesopotamia.

Consider one index of interaction of major archaic Neolithic settlements such as Abu Hureyra. That index is trade. Since Abu Hureyra was an unusually large settlement by the second half of the seventh millennium, it can logically be viewed as the focal settlement of a region that embraced many smaller settlements. The archaeological evidence at Abu Hureyra suggests that the components of this regional complex were in contact not only with one another but also, and for the first time, with other regions. Abu Hureyra regularly received the volcanic glass obsidian from several sources in Turkey, along with jadeite, serpentine, agate and malachite. Soapstone came from the Zagros Mountains, turquoise from the Sinai and cowrie shells from the Mediterranean or the Red Sea. As for what Abu Hureyra and its satellite villages offered in exchange for these desirable raw materials, the archaeological record is, alas, silent.

What factors were responsible for the unusual growth of Abu Hureyra as a regional focal point? The single most important factor must have been the development in the Levant of a new kind of agricultural economy. At Abu Hureyra, as at a number of other early Neolithic sites, this new economy rested on the cultivation of cereals and pulses. Whereas it is only probable that the Mesolithic predecessors of the archaic Neolithic villagers cultivated einkorn wheat, barley and lentils, it is certain that the Neolithic villagers did. What is perhaps equally important is that they also cultivated in addition to einkorn a new kind of wheat: emmer. Although certain other crops were grown, four—two wheats, one other cereal and lentils—were the mainstay of cultivation.

Another factor in the growth of Abu Hureyra was its location at the junction of the Euphrates floodplain and the open steppe to the south. This allowed the human population to exploit the resources of two complementary environments. More than that, the people of Neolithic Abu Hureyra may have taken advantage of an accident of local topography to put under irrigation the low ground to the west of the settlement.

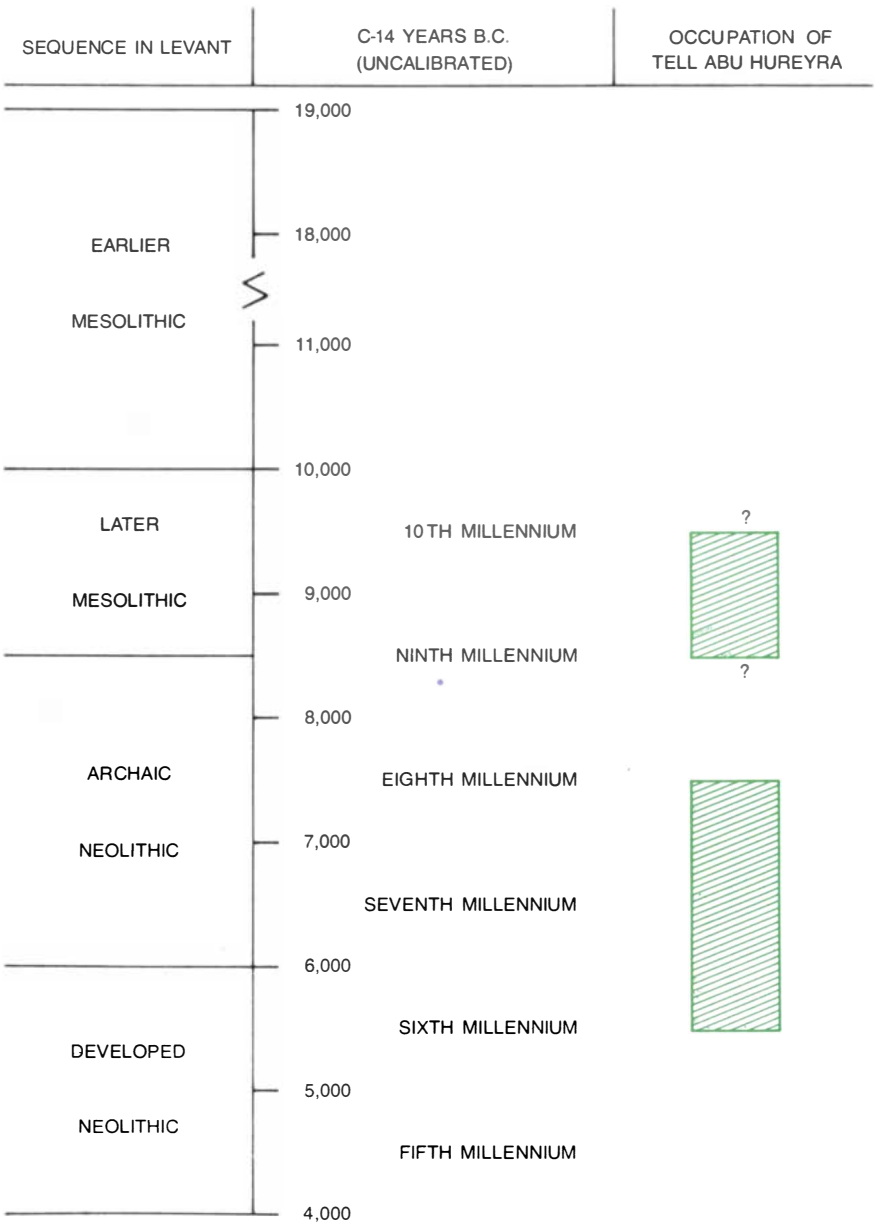
In this part of the Euphrates valley today and in the recent past irrigation

has been possible only with the aid of pumps or with canals many kilometers long. Neither of these methods was available to the archaic Neolithic population. A little to the west of Abu Hureyra, however, is a small valley, the Wadi Hibna. Today it is dry except after a heavy rain, but under the wetter climatic conditions of archaic Neolithic times it probably carried a perennial stream that might easily have been diverted. By so doing the people of Abu Hureyra could have irrigated both the bottom of the wadi and the floodplain west of their settlement, thereby substantially increasing crop yields and meeting the demands of a growing population.

As for the animal sector of the archaic Neolithic economy, the same three species that were hunted as game by the Mesolithic occupants of the site are present: gazelle, sheep and goat. Now, however, all three species were probably being herded rather than hunted. The bone remains show that at first many more gazelles than sheep or goats were being slaughtered. Then suddenly the number of gazelles declined and the number of sheep and goats increased proportionately. By the time the developed phase of the Neolithic had replaced the archaic phase, early in the sixth millennium B.C., cattle and pigs had been added to the settlement's herds and the gazelle had virtually disappeared.

For the remaining centuries of the archaic Neolithic, however, the original three animals were the major source of meat at Abu Hureyra, and the hunting and gathering of supplemental foodstuffs continued to be part of the subsistence strategy. Wild plants and wild game added to the variety of the diet, as did fish and shellfish.

What did Neolithic Abu Hureyra look like? Its numerous rectilinear houses were built close together, with only narrow lanes and little courts between them. Each house, made of mud brick, consisted of several small rooms connected by doorways; some of the doorways were conventional but others had very high sills. Many rooms had burnished plaster floors that were colored black and occasionally had red schematic motifs. The walls of the houses were covered with a mud plaster; some were given a coat of whitewash. It seems likely that each house was occupied by one family.

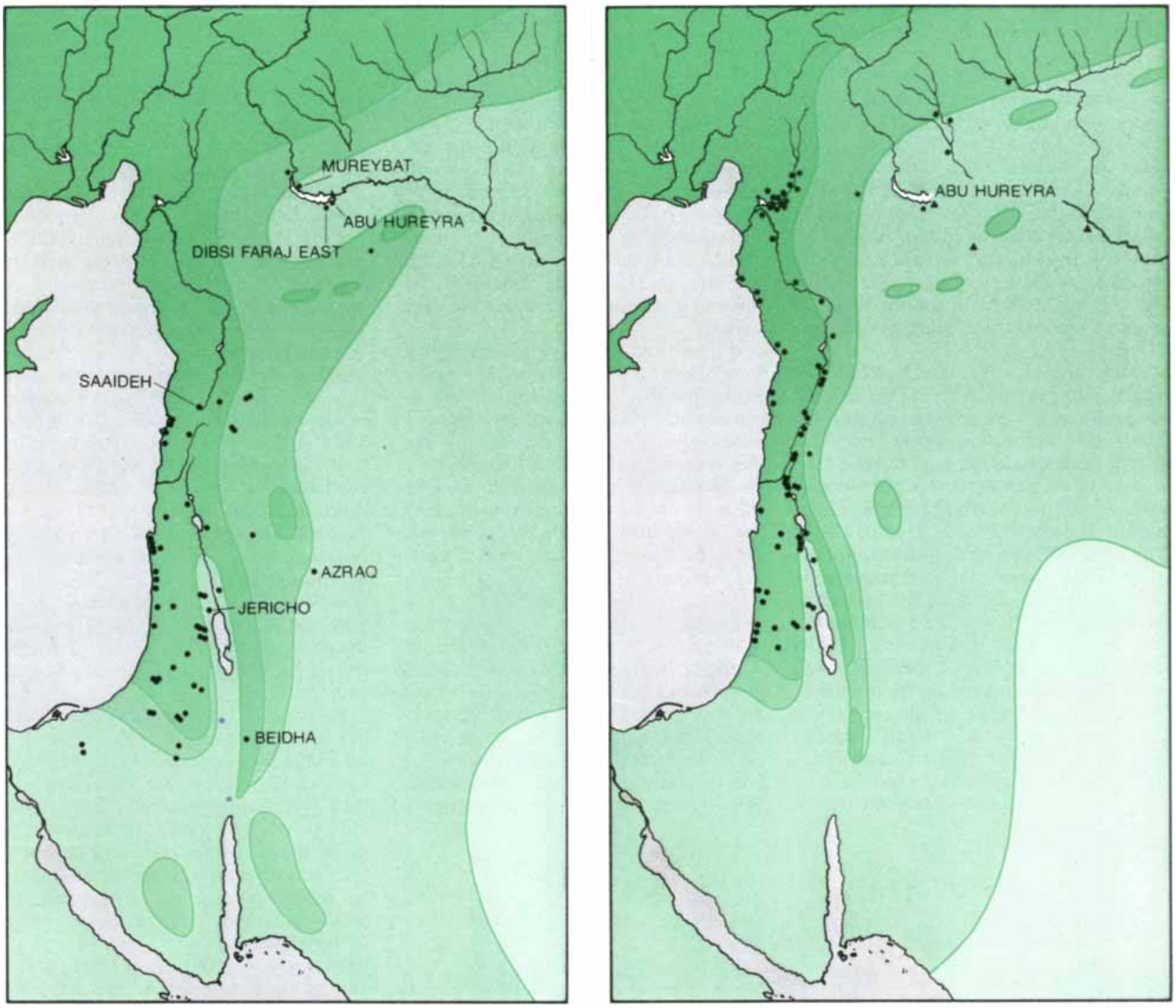


THREE-PHASE OCCUPATION of Abu Hureyra (*colored bars*) began during the later Mesolithic period in the Levant, some 11,500 years ago. The Mesolithic villagers abandoned the site for unknown reasons some 10,500 years ago. About 1,000 years later Abu Hureyra was reoccupied by a group of farmers whose subsistence strategies were characteristic of the archaic Neolithic period. The new settlement soon came to be the largest of its kind in the Levant. A transition from archaic to developed Neolithic came about 1,500 years later. The settlement was deserted not long afterward as a consequence of diminishing rainfall and ecological decline.

The artifacts made by this group included flint and, less commonly, obsidian tools with a variety of new shapes. Most were made from blanks produced by removing large blades from a "core." The commonest blade artifacts are arrowheads, knives, end scrapers (tools with the working surface at the end rather than on the side) and burins (blades flaked to retain a sharp angular projection for engraving). A smaller number of blades were made into reamers and sickle edges. The arrowheads must have been intended primarily for hunting, but we found one lodged in the breast of a human skeleton buried in the settlement, suggesting that the bow and arrow also served in combat.

Lawrence H. Keeley of the University of Illinois and Emily Moss of the University of London have examined some of these tools at high magnification for signs of wear indicating the purpose the tools served. They have been able to show that the end scrapers were used for dressing hides, the burins for engraving wood and bone and the reamers for making holes in those same materials.

The roster of flaked stone tools is completed by choppers and scrapers made from large flakes and by stone hammers. Axheads and chisels were made of stone that was not flaked but polished. The rubbing stones and milling stones used to process cereals were pecked or ground into shape. A few fine bowls were made of polished stone; some coarser stone vessels were formed by grinding. Although most Neolithic



CLIMATIC CHANGE in the Levant between the ninth millennium B.C. (*left*) and the sixth millennium (*right*) is evident from a shift in vegetation zones. Abundant rainfall in later Mesolithic times brought an eastward extension of the Mediterranean forest zone, with more than 300 millimeters of rain per year (*dark color*) and of the intermediate zone of open forest and the steppe beyond, with 300 to 150 millimeters of rain per year (*lighter colors*). Whereas most of the 70 known later Mesolithic sites (*black dots*) were in the zone of greatest rainfall or in the intermediate zone, some were in particularly favorable steppe niches. For example, Abu Hureyra and Dibsi Faraj East were on the Euphrates and Azraq was in an oasis. Three millenniums later the forest, open forest and steppe withdrew westward as rainfall diminished and the desert (*lightest color*) advanced. Even the favorably situated steppe sites (*triangles*) were then abandoned. Of the 80 known sixth-millennium Neolithic sites most were within the area most favorable for mixed farming, the zone of Mediterranean forest.

sites contain pottery, Abu Hureyra and a number of other sites in southwestern Asia show no pottery at all at this stage of their occupation. One of the main functions of pottery is the storage of food, but these Neolithic settlements apparently stored food at least one other way. They had large rectangular vessels made of thick white plaster, so heavy that they must have been permanent fixtures of the rooms in which they were found. It is noteworthy that the proportions of the various kinds of stone tools differ markedly between one part of Abu Hureyra and another, indicating that the archaic Neolithic inhabitants of the site did different things in different parts of the settlement.

The bone artifacts of the archaic Neolithic at the site repeat the Mesolithic inventory of needles, borers and spatulas. A bone fishhook and a bone hook and eye, possibly used to fasten clothing, are archaic Neolithic additions. The excavators also discovered traces of baskets and mats. These must have been among the commonest craft products of the settlement; the baskets would have complemented the rectangular plaster vessels for the storage of food.

Where the repertory of decorative objects in the Mesolithic levels of the site was limited, the number and variety of such artifacts in the archaic Neolithic levels were substantial. There are beads and pendants made of baked clay, bone and shell. Other decorations are fashioned from a variety of stones: simple limestone, various greenstones, including serpentine, and such semiprecious stones as agate, carnelian and turquoise. The finest of all these objects, nearly all of them found in graves, can be described as butterfly beads. Trapezoidal or round in form, they are generally made of serpentine or agate. The beads measure as much as five and a half centimeters in their largest dimension, but they were ground down until they are no more than a few millimeters thick.

The use of such materials, many of them imported, for the manufacture of decorative articles suggests that one incentive for the maintenance of trade with distant communities was pride of person. This suggestion is strengthened by the discovery of an end use for another imported material: malachite. We unearthed one small cockleshell that contained traces of powdered malachite,

a popular cosmetic during later periods.

The presence of beads made of baked clay among these decorations is also noteworthy. Other objects made of clay were found in the archaic Neolithic levels of the site, notably figurines in human and animal form. These clay figures, however, were not baked. Nevertheless, if the clay beads were locally made rather than imported, the inhabitants of Abu Hureyra were familiar with two key properties of clay that should have brought them to the threshold of pottery manufacture: plasticity and fire hardening.

The archaic Neolithic settlement reached its maximum size in the seventh millennium B.C. Its growth reflected the sharp increase in the population of the Levant as a whole that accompanied the spread of the new agricultural economy. The occupation of Abu Hureyra continued after the start of the developed Neolithic period in the sixth millennium, but the village shrank to little more than half its former size. Like the preceding settlement, the developed Neolithic village was also built of mud brick. Little is known about its layout, because it has been almost destroyed by thousands of years of exposure to the elements. The buildings seem to have hardly changed in plan from those of the archaic Neolithic, but their inhabitants took to digging numerous pits among the structures. These they used for a time as hearths and then filled with rubbish.

The artifacts made by the last inhabitants of Abu Hureyra were little different from those of their archaic Neolithic predecessors. One exception is that pottery now appears. Even that innovation seems to have been modest in scale; very few potsherds were found in this part of the excavation. Eventually, in about 5500 B.C. or perhaps a little earlier, the site was abandoned. It remained uninhabited until our group arrived to excavate it.

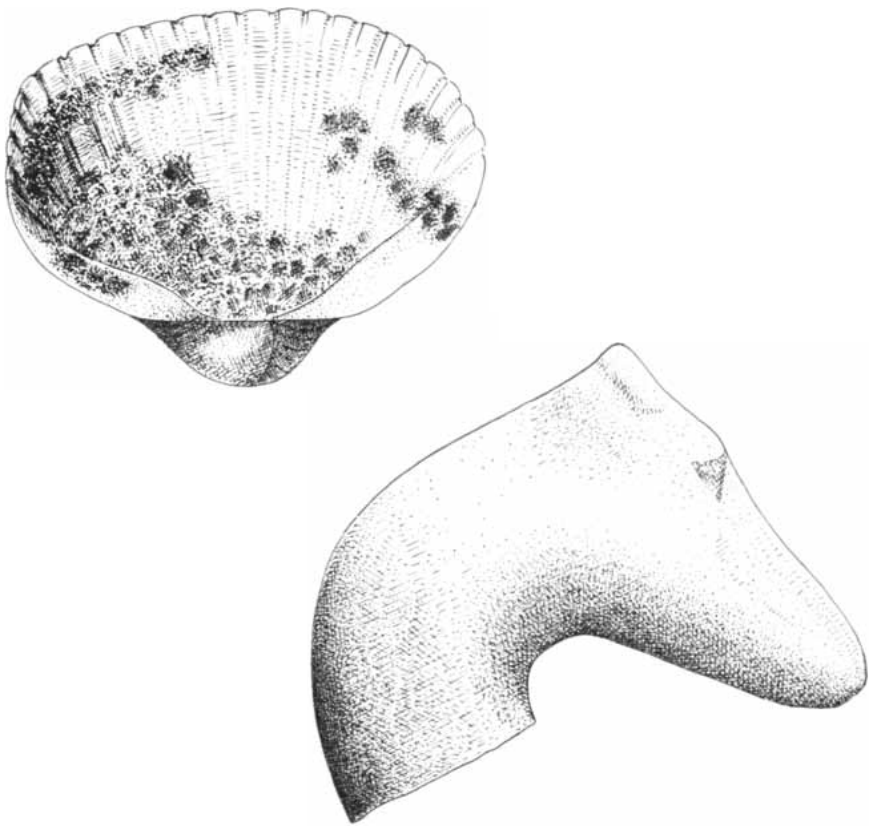
Why was Abu Hureyra deserted by its inhabitants? The results of the excavation show that over the long life of the settlement its agricultural economy was modified to become increasingly productive. One might seek the answer to the abandonment in purely local terms if it were not for the fact that throughout the Levant other sites in the steppe border zone were also abandoned at about the same time.

I suggest two reasons for these abandonments. The first is that from late in the seventh millennium the overall environment of the Levant was becoming less attractive to a farming population. The temperature had been rising since the retreat of the Pleistocene ice, but the rainfall was now diminishing. The change brought about a gradual contraction of the zone of Mediterranean forest. The steppe expanded into the forested areas, and the desert in turn began to encroach on the steppe. Thus as the Levant became more arid the pattern of settlement along the border of the steppe was affected.

The second cause of the abandonments, in my view, is that the growth of population during the archaic Neolithic put too great a strain on the delicate balance of the overall economy. As we have seen, that economy had a dual character. The yields of agriculture and animal husbandry provided for much of the population's needs, but at the same time the exploitation of wild resources made its own important contribution. The plant remains at Abu Hureyra show that the natural vegetation (and therefore the soil cover) became steadily more degraded over the life of the settlement. The inhabitants were evidently overgrazing their pasture and overexploiting the natural vegetation as a source of food and fuel. It seems probable that they also seriously reduced the game population at a time when the change in climate was accelerating the same process.

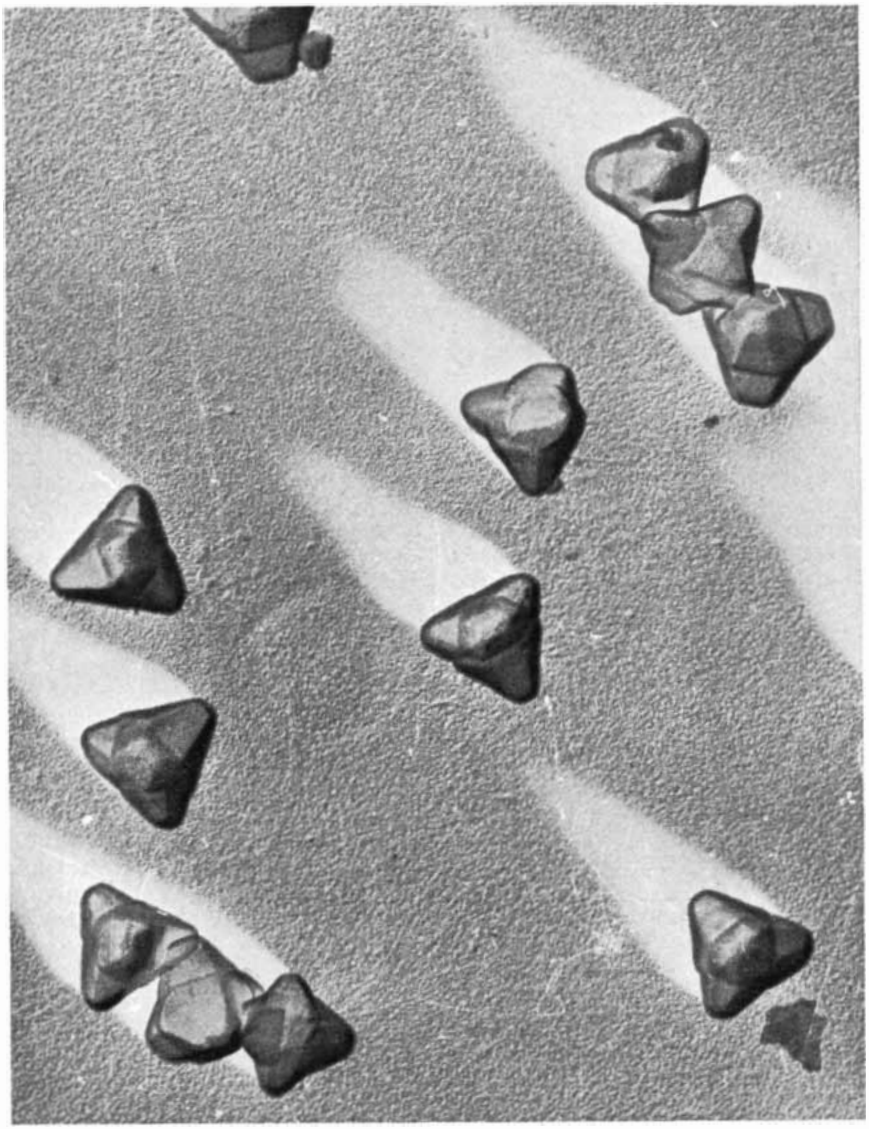
Early in the sixth millennium, possibly in response to the dual pressures of an altered environment and an upset ecological balance, the inhabitants of the Levant settlements, Abu Hureyra included, ceased to rely on hunting and gathering as an important supplement to the farming economy and came instead to depend almost exclusively on the produce derived from mixed farming. In their planting they concentrated on the more productive of the cereal grains and on those pulses that like the lentil were an abundant source of beans and peas. Their herds now included cattle and pigs in addition to the ubiquitous sheep and goats.

This kind of mixed-farming economy was far better suited to those parts of the Levant where the Mediterranean forest still existed, namely the well-watered coast and the northern plains. The inhabitants of the once thriving steppe-border villages gave up their outworn settlements and moved north and west to build new homes in more promising surroundings. The fact that the overall population of the Levant in the sixth millennium B.C. remained about the same as it was at the end of the expansionist seventh millennium indicates that the new adjustment was a successful one, whatever the costs of the dislocation may have been. Although the Neolithic period was already well advanced in the Levant before a mixed-farming economy was developed, the ultimate adoption of that economy early in the sixth millennium B.C. was so successful that it remained the basis of human life in the region until modern times.



ORNAMENTATION in Neolithic Abu Hureyra is manifested by a cockleshell (*left*) that held powdered malachite, an imported cosmetic, and by a miniature gazelle head made of granite.

AgI crystals



If our studies on them should happen to shed new light on how tooth, bone, and snow form, it would be nice

Our purpose here is not at all devious. It is simply to incline you favorably toward our trademark when buying film.

When you plunk down the money for a cartridge of Kodak film, a very small fraction of one of the pennies goes to pay Richard L. Daubendiek's salary and the expenses of his trip last year to Copenhagen to lecture an international conference on "AgI Precipitations: Effects of pAg on Crystal Growth." The professors and sundry academic types from various countries who attend such a conference couldn't care less about how we make film. Nor could Dr. Daubendiek tell them even if he knew all about it.

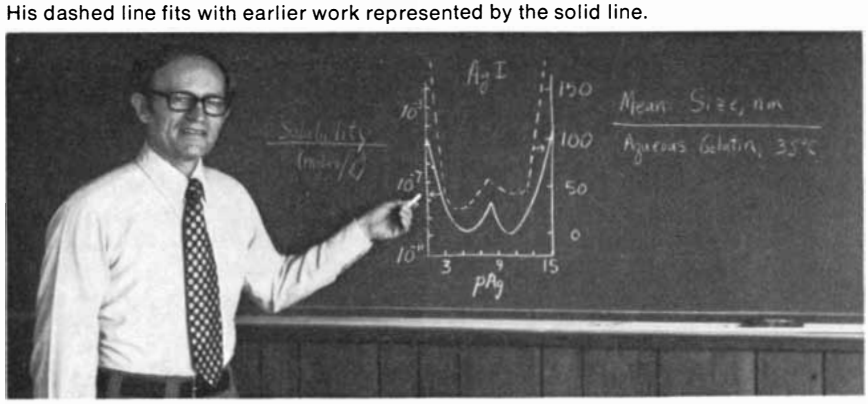
What Dick Daubendiek and a few of his associates do know a lot about is the traffic in ions between crystals and the solutions they sit in. Since our business has rested largely on the growing and behavior of small silver halide crystals, we feel we (and you) can afford Dick.

Although pure silver iodide crystals began to fade in importance as the basis of photography after Daguerre settled down with his 6,000 francs a year for life from Louis Philippe, King of the French, they subsequently turned out to be just right for studies of nature's rules for assembling and disassembling crystals of all kinds–biological, meteorological, or merely industrial.

Now suppose you arrange for a jet of silver ions to meet a jet of iodide ions in an aqueous solution of gelatin at 35°C, and suppose you vary their relative concentrations over a broad range. The size of the crystals that precipitate can be measured by electron micrography from carbon replicas shadowed with platinum-palladium, like the ones shown above that were formed at pAg 6.5 and consist of four hexagonal pyramids each. If you plot their mean size against pAg in a constant ionic environment, as Daubendiek has done here, why do you suppose you get that sharp cusp?

Kodak

His dashed line fits with earlier work represented by the solid line.



SCIENCE AND THE CITIZEN

By Jove

The rich harvest of pictures and data from *Voyager 1* following its close encounter with Jupiter on March 5 is now being supplemented by a second flood of information from *Voyager 2,* which made its closest approach to the planet on July 9. Before proceeding on to Saturn *Voyager 1* returned some 18,000 pictures of Jupiter and its moons. From 60 to 12 days before encounter the entire disk of the planet was photographed in color five times during each Jovian revolution, showing in great detail the evolution of cloud patterns in the planet's swirling multicolored atmosphere. In the dozen days before closest approach, when the cameras were too near for whole-disk mosaics, selected features were photographed with a resolution as high as six kilometers per pair of scanning lines. *Voyager 1*'s cameras and instruments disclosed such a wealth of unexpected findings that according to one team of investigators "the sense of novelty would probably not have been greater had we explored a different solar system." At periapsis, or closest approach, *Voyager 1* was 348,890 kilometers from the center of Jupiter, or some 276,000 kilometers from the top of the planet's clouds.

The spacecraft was targeted to photograph five of Jupiter's 12 satellites: tiny innermost Amalthea (discovered in 1892) and the four Galilean satellites (first seen by Galileo in 1610): Io, Europa, Ganymede and Callisto. Ellipsoidal Amalthea measures about 265 kilometers by 140. The four Galilean satellites are markedly different in appearance. Ganymede, Jupiter's largest satellite, has such a low density that it is probably 50 percent water. Callisto, slightly smaller, much darker and even less dense than Ganymede, presumably has a still higher fraction of water. Its most striking feature is a circular impact basin larger than any seen on the earth's moon or on Mercury. Europa was photographed from the greatest distance and therefore with the poorest resolution. Nevertheless, its appearance is also distinctive. Although it is lacking in obvious impact features, it exhibits an enormous pattern of intersecting lines from 50 to 200 kilometers wide and as much as thousands of kilometers long.

Io is far and away the strangest object yet seen in the solar system. Slightly larger than the earth's moon, it is devoid of craters and other moonlike features. Its heavily mottled surface is a crazy quilt of red, orange, brown, gray and near-white. Judging from the lack of craters Io is evidently resurfaced by erosion or deposition every 10 million years or less. The resurfacing is probably explained by the most spectacular finding of the *Voyager 1* mission: the discovery on Io of at least seven active volcanoes. Gas vented from the volcanoes is undoubtedly rich in sulfur or sulfur compounds. In its various molecular forms sulfur can account for the entire range of colors seen on Io's surface.

Volcanic gases can also account for a doughnut-shaped ring of neutral sodium gas that fills the orbit of Io, a discovery made with terrestrial instruments in 1973. *Voyager 1*'s instruments showed that an intersecting torus of ionized gas at the orbital distance of Io is rich in ionized sulfur and oxygen. Because the plasma torus is under the influence of Jupiter's magnetic field, whose equator does not exactly coincide with that of the planet itself, at the distance of Io the plasma torus intersects the torus of neutral sodium at a slight angle.

According to one hypothesis the internal heating necessary to keep Io in a constant state of volcanic activity could be supplied by tidal forces. On this model Europa and Ganymede induce a wobble in Io's orbit, which causes a variation in the amplitude of the large fixed tides generated by Io's synchronous rotation about Jupiter. (In synchronous rotation one face of a satellite always faces the satellite's planet.) Another source of heat has now been suggested by *Voyager 1* measurements of intense electric currents flowing in or near the magnetic "flux tube" that links Io with Jupiter's magnetic field. The flux tube is simply the ensemble of the lines of magnetic force from Jupiter's magnetosphere that pass through Io.

It had been postulated some 10 years ago that as the magnetic lines of force sweep through Io they generate a current, thereby making the satellite a giant electric-power station in space. *Voyager 1* was targeted to pass through the Io flux tube and measure the current. It turned out that the current flow is so large that it displaces the flux tube slightly from its predicted position, with the result that the spacecraft only skirted the tube. Nevertheless, a current flow of about five million amperes can be inferred. The current loop set up by Io's interaction with Jupiter's magnetosphere implies that 10^{12} watts of heat must be continuously dissipated near or within the satellite. This is an amount of heat equivalent to that calculated for tidal forces acting on Io. Some of the flux-tube current would also be available for heating the Io plasma torus.

The most intense heat registered by *Voyager 1*'s instruments is also the highest temperature ever recorded in the solar system. As the spacecraft entered the outer reaches of Jupiter's magnetosphere, 4.9 million kilometers from the planet, it encountered a thin plasma with a temperature of 300 million degrees Kelvin. Analysis showed that the plasma consists primarily of ions of hydrogen, helium, oxygen and sulfur, with lesser amounts of sodium. It seems probable that much of the sulfur and sodium and some of the oxygen originates in volcanic emissions from Io. The gases are presumably ionized and hurled out into the Jovian magnetosphere.

Conspicuous Consumption

The portion of the Federal budget dedicated to military expenditures is expected to increase by approximately $10 billion in the next fiscal year, continuing the upward trend of the past few years. The total amount of money requested for the Department of Defense and for the military programs of the Department of Energy in the Carter Administration's proposed budget for fiscal 1980 now exceeds $138 billion. Measured either in current dollars or in constant (inflation-adjusted) dollars, that means the military budget will have increased each year since 1976, when President Carter was elected.

The 1980 military budget has been characterized as "the NATO budget" in view of the fact that the Administration and the governments of the other nations that make up the North Atlantic Treaty Organization have made a commitment to raise their total military outlay by at least 3 percent over and above the prevailing inflation rate in the coming year. Actually, however, the big increases in the new budget are mainly in the area of nuclear weapons. According to *The Defense Monitor,* the newsletter of the Washington-based Center for Defense Information, "the NATO emphasis, while still very costly, seems to be fading as a rationale for big military budgets. Despite the commitment of President Carter to lessen U.S. reliance on nuclear weapons, a nuclear buildup has now become the focus of spending increases." Funding for conventional military forces, measured in constant dollars, will actually decline slightly in the 1980 budget, whereas funding for strategic nuclear-weapons programs will climb by some 19 percent (again in constant dollars). In order to provide warheads for the new generation of nuclear weapons funding for the production of nuclear explosives (a Department of Energy program) has more than doubled since the Carter Administration took office.

Among the new nuclear-weapons programs that the newsletter lists as "obvious targets for those who want to control escalating military expenditures"

Mumm's the word.

For 150 years, people who know how to live have been celebrating life with Mumm premium French Champagne.

are the MX missile ($670 million in the fiscal 1980 budget), the Trident submarine and Trident II missile ($1.5 billion), the Pershing II missile ($145 million), the sea-launched cruise missile ($107 million), a variety of antisatellite programs (about $80 million) and the planned modernization of nuclear-weapons production facilities ($81 million).

Other recommended "opportunities for cost savings in the fiscal 1980 military budget" include cutbacks in such programs as the CVV medium aircraft carrier ($1.6 billion), the advanced attack helicopter, or AAH ($176 million), the SSN-688 nuclear attack submarine ($523 million) and the Patriot antiaircraft missile ($569 million). Reimbursements for U.S. forces in Japan ($1.1 billion) and in West Germany (at least $1.1 billion) could add to the savings.

The independent analysts of the Center for Defense Information estimate that a total of $10 billion could be saved from the fiscal 1980 budget "without reducing the security of the United States." In sum, they write, "controlling the nuclear arms race, through SALT agreements and by stopping unneeded programs, and controlling unnecessary Federal spending are becoming more and more overlapping objectives. Slowing spiraling government budgets requires slowing the spiraling nuclear arms buildup. We will be improving our security through both steps."

Continuous Coal

With the apparent inevitability of a greater dependence on coal for energy in the U.S., increasing attention is being paid to the technology and economics of this fossil fuel. The Carter Administration's energy plan calls for a doubling of coal production in the U.S. (to 1.2 billion tons per year) by 1985. To reach even a fraction of that increase will require a huge development effort: new mines, more miners and an expanded system for transporting coal to the plants (mainly stations generating electricity) where it will be burned or converted into liquid or gaseous forms. Most of the coal is now carried by railroads, and they have taken the position that they will be able to keep up with any increase in the traffic. An alternative that is receiving much consideration is the transportation of coal as a slurry (a mixture of powdered coal and water) in pipelines. It is argued that a pipeline system would ensure the movement of coal without excessive strain on the existing transportation system as production increases and that pipelines, by competing with the railroads, would help to keep railroad rates down. The subject is reviewed by the General Accounting Office in a report, "Coal Slurry Pipelines: Progress and Problems for New Ones."

The progress can be summed up quickly: two coal-slurry pipelines have been built and seven are in the planning stage. One of the existing lines—the 108-mile Ohio pipeline from West Virginia to Cleveland—was shut down after competing railroads reduced their rates enough to undercut it. The line still in operation is the Black Mesa pipeline, which runs for 273 miles from Arizona to Nevada and has the capacity to move 4.8 million tons of coal per year. The most heroic of the lines in the planning stage is the only one in the East: the Florida pipeline, which would run for 1,500 miles from eastern Kentucky to southern Florida with a capacity of 53 million tons of coal per year. The other lines, ranging in length from 183 to 1,378 miles and in capacity from 10 to 26 million tons, would originate at Western mines and run west or south.

The two major problems are water and rights-of-way. It takes a ton of water to move a ton of coal. Most of the projected pipelines originate in the arid West, where water use is a frequently contested subject. According to the GAO report "most sources agree that there is presently enough unused water physically available," but in some instances "it is not legally available." The solution appears to lie in the fact that pipeline companies can "look beyond... conventional sources to water that is unusable or too expensive for others to consider."

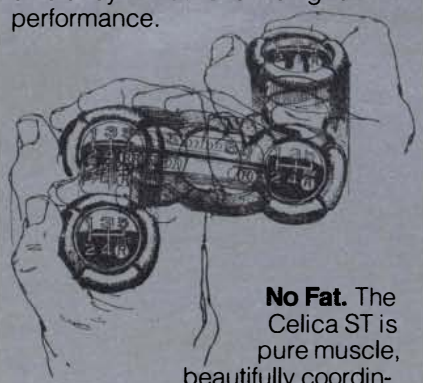
The right-of-way problem is most acute in the East, where much of the land is privately owned and the railroads will usually not let a pipeline cross their right-of-way. The pipeline developers have therefore been seeking eminent-domain legislation, whereby states or the Federal Government can compel private owners to provide rights-of-way for pipelines. Eight states have adopted such laws, but a proposed Federal law, the Coal Pipeline Act of 1978, was defeated in Congress.

Hard Lines

The electromagnetic radiation emitted by celestial processes includes not only radio waves, infrared radiation, light, ultraviolet radiation and X rays but also gamma rays. Although processes that emit gamma rays have been known for some time, it is only recently that it has been possible to detect spectral lines in these "hardest," or most energetic, of electromagnetic waves from space. Sensitive detectors for the purpose have been developed that can be lifted in a balloon or satellite above most of the earth's atmosphere, which is quite opaque to gamma rays. Spectroscopic studies of gamma rays are expected to yield basic information about highly energetic phenomena in the sun and other stars, including supernovas. According to a review article in a recent issue of *Nature,* although these studies

MAKE HASTE, NOT WASTE.

Even with Toyota's lowest priced Celica, you can get clean efficiency without sacrificing real performance.



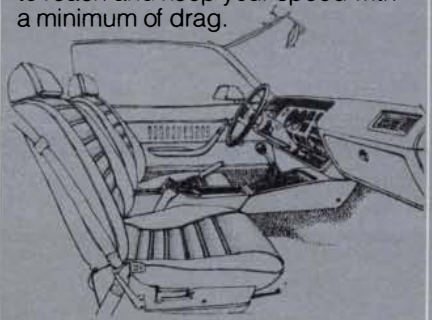
No Fat. The Celica ST is pure muscle, beautifully coordinated by its standard 5-speed overdrive gear box. As *Road & Track* test figures show: the Celica accelerates from 0-60 in 10.4 seconds flat. And can stop in a clean 270 feet from a speed of 80 mph.

Not only won't you get this kind of clean performing precision out of many other sports coupes—you won't even get it out of the much more expensive Porsche 924!

No Waste. Like every Toyota, the Celica doesn't waste an ounce of gas. Its engine burns and energizes every drop of fuel you put into it—efficiently and cleanly.*

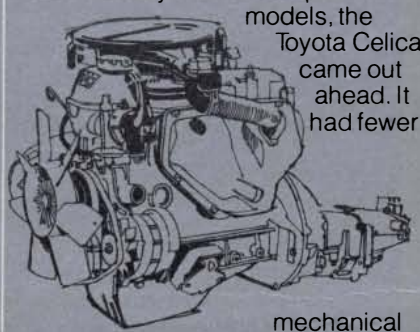
Great Shape. And look at the way the Celica has been designed. Purposeful, not wasteful.

Outside, the lines are so clean, their aerodynamic shape allows you to reach and keep your speed with a minimum of drag.

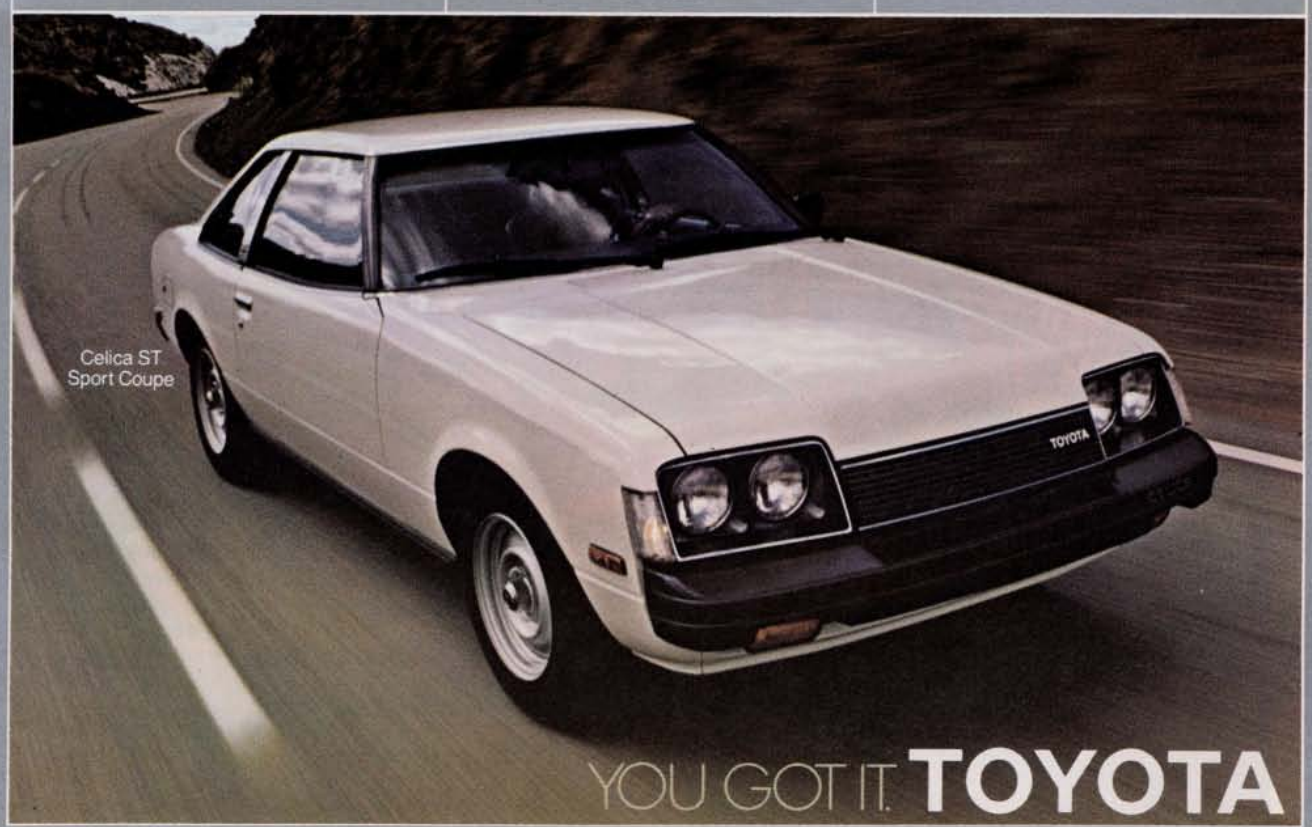


While inside, the clean design is apparent in everything you can see and touch—from thoughtful instrumentation to reclining front bucket seats.

Staying in Shape. Based on an owner's study with 10 competitive models, the Toyota Celica came out ahead. It had fewer



mechanical problems—which is bound to save you money. So will the fact you can get all this efficiency and performance out of Toyota's lowest priced Celica. The Toyota Celica ST—Make haste!



* Based on Clean Air Act Standards for 1979

Guess who's coming to class. The Compucolor II.

If the face doesn't look familiar, you really should get to know the Compucolor II. It's a microcomputer with vivid color graphic displays. And it's flexible enough to fit into any classroom situation.

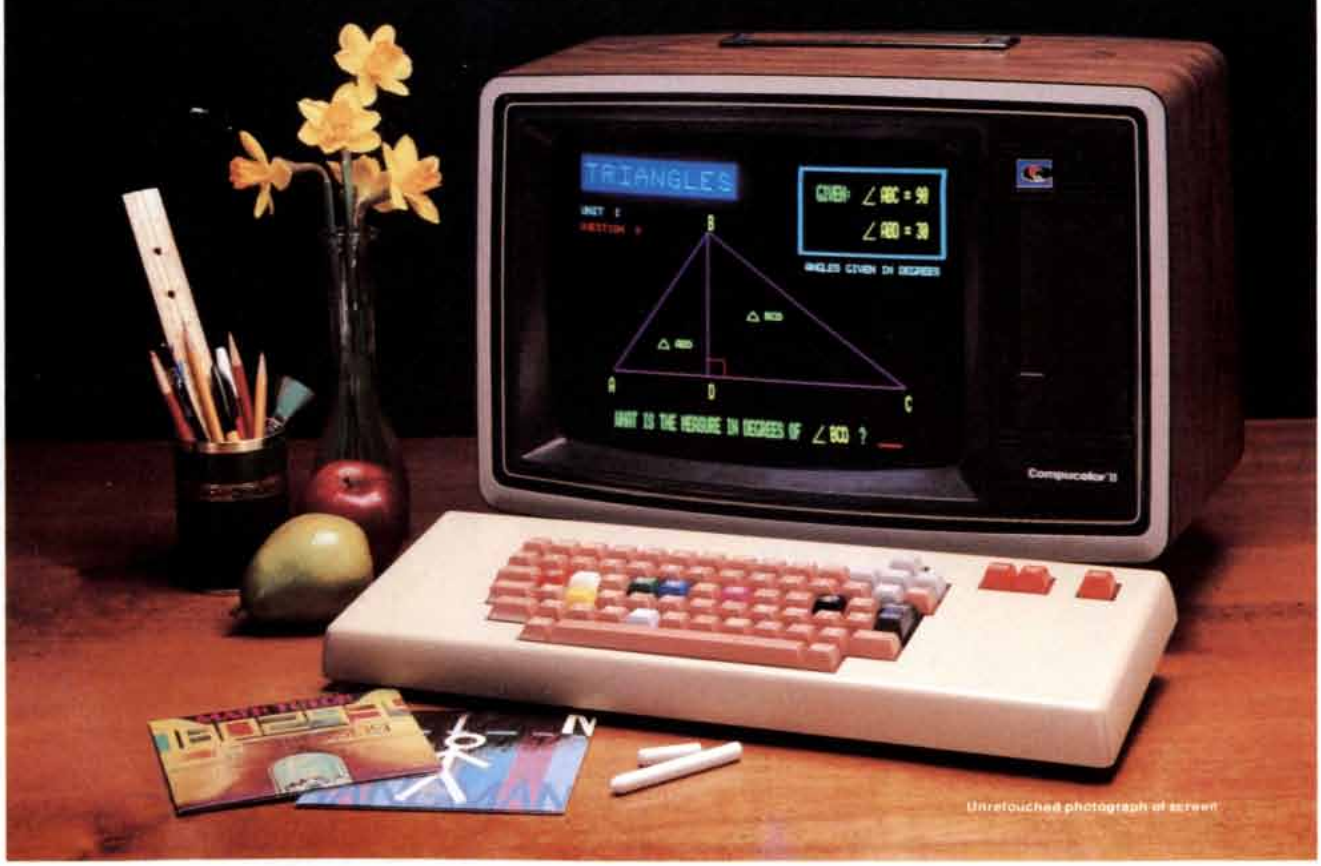
The Compucolor II is completely programmable in BASIC, the easy-to-learn computer language. Even very young students can actually learn to program the computer. Or they can use a pre-programmed Sof-Disk ™ for individualized instruction, sharpening their skills at their own pace while teachers are free to spend more time explaining important concepts.

The Compucolor II is completely portable and self-contained. No clumsy peripherals, no complicated interface and no expensive time-sharing contracts. Just plug it into any standard outlet and it's ready to quiz your students on everything from trigonometry to vocabulary. After the bus leaves, the Compucolor II helps you record grades, figure averages and schedule assignments.

Even with all these features, the price is very competitive. Just *$1495 includes everything you need to introduce your students to tomorrow. Contact your nearest computer store for a demonstration. Once you see it, you'll want the Compucolor II coming to your class.

Compucolor® Corporation

*U.S. Domestic price



Unretouched photograph of screen

are only in their infancy, gamma-ray lines have already been found in solar flares, in the center of our galaxy, in the radio-wave-emitting galaxy Centaurus *A* and in an unidentified transient event in the direction of the Crab Nebula.

Cosmic gamma-ray lines were first observed in solar flares. The source of the lines is probably the nuclear interactions of the particles that the flares accelerate in the chromosphere and lower corona. The strongest line, at a wavelength equivalent to an energy of 2.223 million electron volts (MeV), is the result of the capture of a neutron by a nucleus of hydrogen. Two lines, at 7.632 and 7.646 MeV, are the result of the capture of a neutron by a nucleus of iron. The second-strongest line, at .511 MeV, is due to the mutual annihilation of a positron (a positively charged electron) and an electron.

The .511-MeV line has also been observed at the center of our galaxy, write M. Leventhal of Bell Laboratories and C. J. MacCallum and P. D. Stang of Sandia Laboratories in *Astrophysical Journal Letters*. The line, whose intensity is five times greater than that of the background fluctuations, was spotted by a large detector flown in a balloon over Alice Springs in Australia, which lies directly under the celestial path of the center of our galaxy. Because the angular resolution of the detector was only 15 degrees the line cannot be linked to a particular source of radiation. If the source is exactly at the center of the galaxy, it is giving off a tremendous amount of energy: about 10^{30} joules per second (one joule being the energy needed to lift one kilogram 10 centimeters). This amount of energy is several thousand times greater than the total energy output per second of the sun.

The positrons whose annihilation gives rise to the .511-MeV line may have been created in the nuclear reactions of supernova explosions. The explosive death of a massive star hurls into the interstellar medium nuclei of nickel, cobalt, sodium, titanium, iron and aluminum. As some of these nuclei decay radioactively, positrons are given off. Another possible source of positrons is the magnetosphere of a pulsar, although observations of one particular pulsar do not reveal any gamma-ray lines.

Gamma-ray line emission has also been associated with various transient events. Intense bursts of gamma rays, some lasting for perhaps 10 seconds, have been known for some years, although they have not been examined with enough resolution to reveal whether the gamma rays are emitted in lines or in a continuum. Another kind of transient event, however, has all its observed radiation in the form of lines. Such an event, lasting for about 20 minutes, was found in the direction of the Crab Nebula by a balloon-borne detector on June 10, 1974. The intense gam-

King Carl XVI Gustav of Sweden—on the right in these photos—presents the 1978 Nobel Prize in Physics to Bell Laboratories scientists Robert Wilson (top photo) and Arno Penzias.

What does the Nobel Prize have to do with your telephone?

The two scientists on the opposite page are receiving the highest honor a scientist can earn—the Nobel Prize. They are the sixth and seventh laureates who did their prize-winning research at Bell Telephone Laboratories. These scientists shared a common goal—the search for new knowledge to further advance the art of telecommunications.

Clinton Davisson shared the Nobel Prize in 1937 for demonstrating the wave nature of matter. In 1956, John Bardeen, Walter Brattain and William Shockley were honored for their invention of the transistor. Philip Anderson's theoretical work on amorphous materials (such as glass) and on magnetism led to a Nobel Prize in 1977. And in 1978, Arno Penzias and Robert Wilson received the Prize for detecting the faint radiation from the "big bang" explosion that gave birth to the universe some 18 billion years ago.

The search for knowledge

These scientists and their colleagues at Bell Labs, given the freedom to explore, have proved time and again the value of investment in research—not only for telecommunications but for society in general. The transistor, for example, revolutionized communications and brought into being entire new industries—indeed, a new industrial society—based on solid-state electronics.

Other Bell Labs advances—products of this same research environment—have included high-fidelity recording, sound motion pictures, long-distance television transmission in the United States, the electrical digital computer, information theory, the silicon solar cell, and the laser. The impact of this work—on almost every field of commerce, industry, education and even medicine—has been incalculable.

The innovation process

Research done at Bell Labs in the past is the basis for the products and services the Bell System offers its customers today, just as the research going on now is the foundation for tomorrow's telecommunications.

Bell Labs scientists—specialists in physics, chemistry, mathematics and many other disciplines—team their efforts with those of our systems, development and design engineers. They, in turn, work closely with Western Electric manufacturing engineers and with the people of the Bell System operating telephone companies.

This technical integration is the foundation for true innovation. One idea feeds another. A basic scientific discovery can make possible entire new technologies and products for telecommunications, and a concept for a new product or system can stimulate the research to find even more new knowledge. That interaction, that teamwork, has been extremely productive: Bell Labs people have received 18,645 patents between our founding in 1925 and the end of 1978.

Sometimes, the search for knowledge may lead to a Nobel Prize. Often, it benefits all of society. And always, its ultimate aim is better service for Bell System customers.

Bell Laboratories
600 Mountain Avenue
Murray Hill, N.J. 07974

Bell Laboratories

Keeping your communications system the best in the world.

ma-ray emission fell in four comparatively narrow energy bands. The radiation was strongly red-shifted, that is, its wavelength was increased by the loss of energy in moving against a powerful gravitational field. This suggests that the transient emission is associated with an object having an ultrastrong gravitational field, such as a neutron star or a black hole.

Second Bass

As any dog trainer knows, sounds that are too high for man to hear are perfectly audible to dogs. The upper limit of human hearing is about 20,000 hertz (cycles per second), of canine hearing about 40,000 hertz and of bat and mouse hearing about 80,000 hertz. The lower limit of human hearing is about 125 hertz at normal levels of intensity. Are any animals able to hear lower sounds than man can?

The answer is yes, according to Rickye and Henry Heffner of the University of Kansas at Parsons and Ned Stichman of the Ralph Mitchell Zoo in Independence, Kan. Indeed, it seems likely that all animals with a head larger than man's have this ability. The three investigators recently tested the hearing range of a seven-year-old Indian elephant in the Independence zoo. Reporting on their work at a recent meeting of the Acoustical Society of America, they note that their subject, Lois, could not hear sounds higher than 12,000 hertz but could hear sounds as low as 16 hertz at levels of intensity that were inaudible to human beings. The investigators conclude that the range of hearing in animals is in inverse proportion not to the size of the ear but to the size of the skull, which determines the distance between the ears.

The testing technique used with Lois involved a reward when the elephant correctly punched a "yes" or "no" button with her trunk, depending on whether a sound was audible or inaudible. In the course of their experiment Stichman and the Heffners rewarded Lois with a total of eight gallons of watered-down Kool-Aid, half a cup at a time.

McIlroy Was Here

Most people go to the hospital only when they have to, but a few individuals have a strange compulsion to go repeatedly, even if they have to fake the symptoms of illness. This type of pathological lying is common enough to have a name: Munchausen syndrome (after the classic collection of tall tales titled *The Adventures of Baron Munchausen*). The most remarkable case history to date is that of Stewart McIlroy, whose exploits are chronicled by C. A. Pallis and A. N. Bamji in a recent issue of *British Medical Journal*.

McIlroy's success as an imaginary invalid owed much to minimal signs of neurological disorder: a slight wasting of the muscles in the left hand, a lack of reflexes in the arms and a loss of sensation in areas scattered over the arms and upper trunk. Beyond this McIlroy had a genius for simulating symptoms, appearing on various occasions to be suffering from a collapsed lung, a variety of dramatic neurological disorders, acute respiratory failure (which necessitated a permanent tracheotomy and a silver speaking tube), severe abdominal pain, recurrent chest pain and retention of urine. All these symptoms appear to have been completely counterfeited.

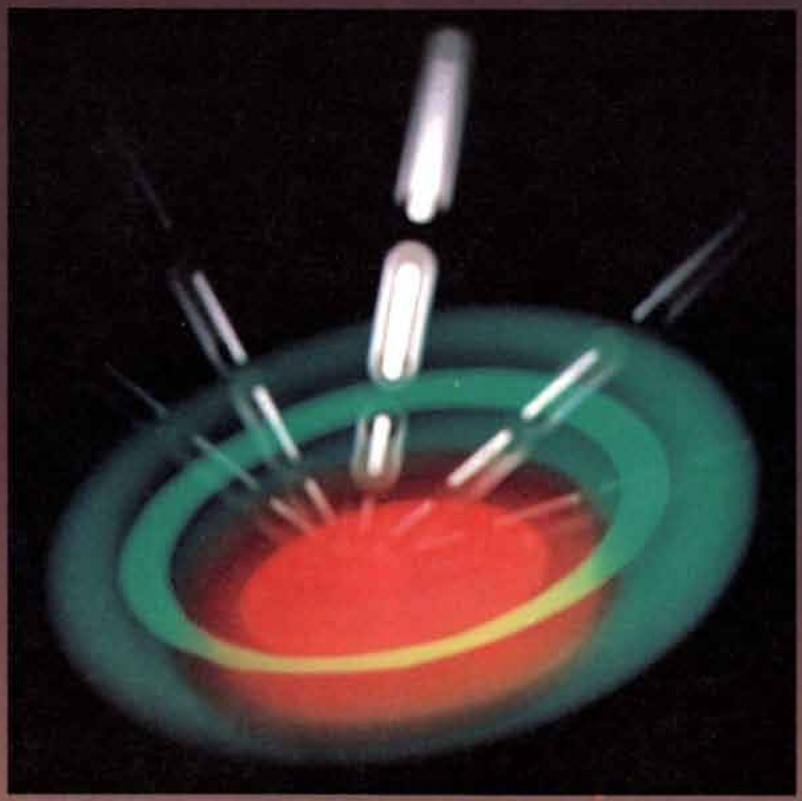
Beginning in 1944, McIlroy spent more than 10 of the next 34 years as a patient in at least 68 different hospitals throughout the British Isles: eight hospitals in Northern Ireland, seven in the Republic of Ireland, eight in Scotland, two in Wales and 45 in England. At least 207 admissions have been documented, and there is circumstantial evidence for another 10. In his long career McIlroy cost the National Health Service a sum running into millions of pounds and perhaps tens of millions.

McIlroy's wide experience with hospitals gave him a remarkable grasp of medical technology and a considerable sophistication in the simulation of various diseases. He developed the giving of his medical history into a fine art, including "facts" going beyond his own history; on several occasions he mentioned the death of a sister, giving one of three different causes. When one array of symptoms was threatened with exposure, he would switch to another, and if a showdown seemed imminent, he would quickly decamp, his paraplegic or hemiplegic disabilities miraculously cured. He discharged himself from the hospital a total of 133 times, managing to escape detection by employing 22 different last names and eight first names in different hospitals.

Quite apart from bed and board, McIlroy cost taxpayers dearly in diagnostic and operative procedures. His skull showed surgical burr holes from early in his career, and his numerous scars bore witness to a variety of orthopedic procedures and abdominal operations. The number of ordinary X-ray examinations and blood tests must have numbered in the hundreds if not thousands, and he underwent at least 48 spinal taps, three air encephalograms and numerous myelograms. His collected medical notes occupy four inches of shelf space.

In 1978 McIlroy disappeared, and inasmuch as it has been more than a year since he was last seen, he is presumed to be dead. Pallis and Bamji conclude that the fact McIlroy lived as long as he did "bears testimony to the resilience of the human frame and to the relative safety of our hospitals." They add that his career has "taught many lessons to those who were deceived."

GERMAN TECHNOLOGY TODAY

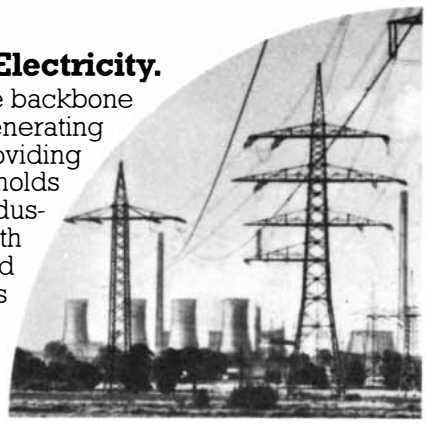


The economic success of the Federal Republic of Germany has captured the imagination and the respect of the industrialized world. Technology in the form of innovation, the improvement and upgrading of products has played a major role in this success: it is the subject of this Special Report by economic and financial writer Andrew Hargrave.

VEBA

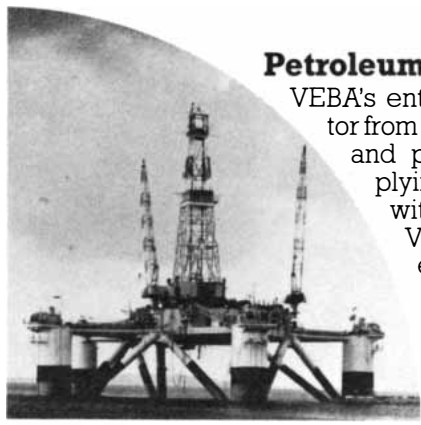
Our Strength - a broad Base

The VEBA Group is one of Germany's major energy producers.
It is at the same time a broadly based industrial and service organization with
total sales of more than DM 30 billion, 80,000 employees
and nearly 1 million shareholders.



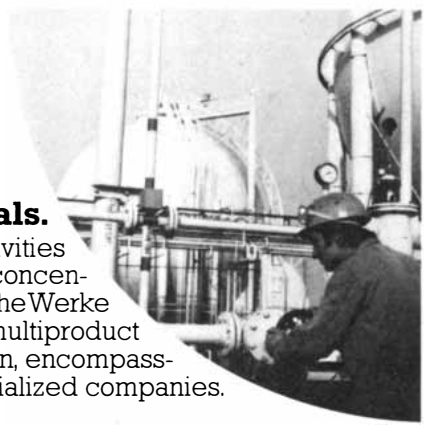
Electricity.

VEBA's corporate backbone is electricity generating and supply, providing millions of households and numerous industrial users with power generated from various fuels such as coal, oil, natural gas, and nuclear energy.



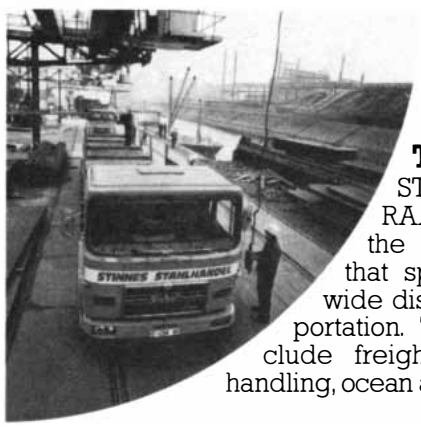
Petroleum.

VEBA's entire petroleum sector from the search for crude and processing to supplying of end-users is within the scope of VEBA OEL. Global exploration is the function of DEMINEX. VEBA's refineries process 16–17 million tons of crude annually. Distribution is with RAAB KARCHER and ARAL.



Chemicals.

The chemical activities of the Group are concentrated in Chemische Werke Hüls (hüls), a multiproduct chemical concern, encompassing several specialized companies.



Trading and Transportation.

STINNES and also RAAB KARCHER are the Group companies that specialize in worldwide distribution and transportation. Their activities include freight forwarding and handling, ocean and inland shipping.

VEBA

Energy is our Business

For further information and your copy
of the Annual Report please get in touch with VEBA AG,
P.O. Box 300306, D-4000 Düsseldorf 30

Foreword by the Federal Minister for Research and Technology

High performance in the fields of research and development coupled with an effective policy on research and technology are vital for a highly developed industrial country such as the Federal Republic of Germany. There can be no resting on the laurels of a great scientific past: consequently all involved in science, in the economy, in government must redouble their efforts to achieve the necessary innovations. Those in authority have recognized the urgency of this task: the level of achievement is reflected in this Special Report.

Well over half the research and development activities in industry in the Federal Republic (63 per cent last year) were carried out by business enterprises which also financed them to a considerable – 49 per cent – extent. Altogether research and development expenditure last year amounted to 2.3 per cent of the Gross National Product and to 3.2 per cent of the combined budgets of the federal and state governments and municipalities.

The objectives of research and technology correspond closely to the overall aims of the federal Government and may contribute to

- extending and deepening the level of understanding for science;
- maintaining and developing the performance and competitive ability of our economy;
- protecting resources and what is generally known as the "good life"; and
- improving the living and working conditions of the people, recognizing the consequences and inter-action of technological developments and justifying and discussing decisions in their use, especially in weighing up the opportunities and risks inherent in those developments.

Official promotion of research may influence the content of research, which for sound reasons is organized along pluralistic lines, only to a limited degree. Research and development in industry and its commercial exploitation is primarily the responsibility of business enterprises. In meeting public needs, on the other hand, co-operation with various institutions is essential. In basic research the choice of projects and working methods is the sole responsibility of scientists.

Last year the State – the federal Government in the main – spent DM 4.2* billion on the promotion of research and development in industry. This corresponds to 23 per cent of industry's total expenditure on research and development. Moreover, tax concessions to promote research and development, amounting to DM 260 million in 1977, have now been extended to small and medium-sized firms.

Direct promotion of research and development occurs mainly where the aim is to eliminate universally known economic and social bottlenecks, although the execution of individual projects is generally left to business. Promotion by the State itself follows only when private enterprise is unable to carry out the necessary research and development in the given time or fails to carry it out altogether.

This applies especially to activities carrying high scientific-technical or business risks; projects which require large investment and whose development is over a long period or where demand is as yet insufficient.

The choice, concept and execution of such projects, sponsored usually within priority programs by the federal Government, require the creative co-operation, expertise, market knowledge and experience of business corporations, unions and trade associations as much as the knowhow of scientists.

In many ways, research and development in the Federal Republic of Germany and, in particular, the research policy of the Ministry of Research and Technology, operates in an international environment. This applies, above all, to the European Community; but co-operation also exists within OECD and UNESCO as well as with organizations covering such varied fields as energy and nuclear research, space research, astronomy, molecular biology. Finally, there is co-operation at two-nation or multinational level in single projects or part programs, including fast reactors, uranium enrichment, reactor safety, medicine, satellites, rockets and space laboratories. The growing significance of co-operation with Third World countries is underlined by several recent agreements in which the Ministry of Research and Technology has been involved.

This brief introduction to "German Technology Today" indicates the range of State participation in research and technology and its contribution to the solution of urgent problems which transcend national boundaries. *Volker Hauff*

* *Approximately 1.9. Deutsche Mark = one US dollar, or roughly two to one.*

Technology: an unceasing quest

Technological excellence, as Federal Minister Hauff points out in his Foreword, is a far from theoretical issue in the Federal Republic of Germany. A nation which is, to a very great extent, dependent on imported raw materials, including energy, to ensure continued prosperity and rising living standards, is therefore faced with a dual task. It has to utilize to the utmost its own slender resources and, at the same time, search for the most appropriate and economic solutions to convert the imported raw materials into high-grade products, a substantial proportion of which would then have to be exported to pay for imports.

Technology, defined here as the application of research to industry and commerce carried out in a suitable human environment, plays a major role in both these tasks. Sometimes, as for instance in the two spectacular upsurges in energy prices during the past six years, it has had to tackle unforeseen and unpredictable twists in world economic patterns beyond the control of not only Germany but most other advanced industrial nations.

Germany, along with Japan, a remarkable example of post-war "economic miracle", has managed to absorb the shocks of energy price rises following the 1973 Middle East War and the 1978 Iran crisis in a significantly short time. It has maintained a healthy balance of payments surplus throughout, partly by playing a major role in the process of re-

cycling oil revenues by selling goods and knowhow to the oil producers. (This has, incidentally, become a source of headache to many German companies involved in trade with and investment in Iran).

It has been said that the second energy jolt came just at the right time to concentrate German minds on one of the biggest – if not the biggest – political/economic/social problems it faces over the next few years. Failure to face the issue of nuclear power squarely and in good time has been all but admitted by Federal Minister Hauff in a magazine interview: "In the debate over nuclear power, political institutions, parties, parliamentary groups have reacted only when the waves of protest were already rolling."

Nuclear power is a significant factor, directly or indirectly, in all three major areas of controversy which, in his view, are troubling the nation today. He defines them as the relationship between technology and environment, the danger to employment through technological change (with microprocessing as the focal point) and risks to safety arising from technological progress.

The nuclear debate is not, of course, confined to the Federal Republic. That is why Herr Rudolf von Bennigsen-Foerder, chief executive of the major energy, chemicals and trading group VEBA, is enthusiastic about the result of the Swiss referendum in February approving the building of nuclear power stations. But that was before the Three Mile Island accident in Pennsylvania which did not go unnoticed by the anti-nuclear lobby in Germany. All the country's existing or planned reactors are of the pressurized water type, similar to Three Mile Island's.

In Germany the debate is further complicated by the often conflicting desires to maintain and, if possible, improve the high living standards and, at the same time, adhere strictly to democratic procedures in allowing nuclear development to go ahead. But while the debate continues, the authorities as well as industry are trying to make sure that alternative sources of all kinds, particularly coal – Germany's only major indigenous fuel – are used to their best advantage.

For instance, the German chemical industry which includes three of the world's top five concerns, is already planning to make a bigger use of coal as an alternative raw material to oil. Professor Dr. Herbert Gruenewald, chief executive of Bayer, foresees "an intensification of the trend in favor of coal-based chemistry" in about ten years' time, increased utilization of oil shale and sands in the decade to follow and "large-scale commercialization of the manufacture of chemicals from biomasses" in the next century. These views are echoed by Professor Dr. Matthias Seefelder, chief executive of BASF, another of the chemicals trio.

In the short term, however, better energy saving and fuel utilization, including oil, are the major preoccupation for German industry, from chemicals to steelmaking and vehicle manufacture. So

TABLE I **Regional Structure of German Foreign Investment, 1960–1978** (In DM billions)

Countries	1960	1970	1975	Mid-1978
Industrial Countries	—	14.9	29.7	38.5
of which				
EC TOTAL	0.42	7.3	14.8	18.6
Belgium/Luxembourg	0.09	2.4	4.2	5.4
France	0.14	2.1	4.3	5.4
The Netherlands	0.08	1.1	2.8	3.3
Great Britain/Ireland	0.04	0.6	1.7	2.2
Italy	0.06	0.8	1.6	1.8
Switzerland	0.36	2.2	4.3	5.0
Spain	—	0.9	2.9	3.0
Austria	0.08	1.0	1.5	1.8
US	0.28	1.8	4.2	7.4
Japan	—	0.1	0.3	0.3
Developing Countries	—	6.2	12.3	16.3
of which				
Brazil	0.53	1.5	2.9	4.2
Oil producers	—	0.1	0.9	1.4
Eastern Bloc	—	—	0.06	0.3
TOTAL	2.93	21.1	42.0	54.8

is the onset of microelectronics, both to be discussed in some detail later in this report.

There are, of course, problems other than energy. Dr. Wilfried Guth, joint spokesman for Deutsche Bank, Germany's largest commercial bank, and the man who was, at the time of writing, tipped as the next president of the Deutsche Bundesbank, Germany's central bank, sees a series of potential threats to German prosperity. These include a slowing down of economic growth, the share of wages and salaries within the nation's total income rising at the expense of income from enterprise and capital accumulation, increasing State expenditure (a phenomenon common to all advanced industrial countries), a "rougher social climate", the shift of center of economic growth to OPEC and more developed Third World countries, particularly those of Southeast Asia and Latin America, the uncertainty over currencies (including the continued upward trend of the Deutschemark), sharpening international competition (between a quarter and a third of Germany's industrial output goes to exports) and a trend towards protectionism in world trade.

Dr. Guth, in referring to the US as a "sleeping giant", believes it is only a question of time before the US makes full use of its new opportunities in the world markets, "something we should welcome in view of its balance of payments position and the state of the dollar."

One significant effect of the rising value of the Deutschemark and of domestic pay levels has been the steep increase in investment by German companies outside Germany. This rose – as Table I indicates – from an admittedly slow post-war start seven-fold between 1960 and 1970 and a further two-and-a-half times beween 1970 and mid-1978:

A further analysis of foreign investments in the same report provides a breakdown of industries and services. It comes as no surprise that industries with the highest per capita as well as overall investment in research and development and deep involvement in technology transfer are prominent as foreign investors as well. For these are the very industries which stand to lose most heavily if they are priced out of world markets, with a consequent threat to profit and employment.

Several leading German companies have made news in recent years with sizeable foreign investments. Some of these will be examined in detail elsewhere in this report but they include Hoechst, the third of the big chemicals trio, which now invests more abroad than in the home country. Siemens, the top German electrical and electronics group, invests more than a third of its world total abroad, with the US as the major recipient.

There are, of course, considerable risks inherent in foreign investment offsetting, to some extent, the obvious advantages. Apart from such unforeseen risks as the revolution in Iran, technology transfer often produces fresh competition, sometimes in the investor's own market. But as Dr. Hans Friderichs, chairman of the managing board of Dresdner Bank and former federal Economics Minister points out, "We cannot very well build factories with our superior plant technology knowhow and

then tell the countries in which they are built that their products may not be sold in the German market."

All the same, Dr. Friderichs sees another side to the coin. New opportunities arise in those countries precisely because products incorporating high technology create a new demand which may in turn be exploited by German exporters as an additional means of protecting markets and, incidentally, Germany's balance of payments surplus as well. (Last year, in spite of the usual deficit on the services account, the surplus amounted to a healthy DM 16.2 billion).

Nor can one ignore the millions of jobs which depend on the continuing success of Germany's exports.

The growing tendency of German companies to invest abroad lends additional urgency to the urge to upgrade manufactures through increasing input of research and development. The expanding network of foreign manufacturing associates, subsidiaries and branches not only produce export-displacing or replacing goods but often surface – Dr. Friderichs' point – as direct competitors to the parent companies, even in the home market. It is an experience common to all western nations, but for one depending so much on exports for economic growth such as Germany, it constitutes an additional threat.

One small example: Volkswagenwerk imports about 100 "Beetles" a week from its Brazilian plant. The once popular model is no longer manufactured in Germany and has been replaced by the more sophisticated and up-to-date Golf, the Rabbit in the US. The imported "Beetles" for which there is still a minor demand, thus appears as a competitor in the home market.

The role of technology in upgrading the German product is the main theme of this Special Report. It will analyze industry's overall research and development (R & D) input and the Government's contribution. The efforts of private enterprise will be examined through the experience of some 30 companies, mainly though not exclusively large and powerful. This analysis will be supplemented by the contributions of other organizations such as banks (particularly their role in providing risk capital), exhibitions and research bodies.

Throughout the report close attention will be paid to the interlocking nature of technologies spreading right across the sector boundaries of industry. Thus data processing has intruded into virtually every corner of the engineering, steel, electrical, automotive and aerospace industries (among others) and the energy issue affects not only power generation but, as already noted, chemicals, steelmaking, vehicle manufacture and many others. Nor can the already mentioned environmental problems and their sometimes inhibiting, sometimes beneficial effects on technological development be ignored: and as Germany's experience in all these fields is shared to varying degrees by other leading industrial nations, some international comparisons will also be included.

TABLE II **Sector Structure of Foreign Investment, 1971–1978**
(In DM billions)

	1971	1973	1975	1977	Mid-1978
Industry	19.9	25.0	32.2	39.9	41.5
of which					
Chemicals	5.5	6.2	8.0	9.4	9.7
Electricals/electronics	2.8	3.5	4.4	5.6	5.6
Iron and Steel	1.4	2.8	3.3	4.1	4.2
Engineering	1.7	2.3	3.1	4.0	4.2
Oil exploration/processing	1.0	1.4	2.5	3.4	3.6
Automotive	2.2	2.3	2.8	3.4	3.5
Services					
of which					
Banks and Insurance	1.6	2.7	3.9	5.4	5.9
Trade	0.3	0.5	0.5	0.7	0.8
Other sectors	0.7	0.7	1.0	1.2	1.4
TOTAL	23.8	32.2	42.0	52.1	54.8

R & D: the overall effort

Germany's total scientific expenditure came to about DM 40 billion last year, of which the Federal and State governments funded 63 per cent and business the rest. It amounted to 3.1 per cent of the Gross National Product compared with 2.6 per cent in 1969.

Research and development accounted for nearly three-quarters of the expenditure, DM 29.2 billion, shared in almost equal proportion by business and the State: only three per cent of it was funded by institutions. Of the total DM 18.4 billion was spent on R & D in industry, 77 per cent of it funded by business companies and their associations. In the case of Siemens, Germany's largest single spender on R & D (DM 2.3 billion in 1977/78), almost 90 per cent of it was self-financed.

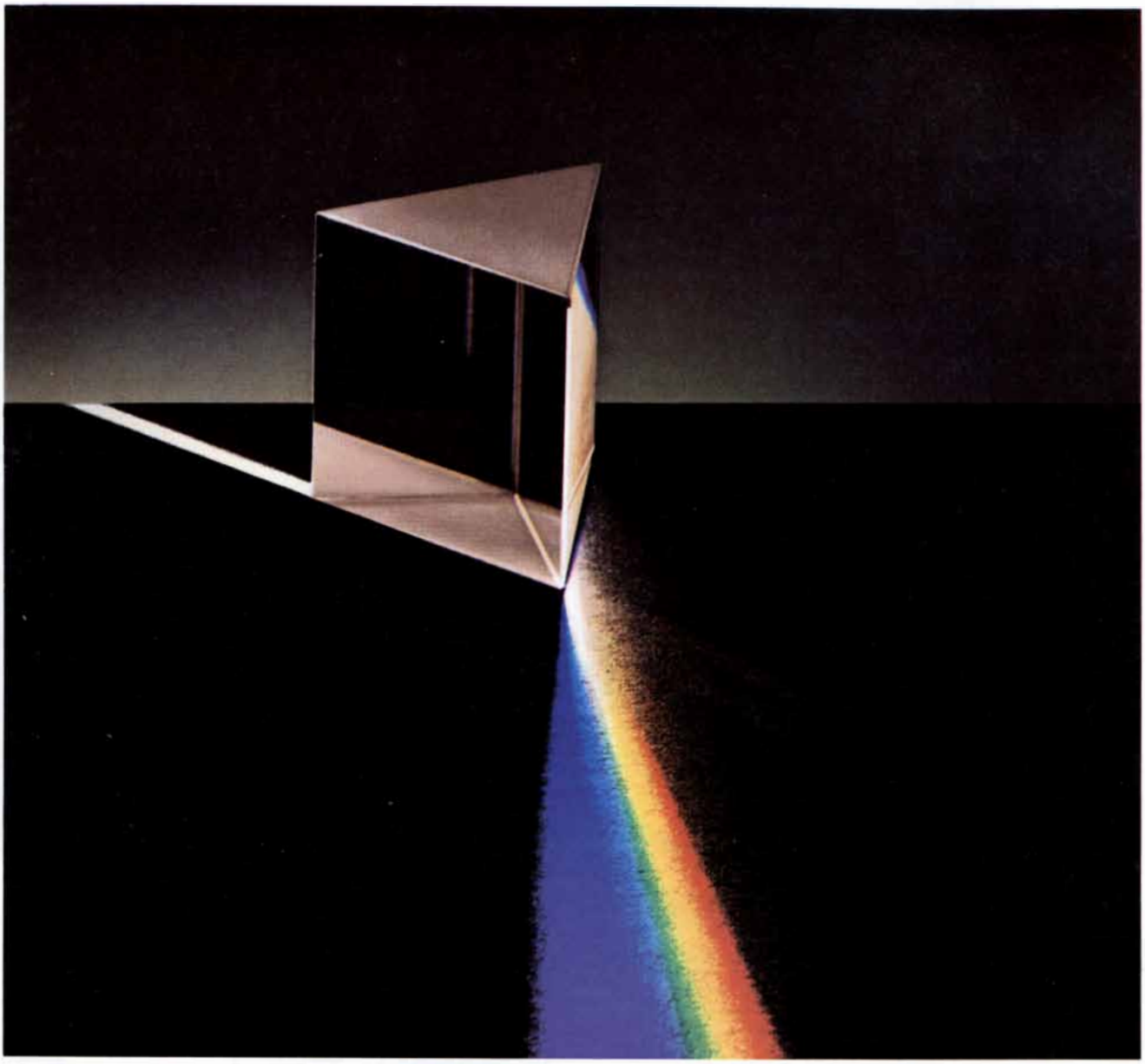
Business's expenditure on industrial R & D, DM 14.4 billion in 1978, compares with its expenditure of DM 6.4 billion in 1969 when total spending on industrial R & D, including State contribution, was only DM 7.5 billion.

The total scientific, technical and other personnel employed in R & D in industry peaked at 199,000 in 1971, declined to 186,000 by 1975 and has remained around that level since. However, the Ministry for Research and Technology claims in its latest annual report that "industry in the Federal Republic still has a great R & D potential in terms of personnel and funds."

R & D expenditure within Germany's Gross National Product rose from 2 per cent in 1969 to 2.4 per cent in 1975, slowing down to 2.3 per cent in 1977. However, the latest OECD figures show that the US spent the same proportion of its output on R & D in 1977 which compares badly with the 2.8 per cent invested in 1969, year of the first moon landing.

The same OECD figures quote the latest available R & D percentages in other advanced western nations: 2.1 per cent by Britain (1975), 2 per cent by Japan (1974), 1.9 per cent by France (1975).

The figures appear to bear out the contention that the West, including the US and Germany, are indulging in a "technological breather." This theory is challenged by at least some top German industrialists.



To find the spectrum, you need the prism.

A universal bank, such as Deutsche Bank, acts as a prism. It takes in the widely varied problems of a multitude of clients (individuals and companies; governments and institutions) and carefully breaks each one into its various components in order to find the simplest, quickest and most efficient solution.

Come to Deutsche Bank, when you want the full spectrum of banking services focused on your problem.

Deutsche Bank
A century of universal banking

Deutsche Bank AG New York Branch, 9 West 57th Street, P. O. Box 890, Radio City Station, New York, N.Y. 10019 U. S. A., Tel.: (212) 940-8000, Telex: TRT 177747.

Central Office: Frankfurt (Main) / Düsseldorf; Branches: Antwerp, Asunción, Brussels, Buenos Aires, London, New York, Paris, São Paulo, Tokyo; Subsidiaries: Hong Kong, Luxembourg, Singapore; Representative Offices: Beirut, Bogotá, Caracas, Hong Kong, Istanbul, Cairo, Lagos, Madrid, Milan, Mexico, Moscow, Osaka, Rio de Janeiro, Santiago, San José, Sydney, Tehran, Tokyo, Toronto.

Diesel-powered cars seem like a startling new idea to just about everybody these days–except Mercedes-Benz.

Mercedes-Benz has built over two million Diesel cars since introducing the world's first production Diesel 43 years ago. Priceless expertise, epitomized by the 5 Diesel thoroughbreds shown here. Read all about them, and about the crucial differences between a Diesel – and a Diesel from Mercedes-Benz.

Cynics were still snickering that "the Diesel engine can never work in an automobile" when Mercedes-Benz unveiled the world's first production Diesel passenger car.

That was in 1936. By 1958, Mercedes-Benz had introduced its Diesel cars to America. And while skeptics warned that "Americans will never buy Diesels," Mercedes-Benz proceeded to sell more than 155,000 of them over the following 21 years.

8 generations of Diesel cars

Today, Mercedes-Benz has progressed to an *eighth* generation of Diesel-powered cars, in a range of five models.

Five singular cars, each a unique rendering of the Diesel idea. Each backed by those 43 years of Mercedes-Benz Diesel development and the experience gained from building more than two million Diesel cars. Their engineering proven in labs and on test tracks, of course – but also by millions of miles of everyday use in 142 countries around the world.

Same principle, different Diesels

All Diesel-powered automobiles share the same technical principles – those laid down by inventor Rudolf Diesel almost a century ago.

But if technical principles were all that counted, all Diesel cars would be alike. It also counts how a Diesel car is *designed* and how it is *built*.

Diesels are no sudden enthusiasm, no production sideline at Mercedes-Benz. Built in the thousands year after year over the decades, Mercedes-Benz Diesel automobiles are almost the company's lifeblood. They reflect the efforts of engineers who have made the Diesel-powered automobile a career.

You can't argue with results

This commitment and that 43-year fund of Mercedes-Benz Diesel expertise combine to pay welcome dividends.

Diesel engineering breakthroughs seem almost a Mercedes-Benz habit. For example: In 1975, the world's first five-cylinder Diesel passenger car; in 1978, the world's first turbocharged Diesel passenger car – its engine, cousin to that used in a 200-mph Mercedes-Benz Diesel coupe that shattered 9 absolute world speed and distance records.

Other dividends take less spectacular but still impressive forms. The special precombustion chamber provided for each cylinder of a Mercedes-Benz engine, for instance: Air and fuel are mixed and burned in *two* stages, for running smoothness and more complete combustion.

Five different Diesels

The Mercedes-Benz Diesel range consists of more than varied body styles. Here are profiles of five different kinds of Diesel cars:

The 300 SD Turbodiesel Sedan is a *performance* Diesel – the first production Diesel car to employ turbocharging. It is also the roomiest Diesel sedan Mercedes-Benz has ever built, and the most sumptuous.

The total effect is breathtaking. You are literally turbocharged away from stoplights, up long grades, through passing maneuvers, with a feeling of power to spare. "Diesel lag" is gone.

The new 300 TD station wagon, just announced, had to behave like a Mercedes-Benz first and foremost: it is meant to be not just another station wagon, but a brisk-handling machine that is rewarding to drive – even under heavy load. The 300 TD is as *solidly built* as a Mercedes-Benz car, and interior space has been so ingeniously used that all of the passenger seats can fold away to create the maximum possible cargo space.

The 300 CD Coupe introduces Diesel efficiency into the elegant realm of the two-door, limited-production coupe.

The extraordinary result: a rival to some of the world's most exclusive two-door cars, rich with creature comforts and ideal for extended highway cruising – yet powered by a five-cylinder Diesel engine that purrs along on the cheapest automotive fuel you can buy.

The 300 D Sedan is that rarity of rarities, a four-door automobile that is also a *driver's* car.

One spirited run down a challenging mountain road in the 300 D should forever quash the old prejudice that a sedan, or a Diesel, has to be dull.

The 240 D Sedan offers classic, no-nonsense Diesel practicality in its most refined form to date.

With its modest 2.4-liter, four-cylinder engine size, the 240 D's fuel appetite is meant to be lean indeed. The EPA estimates 30 mpg for a 240 D equipped with a manual 4-speed transmission. *Remember:* compare this estimate to the 'estimated mpg' of other cars. You may get different mileage, depending on how fast you drive, weather conditions, and trip length.

Economy is furthered by a manual 4-speed transmission as standard equipment. For those bent on peak efficiency, air-conditioning and certain accessories are omitted from the standard equipment list. They can be ordered as extra-cost options.

Engineered like no other car in the world

The Mercedes-Benz aim is doggedly single-minded. It is to build safe, comfortable, practical cars with as few imperfections as possible.

This philosophy puts engineering ahead of petty economies and precludes the mass production of inexpensive cars.

A Mercedes-Benz is engineered like no other car in the world.

Light alloy wheels on 300 CD and 300 TD – and sunroof shown – are at added cost

Professor Dr. Joachim Zahn, retiring chief executive of automotive manufacturer Daimler-Benz, sees no reason to be pessimistic: "It is a fundamental characteristic of man to strive constantly to improve his environment and thus become a permanent driving force for growth."

Dr. Bernhard Plettner, chief executive of the electrical giant Siemens, puts it differently: "The success of the German economy is due to the very broad range of products incorporating science and technology", he says. "Companies operate in an international environment: despite the high-value Deutschemark and high wages, Siemens exported about one-third of its output last year and made a profit."

A dissenting voice is that of Dr. Guenther Sassmannshausen, chief executive of Preussag, a group active in as diverse fields as energy, metals, construction and transport. He criticizes the federal Government for what he terms as a lack of raw materials policy, of co-ordination and the loss of drive. "We may even have misread our much-praised economic miracle," he says, adding that only by technological progress – which competitors in the Third World cannot imitate – can the West stay ahead.

There are a number of criteria which link technology to economic growth. An analysis by the German Science Foundation (Stiftverband der Deutschen Wissenschaft) for instance shows that the high spenders on research and development are equally successful in increasing their sales and exports.

Thus between 1970–76, the seven above-average growth industries included five of the biggest spenders on R & D. They were, in order of increase in growth rates, chemicals (4.7 per cent of sales spent on R & D); the automotive industry (5.7 per cent); instrument manufacture and optics (4.9 per cent); engineering (3 per cent); and electrical/electronics (6.4 per cent). The exceptions were crude oil processing which topped the list, an industry dominated by multinationals whose research and development is located largely outside Germany; and timber processing, a relatively small industry.

Of manufacturing industry's total expenditure on R & D during that period, three – chemicals, electrical/electronics and automotive – were between them responsible for over three-quarters and engineering for a further one-eighth of the total. In other words, these four industries account for about 90 per cent of all industrial research and development in Germany.

One should not, even in the German context, ignore the relatively small but technologically advanced aerospace industry which spends on R & D an amount yearly which corresponds to over one-third of its total sales. As in other countries, a very high proportion of this is funded by the Government.

The four high R & D spenders are also Germany's largest export earners, accounting for about two-thirds of the country's exported manufactures. This seems to confirm Dr. Plettner's assertion linking high-grade technologically-based products with success in the world market. According to Dr. Manfred Lennings, chief executive of the substantial GHH engineering group, such exports are essential to maintain production capacities, even at the risk of technology transfer. For even if this enables the overseas customer to manufacture his own plant and machinery – and in this he agrees with Dr. Friderichs of Dresdner Bank – it acts as a spur to the German manufacturer to upgrade his own products and offer even more sophisticated ones in the customer's home market.

A further criterion of technological progress is productivity. According to the Federal Economics Ministry, new technology is responsible for about three-quarters of the increase in productivity: the fact that over a five-year period in the 1970's productivity had increased at an annual rate of only 3.8 per cent compared with 5 per cent in the 1960's may be a further sign of a slow-down in technological advance.

Yet an analysis of 3,400 German companies of all sizes carried out by the IFO economic research institute shows that technological development rates fairly high in the forward planning of industry. The analysis, as Table III shows, also reflects the significance of size in the companies' investment intentions.

The table points to some interesting facts, including the not unsurprising one that the larger the company, the more it invests in technology – technical development, quality improvement, changes in methods and patterns of production, improvement in working and environmental conditions, energy saving and new R & D plant. More remarkable is the relatively low percentage of even the largest companies intending to invest in energy and other raw material savings and in new R & D plant.

Perhaps the shock of the energy crisis of 1973/74 may have worn off by the time the survey was taken, in 1976; resistance to nuclear expansion which has virtually paralyzed the building of new nuclear power stations in the last four years was less evident; and the revolution in Iran had not yet taken place. It is therefore possible that a more up-to-date survey would show significant changes.

TABLE III **Investment Planning According to Size of Companies, 1976–80**

Reasons for investment	Companies under 50 Employees %	Companies 50–200 Employees %	Companies 200–1,000 Employees %	Companies 1,000 and upwards %
Rationalization	74.1	84.6	91.3	96.0
Technical development	53.4	63.0	69.5	76.3
Improvement of quality	33.7	51.7	57.0	63.4
Changing structure of demand	29.2	32.3	38.5	43.6
Change to new methods of production	22.5	28.4	37.0	50.3
Conversion/Extension of production pattern	23.6	30.3	36.0	41.0
Improvement of conditions of work	22.0	25.3	31.9	47.6
Environmental protection	14.3	23.2	33.7	55.4
Energy and raw material cost increases	19.7	21.0	26.3	34.5
Too small capacity	9.4	12.1	11.9	13.3
Scarcity of labor	11.3	10.3	11.6	9.3
R & D plant	2.3	4.9	11.4	28.1

The role of the State in R & D

As Minister Hauff says in his Foreword, Government promotion of research and development in industry – DM 4.2 billion last year, or 23 per cent of total

Advanced Technology. We've made it our business.

Today, as throughout its 140 year history, Krauss-Maffei owes much of its success to the consistent application and continuous development of the latest technologies. That's why each one of this Munich operation's five divisions is a worldwide leader in its field.

Krauss-Maffei builds machinery and vehicles for the most demanding applications. But sophisticated engineering and dependable quality are only part of the story. Krauss-Maffei supplies not only individual machines and vehicles, but complex systems as well. And the firm's technological know-how is extensive enough to provide solutions for a myriad of engineering problems.

Take the Porsche 928, for instance. Its rear section and taillights are manufactured on Krauss-Maffei **plastic molding machines.** An innovator in this field for more than fifty years, Krauss-Maffei today offers a complete line of products ranging from injection molding machinery, extruders and polyurethane foam equipment to complete manufacturing installations.

In the field of **transportation,** Krauss-Maffei is actively involved in shaping the future of railroad technology. Not only by developing and manufacturing locomotives of every size and description and producing electronic feedback control systems for safe and economical railroad operation. But also through intensive research in such areas as determining the practical limits of the wheel-and-rail system and developing new transportation systems operating on the non-friction principle of magnetic levitation.

Casting and forging have long been Krauss-Maffei specialties. Particularly the production of complex, hard-to-cast parts. Or the manufacture of ready-to-install rollers possessing the high degree of surface precision required by the plastic, rubber, paper and steel industries.

Over 5,000 intermediate and finished products are processed with the help of Krauss-Maffei filters, dryers and centrifuges. The **process engineering** division supplies equipment for the mechanical and thermal separation of liquids and solids for the chemical, pharmaceutical and mining industries as well as for fertilizers, plastics, foodstuffs and reclaimable waste products.

Another complex field of concentration is **military technology.** A weapon system as sophisticated as the Leopard 2 tank is the result of a concerted analytical approach taking every technologically practicable option into consideration. But there's more. Training, testing and maintenance equipment, for instance. Modification . . . modernization . . . logistical support and – for Krauss-Maffei as Chief Contractor – the mammouth task of coordinating the shipments of more than 1,500 subcontractors for series production.

Krauss-Maffei
Aktiengesellschaft
Krauss-Maffei-Straße 2
8000 München 50
West-Germany

expenditure – is largely indirect, carried out in sponsored projects by private enterprise.

Direct promotion is limited to fields, such as aerospace or energy, where large investment over long periods is required, risks are high and demand needs to be built up.

In view of resistance to nuclear power by sections of the public as well as certain local authorities, the Government has considerably stepped up its support for non-nuclear sources of energy such as district heating, conservation measures, the use of waste heat and projects in search of renewable sources such as solar and wind energy. While in 1972 support for nuclear research, DM 789 million, compared with a mere DM 10 million for non-nuclear energy projects – a ratio of 79 to one – by last year support for non-nuclear research rose to DM 406 million as against DM 1.1 billion for nuclear projects, a ratio of rather less than three-to-one. By 1982, the margin should narrow further, with the rate of increase for non-nuclear research more than four times that of the nuclear one to reach a ratio of two-to-one.

Under the medium-term plan for "modernizing industry", the Government is providing support to the tune of around DM 9.8 billion over the three-year period 1978 to 1980 under the headings shown in Table IV.

Other major areas of Government support include safety and environmental research, the "humanization" of work, transport and – a substantial item – the raising of scientific performance overall.

Looking further ahead in its latest report, the Ministry for Research and Technology wants to see the German electronics industry among the world leaders by 1982. (In 1977 around 90 per cent of all microprocessors and semiconductor storage were sold in Germany by foreign companies). Hence, the Government is stepping up aid to this vital sector, particularly to small and medium-sized companies, through sponsored research, a new technological advisory center as well as tax allowances.

Total expenditure on R & D by the Ministry over the period 1979–1982 will exceed DM 25 billion in order, the report indicates, "to give new impetus to the further modernization of our economy," including innovation "on a broad front" leading to new production processes.

However, the report adds, the Minister for Research "may only provide the initial thrust – help to self-help. Business must take responsibility for further development."

At the same time, as Minister Hauff has emphasized, it was the Government's duty to safeguard employment, especially in view of the additional one million young people leaving schools and universities by 1985. There must therefore be a limit to the transfer of production abroad, opportunities left for technologically less developed products and time provided for the necessary adjustments in such industries. For these reasons alone, he says, "the modernization of the economy, including industry, is a top priority for the Government."

German technology: some comparisons

The rising proportion of the German GNP being invested in research and development by Government and industry over the past decade has been noted above. Although the impetus has slowed down somewhat after 1975, evidence points to German R & D input as a share of GNP reaching par with that of the US which had a long lead as recently as the early 1970's. This should not, however, disguise the fact that the absolute R & D expenditure in the US is still considerably higher.

R & D expenditure by itself does not of course guarantee technological progress, let alone company profitability. Nevertheless, the link between R & D expenditure on the one hand and total sales, exports and productivity on the other, has been clearly established: and so has the high rating given to technology-based investment in the forward planning of German companies, particularly the larger ones. Significant, too, is the speed with which the fruits of research, invention as well as innovation, are developed and translated into finished products and, eventually, into worldwide sales.

Such comparisons are notoriously difficult, particularly as invention and innovation – especially the latter – seldom occur in a single country, let alone a single company. Moreover, the timescale and diffusion of products from invention to worldwide sales are difficult to measure objectively, even if the starting points, processes and products are clearly defined.

It has been estimated that all but 5 per cent of the people employed in German industry's R & D are engaged in applied research which is a higher proportion than in most other industrial countries, including the US. This does not mean that Germany lags behind in basic research as a great deal of such research is carried out in universities, technical colleges, scientific and research institutes. It does signify though that R & D in German industry has a highly practical bent which would seem to suggest that while German companies may not always be first in the field with a product or process, once these surface in the market and achieve some success, the Germans are quick off the mark in exploiting them.

In support of this theory, Minister Hauff has expressed the hope that the German computer industry will be "internationally competitive" in the early 1980's. At present Germany imports more than 90 per cent of its microprocessors, mainly from the US – and this in spite of the Government contribution of DM 3.7 billion to investment in data processing so far. But, says Minister Hauff,

TABLE IV **R & D Support by Government for Modernization of Industry** (In DM millions)

Item	1978	1979	1980
Energy and Raw Materials	1,718.2	1,868.5	1,921.8
Data processing	310.0	314.0	320.0
Telecommunications and Electronics	149.6	159.2	160.2
Key Innovative Technologies	367.4	387.9	403.1
Space Research and Technology	554.0	571.0	585.0
TOTAL	3,099.2	3,300.6	3,390.1
Total Government expenditure on R & D for Industry	7,748.5	8,021.3	8,183.7
	40.0%	41.4%	41.4%

Siemens has increased its share in the German data processing market from 6 per cent to 20 per cent in the last few years and now has a market share of 7 per cent in Europe.

As for profitability, in the last annual report of Siemens, Dr. Plettner proudly points to the fact that the data processing side of the business had made a profit for the first time ever.

All the same there is no denying the fact that IBM is still ahead in the German computer market, let alone the European or world markets. Some German industrialists are sceptical whether this lead could be seriously cut down or eliminated altogether in the short term, especially as Japan has emerged as a formidable new competitor (particularly in the application of micro-electronics) in a broad range of industries.

Dr. Egon Overbeck, chief executive of the Mannesmann steel and engineering combine, recently told businessmen in New York that while the Europeans widely adapted and even improved on technologies originating in the US, he doubted whether they could shake US domination in certain fields, such as reactor techniques, aircraft construction, communication and semi-conductor technologies.

In the motor vehicle field, Professor Dr. Zahn of Daimler-Benz sees big US investment as the main threat. He predicts "a new generation of automobiles from General Motors and Ford, backed by an investment of $50 billion, to hit the world markets in 1980/81.

But while accepting a US lead in certain disciplines, Dr. Overbeck has also pointed to other sectors of industry, including Mannesmann's own business, where Germany may claim a technological edge. Among these, in his view, are engineering, steelmaking and the upgrading and conversion of coal, including gasification and liquification.

Dr. Overbeck's views are certainly based on long, broad experience and success. There have also been a number of studies attempting to put such international comparisons on a more scientific basis.

The two cited in this report, one a joint survey of some 20 technologies conducted by the German center of the Battelle Institute and the IFO research institute of Munich,* the other a seven-na-

* *Development trends in production and manufacturing processes in the Federal Republic of Germany: Opportunities for a State research and technology policy, Volume III, December 1975.*

tion survey of certain key industries (with IFO participation)** have come up with a number of significant findings, in some cases confirming, in others casting doubt on existing notions of comparability. They are slightly out of date, having been published late in 1975 and in 1974 respectively: the pace of industrial change has accelerated considerably since the war. (For instance, 45 per cent of Siemens' entire output in 1977/78 consisted of products less than five years old.) Nevertheless the surveys still retain some relevance to the main theme of this report.

The first study makes a special point of the technological linkages that exist between industries, products and applications. For instance, computerized numerical controls of plant, machinery and tools (which are a subject of the second study as well) are being used in industries as diverse as engineering, foundries, textiles and rubber. Measuring devices and processes, now increasingly incorporating computers and microprocessors, spread across virtually the whole field of manufacturing. High-precision inspection systems overspill from aerospace into the automotive industry. Modular control systems are shown as being used by the manufacturers in their own production as a sales point or model to their customers.

The second study's method of comparison is not so much the discovery of a product or manufacturing process but the time-scale which separates innovation and prototype from large-scale production and marketing. In control systems for machines, it finds that although the US as well as the United Kingdom, Austria and Sweden had a head start in its first use, by 1968 Germany had drawn at least level, except in electronics and the aircraft industry. In both, as Dr. Overbeck has asserted, the US enjoys technological as well as volume superiority. According to the first study wider application and use is one of the factors in the US and also Japan still having an edge in the spread of electronic systems, controls and processes.

Dr. Overbeck has pinpointed continuous casting in the steel industry as one in which Germany has played a leading role. The seven-nation study confirms that although Germany had made an early but relatively slow start, by 1969 it was using the process to a greater extent than any other steel producer except Sweden but including the US. In oxygen steelmaking, another major post-war innovation, Germany was on a par with the US and Japan in terms of proportion of total tonnage produced by this process, and ahead of the rest.

Another basis for technological comparison is the external balance of fees for licenses and patents. The latest figures published by the Deutsche Bundesbank purport to show that Germany's deficit increased steadily, from DM 409 million in 1969 to DM 1,117 million in 1977. However the bank also indicates that within that figure home-based companies had a sizeable (DM 262 million) surplus while others in which foreign enterprises had a substantial stake incurred a deficit of DM 1,379 million.

As already indicated, practically all the research by international companies dominating the German oil scene is being done at their own home base for which their German associates pay a license fee. The same applies, to a lesser extent, to the international electronics industry in which such companies as IBM, ITT or Philips have a strong German representation.

Indeed, the Bundesbank's international comparisons for 1976 reveal that apart from Great Britain (which had a tiny surplus), the only countries with substantial positive balances in patents, inventions, processes and authors' rights are the US and Switzerland, the latter no doubt as a result of its powerful pharmaceuticals industry.

As much as 83 per cent of the total US surplus of DM 9.8 billion had come from transfer license payments: in other words from foreign associates and subsidiaries to the parent companies.

** *Diffusion of new industrial processes, Cambridge University Press, 1974. German edition, Duncker and Humblot, 1978.*

TABLE V **Germans Among the Top Industrial Companies in the World** *

Rank	Company	Headquarters	Sales 1977 ($ billion)
1	General Motors	Detroit	54.96
2	Exxon	New York	54.13
3	Royal Dutch/Shell	London/The Hague	39.68
20	Siemens	Munich	10.64**
21	Volkswagenwerk	Wolfsburg	10.41
24	Hoechst	Frankfurt	10.04
30	Bayer	Leverkusen	9.22
31	BASF	Ludwigshafen	9.12
34	Daimler-Benz	Stuttgart	8.63
37	Thyssen	Duisburg	8.33

IN ELECTRICAL/ELECTRONICS Siemens is fourth after IBM, General Electric, ITT (all US) and Philips (The Netherlands).

IN THE AUTOMOTIVE INDUSTRY Volkswagenwerk is fourth after General Motors, Ford and Chrysler (all US).*** Daimler-Benz is seventh after the above-mentioned four, Renault (France) and Toyota (Japan).

IN STEEL, Thyssen is third behind US Steel (US) and Nippon Steel (Japan).

* *Fortune, August 14, 1978.*

** *The VEBA group of energy, chemicals, etc. which has usually been at or near the top of German companies in terms of sales, is in the throes of reorganization following a deal with Deutsche BP.*

*** *The list was compiled before the merger of Peugeot-Citroen (35th) with Chrysler Europe.*

Exciting prospects for electronics

"Electronization", a term used in Siemens' latest annual report, is one of the most spectacular, exciting, fastest growing and changing aspects of modern society. The changes are so fast that forecasts and even current estimates vary widely, not least because the definitions and data themselves often vary according to who compiles or forecasts and for what purpose.

What is undeniable is that electronics – computers, microprocessors, automated systems, semiconductors, etc – are applied at an ever-increasing rate in industries and services. Furthermore, all estimates and forecasts agree that the US is maintaining its early lead in the

For a new approach to your banking needs - WestLB.

Cologne Cathedral - from WestLB's selection of wonders of the world

WestLB New York offers a broad spectrum of banking services tailored to the needs of corporate clients seeking a flexible and seasoned approach to financing. Whether the requirement be short-term working capital finance under a line of credit or medium-term credit facilities for capital investment projects, WestLB has the capability to package its financing to suit your requirements.

Credit facilities can be quickly arranged. WestLB can also provide loans in major currencies through its offshore banking branch in the Cayman Islands and installations in all major world financial centers.

WestLB in New York is also an active provider of corporate deposit services including Dollar certificates and Eurocurrency deposits through its Cayman Islands branch. The New York branch also furnishes advice on and execution of corporate foreign exchange needs.

At WestLB you'll meet a dynamic, flexible and experienced team of bankers with a proven international record, who can mobilize the local skills of WestLB New York and the US-$ 40 billion in worldwide resources of the tenth largest bank in the world. For a new approach to your banking needs contact WestLB.

Westdeutsche Landesbank
New York Branch
450 Park Avenue, New York, N.Y. 10022
Telephone: (212) 754-9600
Telex: 420736

A strong force in wholesale banking **WestLB**

Westdeutsche Landesbank

Headquarters P.O. Box 1128, D-4000 Dusseldorf 1, Tel. (211) 8261. Frankfurt Office, Tel. (611) 25791
Branches: London, Tel. 638 6141; New York, Tel. 754-9600; Tokyo, Tel. 216-0581
Subsidiaries: WestLB International S.A. Luxembourg, Tel. 45493; WestLB Asia Limited, Hong Kong, Tel. 5-259206
Representative Offices: Latin-America Office New York, Tel. 754-9620; Rio de Janeiro, Tel. 224 7162; Hong Kong and Southeast Asia, Tel. 5-228211; Tokyo, Tel. 213-1811; Melbourne, Tel. 678191
Participations: Banque Franco-Allemande S.A., Paris, Tel. 3590109; Banco da Bahia Investimentos S.A., Rio de Janeiro, Tel. 253823

S 15

world markets and companies like IBM outsell their non-US competitors even in their own domestic markets.

There is a rush of new companies, again mostly from the US, offering microprocessor applications and computers, thus underlining the US lead in this newest and most innovative sector of electronics.

The question arises: why do European manufacturers not co-operate in countering US domination in this technologically advanced field?

There has indeed been an attempt by Siemens of Germany; Philips of the Netherlands; and CII of France to join in the computer sector of the market and so challenge IBM's supremacy, at least in Europe. The attempt has failed, not least because of the incompatibility of the partners and products and, in the case of CII, intervention by the French Government.

As for microelectronics, Siemens, Germany's leading electrical/electronics concern and the world's fifth largest in terms of sales, recognizes in its latest annual report that Germany still has some leeway to make up.

It estimates that only 13 per cent of electronic components used in Germany are microelectronic, half the proportion of those in the US. "If microelectronics is not applied rapidly and extensively, there is a danger that in the long term entire industries may lose their competitiveness," the Siemens report says.

Although data processing and electronic components by themselves occupy a relatively modest place within the total sales of Siemens and of AEG Telefunken, Germany's other major electrical/electronics group, the signs are that applications in the main volume products of electrical and power generation equipment and telecommunications are very much on the increase.

Dr. Bernhard Plettner, Siemens' chief executive, believes that eventually all mechanical steering and control of motors will be replaced by electronics. It is one major reason why 70 to 80 per cent of Siemens' impressive research effort has been integrated, thus cutting across the boundaries of the group's seven operating divisions.

In 1977/78 Siemens' R & D expenditure amounted to DM 2.3 billion, or 8 per cent of the group's total worldwide sales. It was the largest of any single company's in Germany and alone amounted to 12.5 per cent of all industrial R & D in the country.

Furthermore, Dr. Plettner adds, in

spite of substantial contributions from the federal Government, about 90 per cent of the group's R & D effort is self-financed.

The significance of microelectronics for the group is underlined by the microprocessor school founded in 1976, which is part of its customer service and which trained nearly 5,000 people from client companies up to the end of last year.

Dr. Plettner sees the principal applications in microelectronics in telecommunications, automobiles (of which more later), household appliances and engineering. Siemens and AEG are involved in all these fields: and he reckons two-thirds of all industrial employees in Germany will be affected by the advance of microelectronics over the next ten years.

Although at present a high proportion – some say over 90 per cent – of silicon "chips" used in Germany are imported, Dr. Plettner recognizes the need for home-produced chips, precisely because of growing significance of "miniaturization" in electronics.

Dr. Walter Cipa, chief executive of AEG Telefunken, shares this view. The intrusion of microelectronics into nearly every sector of industrial life, he said recently, had a profound effect on industrial change.

In addition to power generation and telecommunications equipment, Siemens and AEG are among the world leaders in the application of automated systems for all forms of transport, including air and spacecraft, ships, railways and motor vehicles. In some of these applications they claim to be ahead of the competition: for instance a world lead in modular display units for pilots incorporating separate or integrated minicomputers is claimed by AEG research chief Dr. Horst Nasko.

The AEG Telefunken group, a famous name in industrial history, has gone through a lean time in recent years, recording substantial losses and paying no dividend since 1974.

Its difficulties have been due partly to an unsuccessful entry into the nuclear power station market and the subsequent partnership with Siemens in the constructors Kraftwerk Union (to be dealt with separately in a later section). The partnership was eventually dissolved, with KWU now a wholly owned subsidiary of Siemens, but AEG has had to pay substantial compensation for losses as well as its share in the business.

AEG has also been feeling the effect of fierce competition in other fields, notably in office equipment, consumer electronics and television, from the US (in the first one), Japan and other Far Eastern nations. Its association with Nixdorf to penetrate the larger end of the computer market has also been unsuccessful.

Nevertheless, despite its troubles and the subsequent streamlining of AEG's activities, R & D expenditure is being steadily increased, with DM 930 million or 6.7 per cent of the turnover spent in 1978. "We have to do this to stay competitive at home and abroad," says Dr. Nasko. Like Siemens, AEG has to export to live while protecting, as far as possible, its domestic markets both in what he calls "bread-and-butter" items such as power station equipment, turbines, generators, control gear but also in the increasingly tough and sophisticated areas such as automation systems for industrial plant, electronics for seagoing vessels and steel mills and complete environmental protection systems.

Specialists

As already noted, Siemens and AEG, like their great international rivals, General Electric and ITT in the US and Philips in the Netherlands, are broadly based groups, with data processing only a part – though an increasingly significant part – of their business. On the other hand, Nixdorf, one of Germany's fastest-growing companies – almost 75 per cent increase in sales between 1974 and 1978, to exceed DM 1 billion for the first time – operates solely in the data processing field.

It was a deliberate decision by this post-war company's optimistic and dynamic founder, Herr Heinz Nixdorf, to attack a segment of the market where its comparatively modest size is of least disadvantage. Nixdorf manufactures computer systems for the small and medium-sized user, with an upper price limit of $ 200,000. It also supplies, if desired, terminals, peripherals and other support for large computers installed by the big international companies. As a justification of this policy, Nixdorf claims a 35 per cent share in the German market for computer systems costing less than DM 250,000 which is, owing to miniaturization and rapid fall in prices, the fastest growing segment of the market.

Although the company spends the equivalent of 7 per cent of its annual sales on R & D, the emphasis is on service to customers. "We are a bridge between our customers and technology," says deputy chief executive Herr Klaus Luft. Nearly two-thirds of the company's total workforce of 9,500 are engaged on servicing, including (says Herr Luft), "the provision of tailor-made solutions for individual customers."

The unsuccessful AEG-Nixdorf attempt to penetrate the large computer market has already been mentioned. Nixdorf's "people's computers" (as the chief likes to call them) are being offered to banks, commercial enterprises and public authorities. The company also has an eye on word-processing which, Herr Nixdorf thinks, will inevitably expand in the wake of data processing.

Nixdorf increased its exports by 55 per cent last year, the result being that US sales, at $81 million, accounted for 15 per cent of total sales.

One consequence of fast growth was the need for additional capital. Having rejected an approach from Volkswagen, Nixdorf's capital requirement was met by Deutsche Bank, with an eventual buy-back option for the additional shares issued. Some of Volkswagen's own surplus funds were subsequently utilized by the leading automotive company's acquisition of a majority holding in the office equipment and electronics company Triumph-Adler from its parent, Litton Industries, of the US.

Grundig, like Nixdorf, is a post-war company built up by one man. Again, like Nixdorf, Grundig serves a distinct segment of the market, consumer electronics and television. But unlike Herr Nixdorf, Herr Max Grundig, founder of the company, has retired as chief executive to become "president for life."

About half of Grundig's turnover is in color television, a further 18 per cent in recordplayers. In Germany itself, Grundig claims over a quarter of this fiercely competitive market.

During the past year, Grundig has opened (in Nuernberg) Germany's first factory manufacturing exclusively videotape recorders. Its sole product, a four-hour videotape with a recorder storing two programs for up to two days, is one of Grundig's answers to Japanese competition. Another one is a new color TV projector with which the company hopes to fend off competition from not only Japan but also the US.

Backing up these new as well as existing products is the technology center which has created the module principle.

We built a 12-cylinder engine. We built a 8-cylinder engine.

But the future has taken our decision for us.

It takes a long time to develop an automobile. Any serious manufacturer must begin planning ten years ahead of production. At BMW, we have probed the future, studying the demands that will be made of automobile engines, checking out the chances that one concept or another has of meeting these demands.

In the automobile of the future, luxury will never be a justification for waste

In a special research program, BMW have developed various engine concepts through to standard production level. Research and field tests combine to show us the right road ahead.

We have, for instance, built and tested a 4.5 litre 12-cylinder engine. We have also used our years of experience with previous V-8 engines to develop a 3.8 litre V-8 unit. The V-12 produces about 185 DIN kW (252 BHP), while the V-8 produces about 160 DIN kW (218 BHP). Both engines – like the famous BMW V-8 of the '50s – are made of aluminium throughout.

Challenging the best to find the ideal

To match BMW's current superb 6-cylinder unit, and at the same to satisfy the conditions of the future, any new engine must combine optimum power and acceleration with minimum fuel consumption. It must be effortlessly smooth-running, weigh as little as possible, and take up minimum space.

Power means more than performance. It means safety – vital safety

A car like a BMW 7 series or its competitors calls for an output of about 140 DIN kW (190 BHP). Even driven like a sports car, it needs no more than about 190 DIN kW (258 BHP). This means an engine of between 3 and 3.5 litres – a modest capacity for today, and by no means excessive for the future. Such a capacity calls for no more than 6 cylinders in-line – no need here for a V-8 or V-12 unit.

V-12: splendid but wasteful

BMW's V-12 develops splendidly smooth and powerful torque right up to its maximum engine speed. But most of the merits of a V-12 only show up at very high engine speeds – speeds which are of little value today, because high engine speeds mean very high fuel consumption.

Too many cylinders mean other problems as well: excessive weight and size, high production costs, complicated maintenance. The demands of tomorrow are for fuel economy and light, compact engines: the V-12 is not the way ahead.

V-8 versus 6-cylinder in-line: the comparison

The main improvements in smoothness come in the transition from 4 to 6 cylinders. The jump from 6 to 8 is a minor change by comparison. So two more cylinders mean little advantage – but significantly affect fuel economy.

The basic rule is simple: the more moving parts – or cylinders – an engine has, the more energy it takes to move them. An engine with more cylinders also has more bearing points, and so more friction to overcome. This alone means higher fuel consumption. But more cylinders usually mean more weight, as well – and the extra weight and size increase the thermal capacity which has to be achieved after the engine has been started. The result is again greater consumption, both initially, as the engine warms up, and, most noticeably, in stop-go traffic.

6 cylinders in-line: the engine concept of the future

6-cylinder engines will offer significant advantages as exhaust-emission regulations become stricter. Exhaust-gas catalysts with lambda probes are far more efficient with 6 cylinders than with a V-8, and it is far easier to use a turbocharger with 6 cylinders. So here, too, the 6-cylinder unit scores as the optimum solution for the future.

Research automobile with V-8 engine.

Research automobile with V-12 engine.

The engines of the future mu
the realities of the future. W
than the BMW 6-cylinder in-

The BMW 6-cylinder in-line engine is not merely outstanding today, but is the basis of a perfect engine for the future.

The automobiles of the future will have to be economical in every way – no matter how large, or how expensive. And to us, that means just one thing: We must find an optimum synthesis – maximum performance and smooth, effortless running on the one hand ... low consumption, minimum weight, and compact dimensions on the other.

Other manufacturers need larger engines or more cylinders to achieve the same performance and torque as the BMW 6-cylinder unit

To achieve the same low fuel consumption, they must build lighter – less safe – automobiles.

So for tomorrow's solution today, look to the BMW 6-cylinder power unit. Its refinement and its fuel economy will represent the ideal then, as now. And the new versions of the BMW 6-cylinder that power the BMW M1 show just what a performance range this engine and BMW's expertise have to offer: from 204 DIN kW

ust watch
hat solution could be better
ine engine?

(277 BHP) to over 588 DIN kW (more than 800 BHP).

Study the demands of the future in terms of weight and economy, and you will reach the same conclusions as many motoring journalists: In the USA, the trend is from 5 litre or 6 litre engines towards common sense and economy. It is, after all, in the USA – where huge multi-cylinder engines have lost their charisma as they have lost their future – that the BMW 733i has scored its greatest success.

In a survey of all automobiles currently on sale in the USA – a survey conducted every four years by the editors of the largest US automobile journal – the BMW 733i was recently chosen as the best saloon of all.

The reason surely, is this: Efficiency is now the dominant factor in the US automobile market.

The large BMW: standard-setter in engineering and design

BMW Cars
The BMW range of fine automobiles: the ultimate in performance, comfort and safety.

This provides for the majority of modules in each product to be interchangeable, allowing fast replacement of parts and remedying of defects.

The automotive industry: the drive for automation, fuel saving, clean air, and safety

With well over 4 million motor vehicles sold in 1978, nearly half of it for export, Germany is the world's third largest vehicle producer after the US and Japan, ahead of all its European rivals.

Research and development has a high priority in the industry's budgeting. Competition is one major but by no means sole challenge: there is also the threatening scarcity of energy sources forcing the pace of fuel economy; the ever-stricter pollution legislation and regulations, not to speak of matters of safety and the advent of microelectronics. For Gemany, the rising value of the D-mark against competing currencies and the relatively high wage costs are additional risk factors.

Germany's leading manufacturers react to this challenge in a variety of ways but all acknowledge the inherent dangers. Professor Ernst Fiala, management board member of the Volkswagenwerk (VW) responsible for R & D, sums the task up thus:

- A reduction in the size, length and weight of automobiles without reducing the inner space, *or* increasing the inner space without increasing size, length and weight, by optimum choice of materials and shape of vehicle;
- Reduced air and wind resistance to cut fuel consumption at higher speeds, aided once again by optimum size and shape;
- Reduced environmental pollution (air and noise) by altering the composition and nature of fuel (gasoline or diesel oil) or by changing the fuel itself (alcohol);
- Improved quality and comfort (quieter engine, better designed seats, etc);
- Increased use of electronics, including its use for the engine, the electrical parts and information for the driver;
- Fuel economy.

By 1983, Professor Fiala predicts, the mainline volume-produced automobile in Germany will over the next few years be equipped with electronic ignition and carburetor, a dashboard computer, an automatic braking system and "crash sensor" (airbag).

Some of these devices are already incorporated in higher-priced vehicles, particularly those exported to the US where the use of electronics is more widespread than in Europe. In this field the gap is seen not so much in technology but in the level of acceptance by the motorists concerned. Says Dr. Hans Bacher, general manager for technology at Bosch: "It is a question not only of price but of habit."

Nevertheless habits are changing. The price of new devices (particularly through the dramatic drop in the price of microprocessors) as a share of the total cost should decline too, if it were not for the fact that these devices have to withstand a great deal more from road conditions and temperature changes in the automobile than in the static conditions of office or factory or in the relatively calm progress of the aeroplane.

In overcoming customer resistance as well as road and weather hazards, automotive manufacturers work in close touch with Germany's biggest suppliers of automotive electrical and electronic gadgets, Bosch, which itself maintains close liaison with the main equipment manufacturers in the electrical and electronic industry, such as Siemens and AEG Telefunken.

The integration of various electronic controls in the vehicles is still a major problem facing designers. Professor Werner Breitschwerdt, the Daimler-Benz management board member responsible for R & D, in looking to the future has a single instrument in mind controlling tire pressure, ventilation, air/gasoline mixture, emission as well as combining engine, ignition, fuel injection and steering controls while also catering for temperature changes ranging from minus 50 degrees C to plus 120 degrees. "But there are still formidable problems to overcome," says Professor Breitschwerdt.

Last autumn the German automotive industry achieved one of its biggest break-throughs when the anti-blocking system ABS, jointly developed by Bosch and Daimler-Benz, was introduced as a standard fixture in the latter's higher-priced models and also in some of the BMW ones.

Research into the device, a major contribution to road safety, goes back at least 15 years: but it was only in the early 1970's that the problems associated with temperature-resistant materials and the efficient use of electronics was solved.

One deterrent to the more widespread use of ABS is the cost, which may be as high as 7 or 8 per cent of the price of more expensive automobiles. But, says Dr. Bacher of Bosch, a prototype already exists for a device which may cut the price to about one-third of the present one without endangering safety. It may be on offer in popular cars around 1983.

The new emission laws and measures to cut gasoline consumption in the US in view of serious shortages also pose problems to German manufacturers. It is not only a question of safeguarding exports, but also of facing the likelihood of fuel economy measures being gradually adopted in Europe as well. German executives, including Dr. Karlheinz Rademacher, BMW's research chief, find some emission regulations incompatible with fuel economy. He also deplores the mounting volume of new legislation affecting the automotive industry which, he says, rose by 400 per cent between 1965–78.

Fuel economy

Fuel economy may, of course, be a new concept in the US but it has always tested the ingenuity of automobile designers in a Europe without large indigenous energy resources. However, as Dr. Rademacher sees it, the trend everywhere (including the US) is for "big automobiles to get smaller and small automobiles to get bigger" which should stand the fuel economy-conscious German motor manufacturers in good stead.

One significant step towards fuel economy is the adaptation of automobile engines to diesel fuel (it is already widespread in the commercial vehicles field) which, with a relatively minor addition to the price has already produced spectacular savings of up to 60 per cent per gallon, particularly in city center driving. Already 45 per cent of Daimler-Benz's Mercedes automobiles are powered by diesel engines as are a growing proportion of VWs, while diesel-powered BMWs are also in sight. BMW has set up a diesel R & D center jointly with the Austrian company Steyr-Daimler-Puch and, according to Eberhard von Kuenheim, chief executive,

BMW will enter the diesel-powered car market "at the right time, with the right capacity and quality."

In 1978 only 530,000 (or 1.7 per cent) of the world's 32 million automobiles were powered by diesel engines. The forecasts point to an annual increase of 25 to 30 per cent, leading to a rise in the number of diesel-powered cars six to seven-fold, to about 3.5 million by 1985, or 8 to 10 per cent of the car population. It is reported from the US that customers are already willing to pay substantial surcharges on VW Rabbits.

At the same time, German manufacturers are not neglecting the development of traditional gasoline engines. BMW, for instance, has a research gasoline engine on its test bed which aims at cutting three of the six cylinders out while idling, and so achieve even greater fuel economy.

As already noted, technological developments, emission legislation and fuel economy measures in the US are of crucial significance to all German automotive manufacturers. The US is the largest single market for the sports car makers Porsche which takes up around half the company's total output; VW has its own US plant for assembling Rabbits; Daimler-Benz and BMW regard the US as their major export outlet, supplying it with their most highly "electronized" vehicles.

Indeed, Daimler-Benz's DM 10 billion investment program, announced recently, centers on the so-called "Little Mercedes" planned primarily for the US market. It will, says Professor Breitschwerdt, have a lower fuel consumption, more safety, quality and comfort than any previous model. But, he adds, "in view of the high technological input and more expensive materials, it will not be a cheap automobile."

Commercial vehicles

There have been significant incursions into the US recently by German commercial vehicle manufacturers, including Daimler-Benz and MAN, a member of the GHH group. Here again "dieselization" is the key: one of the leading domestic commercial vehicle manufacturers, International Harvester predicts that in the middle-range, Class 6 truck category (from 9 tons to 11.8 tons) alone, last year's proportion of 8 per cent for diesel-powered engines is likely to rise to 15 per cent by 1980 and 35 per cent by 1985 of a predicted sale of 200,000 vehicles. This would account for the entire increase in sales over the 7-year period.

To participate, with its diesel expertise, in the overall growth of the US truck market, Daimler-Benz is setting up its own assembly plant for Class 6 as well as Class 7 (up to 14 tons) trucks in the US. MAN has recently made an offer for a majority holding in White Motors, of Cleveland, manufacturers of trucks powered solely by gasoline engines. MAN is developing an entirely new diesel power unit which, Dr. Friedrich Laussermair, MAN's research chief, hopes will one day be in a position to exploit the growing popularity of the diesel engine in the US.

The third significant German commercial vehicle manufacturer, Kloeckner-Humboldt-Deutz, is a 20 per cent partner in IVECO (80 per cent Fiat) which has set up its own distribution network in the US to safeguard sales.

Traffic control

Leading German automotive and electronics manufacturers, Bosch, the federal Government as well as research institutes are co-operating in a number of experiments incorporating computer control for highway and city traffic. The ALI system developed jointly by Blaupunkt (a Bosch subsidiary), VW and Aachen College of Technology is already on trial in the Ruhr. Parallel to the anti-blocking system ABS, a radar collision warning system was developed by Bosch and AEG Telefunken which not only warns the driver of the vehicle but automatically involves his brakes and throttle. Differing guiding concepts for highways are being developed by the various auto manufacturers with Government support, ranging from computerized guidance to an electronic display road map built into the vehicle.

As an alternative to computer-controlled cars in congested city centers buses and rail-bound mobile cabins are being developed, again with Government support.

A joint Bosch-Daimler-Benz-Dornier project is already on trial in Esslingen, on the River Neckar; it involves a bus operating normally outside the pedestrian zone and then switching to a rail

track with minimal delay. Systems involving cabins guided by overhead cables have been developed by MAN, Siemens and Messerschmitt-Boelkow-Blohm (MBB) respectively.

Alternative fuels

The looming and, in some cases already present oil shortage, has forced German auto manufacturers to consider seriously the use of alternative fuels to gasoline and diesel oil.

One such fuel is liquid gas, propane or butane, which is a byproduct of crude oil refining and which is already in limited use in Germany and elsewhere. But liquid gas is still a product of increasingly scarce oil: so for long-term alternatives, the researchers are having to look elsewhere.

In view of its ample coal reserves, the only major indigenous source of energy, Germany has already used coal during two world wars for both automotive and aviation fuels. It is, however, uneconomic in peace-time and current research aims at more economic solutions. Daimler-Benz has coal-based methanol-fuelled vehicles on trial while VW prefers ethanol, a "biomass" product, in this case based on sugar cane. The experiments are being carried out in co-operation with the international oil and chemical companies, with the objective of reducing the disadvantages of alcohol (methanol or ethanol) such as poor start from cold, lower calorific value and boiling point, the former affecting performance, the last safety. A suitable mixture of gasoline and alcohol is one possible solution, greatly reducing the disadvantages while still saving gasoline.

Hydrogen, also coal-based, is being offered as a further long-term alternative. Daimler-Benz proposes to couple up household central heating systems (the tanks containing the gas supplying heat and air-conditioning) with "tapping" automobile fuel. The company, in co-operation with the University of Kaiserslautern, has already tested a vehicle powered by gasoline/hydrogen and hydride storage.

Another solution, the electric vehicle, has of course been in use for a considerable time in Germany and elsewhere: but until the problem of storage, replacing batteries with a more convenient source is found, the electric vehicle will remain only a short-range, city-bound alternative.

Aerospace lives on R & D

Considering the global size and importance of aerospace industry, with its huge demands on technology, advanced thinking and cash, the German industry is relatively modest. To speak of a "German industry" in isolation is, however, a misnomer: for it is closely integrated on the defence side with other NATO countries on the one hand; with other advanced industrial nations and Governments in commercial aviation and space programs on the other.

It is an industry which in Germany spends, with substantial Government support, the equivalent of about one-third of its sales on R & D. The three main companies active in this field are Messerschmitt-Boelkow-Blohm (MBB), followed by the Dutch-German VFW-Fokker group and Dornier GmbH – all well-known names in aviation history.

MBB, even at its fast rate of progress (DM 2.5 billion are the sales forecast for 1979, a 40 per cent increase in two years) is only a fraction of the size of, say, the Boeing Corporation, of the US. MBB nevertheless claims to be fully competitive in its field, including the technology incorporated both in its products and methods of manufacture.

For instance, an analysis in a recent issue of 'Aviation Week and Space Technology', a specialist US magazine, concludes that among the 15 top European companies and institutes in Europe applying computerized manufacturing techniques, MBB is probably among the leaders. The magazine's advice for US competitors is "to launch a fresh attack to maintain the lead in an area of innovative manufacturing technologies."

As for products, MBB is fairly broadly based, the range including combat as well as commercial aircraft, helicopters, missiles and spacecraft. One employee in five in the workforce of 22,000 is a qualified engineer or scientist.

Indeed, as Professor Gero Madelung, chief executive, points out MBB is one of the few aerospace companies in the world manufacturing its own helicopters (developed jointly with the Japanese company Kawasaki), with a future generation of transport helicopters already on the drawing board. It is a cooperative effort by French, Italian and British companies as well as MBB.

MBB's successful bread-and-butter commercial aeroplane, the A300 Airbus is the joint effort of four nations, France, Britain, Spain and Germany. Some options have already been placed for the sized-down version which is due to take to the air in 1982/83. Professor Madelung, a firm believer in European co-operation in the aerospace industry, adds with some confidence that "even the smaller Airbus is only a beginning. We'll undoubtedly broaden our range further as part of our strategy in facing competition, mainly from the US."

The Tornado multi-combat aircraft is another example of MBB's international involvement (British, German, Italian). In its other aerospace ventures such as a variety of communication, weather and television satellites, too, MBB co-operates closely with European companies, Government agencies and research institutions. And while well over 50 per cent of MBB's turnover is in defence contracts, Professor Madelung emphasizes that the company is spending between DM 35--50 million from its own resources on future development programs such as the new generation of transport helicopters.

Magnetic rail

As already noted, MBB is involved in advanced land-based projects as well as aerospace ones. Such a project is high-speed trains, carried out by the Government-backed Transrapid EMS consortium whose members include MBB, Thyssen Henschel (an offshoot of the well-known steelmaker), Siemens as well as a smaller company, Krauss-Maffei, better known for its combat tanks.

Krauss-Maffei is, like MBB and Siemens, based in Munich and, as chief executive Dr. Hans-Heinz Griesmeier points out, Leopard and Gepard tanks comprise only a quarter of its turnover. The company's future depends on commercial products such as advanced locomotives, process engineering, plastics machinery and others with a fairly high technological content. "We are a thought factory rather than manufacturers", is how Dr. Griesmeier describes his company.

Membership of Transrapid EMS is a logical step for a company whose start way back in the 19th century was associated with locomotives.

A Tradition of Progress

Over a century ago Alfred Krupp's invention of the seamless railway tyre made the name of Krupp known throughout the world. The Krupp three-ring emblem, adopted then to symbolize this invention, has since stood for products of the highest quality and pioneering work in the service of technical progress.

This tradition lives on at Krupp in the inventiveness and expertise by which we develop the products and processes of the future. Examples: coal degasification, water desalination, effluent treatment, plant and machinery for the cement, agricultural and foodstuff industries, marine electronics and fishfinding equipment, environmental engineering, giant wheel excavators for open-cut mining, tonnage, quality and special steels, satellite communications systems.

KRUPP

Fried. Krupp GmbH, P.O.Box 10, D-4300 Essen
Krupp Companies in North America: Krupp International, Inc., Harrison, N.Y. 10528 · Krupp Steel Products, Inc., Houston, Texas 77001
Polysius Corporation, Atlanta, Ga. 30339 · Krupp Canada Ltd., Pointe Claire, Que. H9R 4Z7 · Krupp Industries (Canada) Ltd., Edmonton, Alberta

S 25

Transrapid has developed a series of magnetically "levitated" trains traveling on an air-cushioned track. The latest of these, with a maximum speed of 250 km per hour, has already run on a test track near Emsland. A pre-prototype train with a maximum speed of 400 km is in the course of development.

The Deutsche Bundesbahn, the State railways, intends to renew its locomotive stock in the period from 1985 to 2010 which, says Dr. Griesmeier, "gives us, originally a locomotive manufacturer, our big chance." He adds that, in addition to speed, levitated trains use only between a third and a quarter of the energy of trackbound trains, uses pollution-free electric power and is less noisy

A three-pronged strategy for steel

While the world's steel industry is showing signs of recovery following the steep decline in the immediate post-oil crisis years, the leading western nations are steadily losing ground vis-à-vis the Communist block and some Third World producers. In 1974, Germany's crude steel output of 53.2 million tons, fourth biggest in the world, comprised 7.5 per cent of the world's total steel production. Last year's 41.3 million tons, slightly up on the 1977 figure, amounted to a mere 5.8 per cent share of the world's output.

The five leading western steel producers – the US, Germany, France, Italy and the United Kingdom – plus Japan contributed 53 per cent to the world's steel output in 1974; by last year this share declined to 46 per cent. Meanwhile the Comecon countries increased their share from 26 per cent in 1974 to nearly 30 per cent last year, with the Soviet Union alone raising its share to over 21 per cent, four per cent more than the US.

Other countries, too, especially China (now the world's fifth largest steel producer) and Brazil have been making rapid progress. For the first time South Korea and Taiwan, too, appear as serious competitors.

German strategy to counter the threat to a major basic industry has been threefold, varying from company to company in degree but common to all in principle. It is, firstly, the maximum use of high technology, including automation, process control and energy saving in volume production; secondly, a gradual transfer from volume production to high-quality, high-price special steels; and, thirdly, diversification into other industries, such as plant manufacture, engineering, etc, using steel as a raw material but no longer as a product for outside sale.

These measures are putting considerable pressure on the companies, particularly as increased investment has also to be directed into anti-pollution and noise abatement measures to satisfy tough federal legislation and resistance by local environmental groups.

The industry has its own high-powered and well-endowed research institute. At the same time, all the leading companies – Thyssen, Krupp, Salzgitter, Kloeckner, the German-Dutch Estel and Mannesmann (which no longer considers itself a steel company) – maintain substantial R & D programs.

Dr. Egon Overbeck, chief executive of Mannesmann, puts the case in a nutshell: "All our investments contain innovation: for without innovation, there is no future." He recalls the company's pioneering achievements in seamless piping and continuous steel casting as a proof of an innovating past.

Mannesmann is typical for the third approach in strategy. It is still one of the world's largest manufacturers for steel piping and tubing, including a substantial share in large diameter piping connecting North Sea oilfields with shore terminals: but the emphasis is shifting to plant construction and engineering, a sector in which Germany is already the world's largest exporter. By 1983 nearly half of Mannesmann's total world sales should be in plant construction and engineering compared with one-third in piping and tubing.

The group's overseas manufacturing subsidiaries are expected to make a substantial contribution to this shifting of balance towards plant construction and engineering from the present 15 per cent to 22 per cent of world sales. This would be largely as a result of expansion at the group's Brazilian and US plants.

However, Dr. Overbeck forecasts a continuing high demand in large diameter piping in the North Sea for at least another 15 to 20 years. Mannesmann, through its Demag subsidiary, is also involved in what may become a temporary extension to the oil age, the rich oil shale and sands of North America. A 6 km long conveyor belt has already been delivered, hopefully the forerunner of many more.

Dr. Overbeck has high hopes, too, for expansion in hydraulics in which the group's stake has increased considerably through the acquisition of Europe's largest hydraulics manufacturer, Rexroth.

While the other leading German steelmakers have not entirely followed the Mannesmann path out of steel, they are all diversifying along not dissimilar lines.

The contribution of the Thyssen group, Europe's largest private steelmaker, to high-speed train projects has already been noted. Its R & D activities cover a wide range, from numerically controlled machining and turbine blade improvements to corrosion protection, non-destructive testing of metals and optimization of the chemical composition of steel.

At the same time, high-grade steel for applications as varied as strip for motor vehicles, heavy structural steel for icebreaking bulk carriers and nuclear reactors is still a significant part of Thyssen's business.

For the Krupp concern, special steels have reached a turnover amounting to fully half of all steel sales. Quality and price are becoming ever more significant factors as users, particularly manufacturers of commercial vehicles, turn to cheaper substitute materials, including plastics and lower-grade steel.

Krupp, too, is developing its already considerable engineering side. Herr Heinz Petry, the chief executive, has recently pointed to a "fundamental structural change" in this famous concern, from being just steel fabricators and manufacturers of heavy machinery to one of Europe's foremost constructors of industrial plant.

As the group is increasingly forced to accept local participation in plant construction, this "dual structural change" does not always help in filling capacity in the German factories. "It presents a major test of flexibility to our workforce," says Herr Petry.

Both Thyssen und Krupp have major shipbuilding subsidiaries (as has Salzgitter), living under the shadow of worldwide recession and over-capacity. This makes continued profitability and expansion of other activities, partly through a high technology input all the more important.

The powerful and highly diverse Gutehoffnungshuette (GHH) group has now completely turned its back on steel

Managing the future with the BV Lion

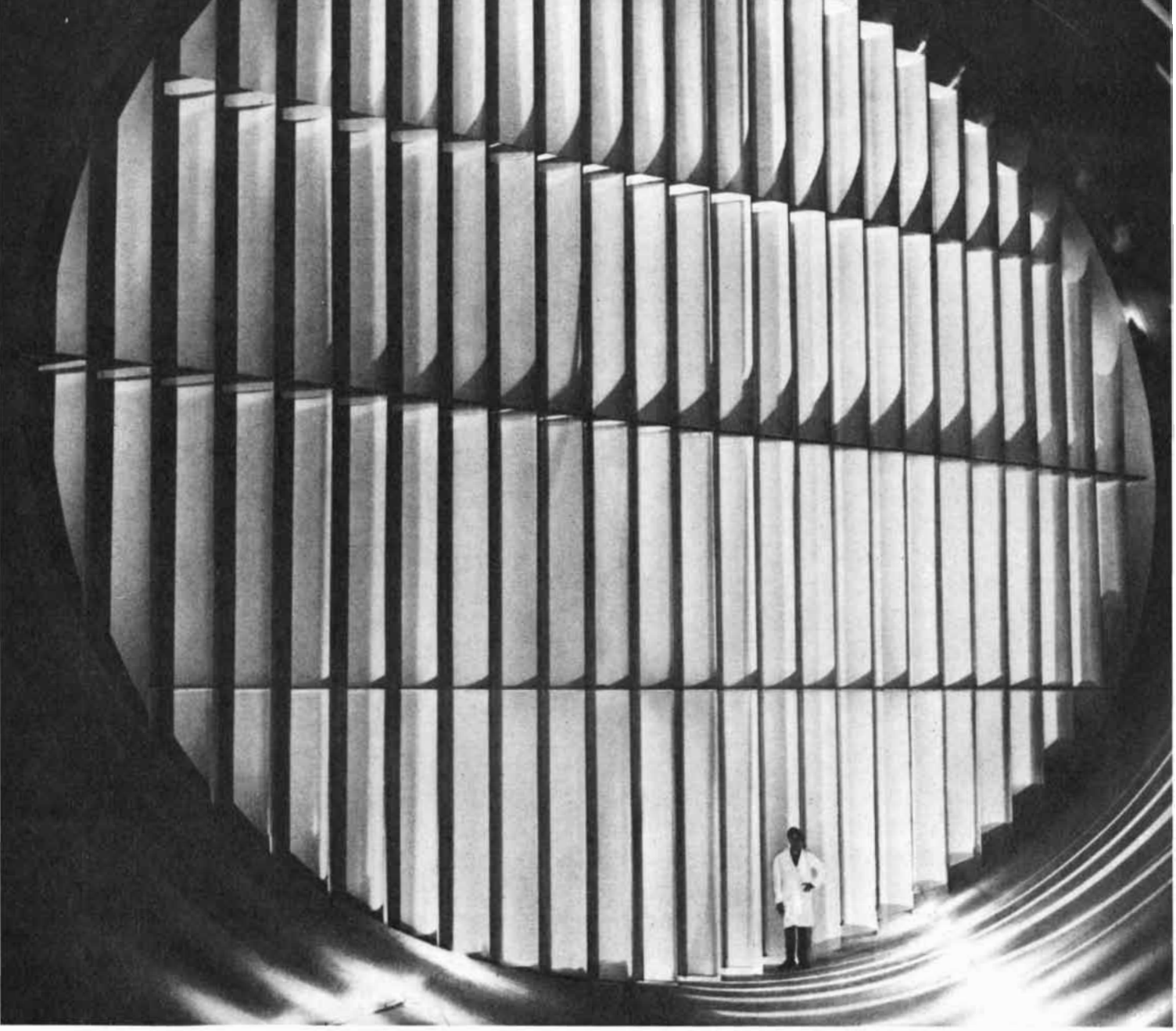
With a strong traditional background Bayerische Vereinsbank's approach to the future is both modern and creative. Hence our readiness to finance progressive new technologies such as the European Airbus, satellite research, transport and communication systems etc. Not only do our resources - BV's consolidated assets exceed DM 75 billion - go to finance the development of such technologies. Our bank's organisation relies on a broad range of sophisticated technical systems which refine and facilitate the innovative services we offer our clients worldwide. Linked internally by an electronic teleprocessing system, BV's international branch network is also connected to the international payments system SWIFT.

There are branches of Bayerische Vereinsbank in London and Tokyo, as well as a US network in New York, Chicago, Los Angeles, Atlanta and Cleveland operating under the name UBB – Union Bank of Bavaria. The Bayerische Vereinsbank's international presence mirrors our far-ranging expansion and diversification, covering financial centres such as Luxembourg, Zurich, Paris, Caracas, Hong Kong, Johannesburg, Rio de Janeiro and Bahrain.
If familiarity with the German markets applied by an international banking operation is what you need, why not get in touch with us?
The BV lion brings Germany nearer to you.

Bayerische Vereinsbank (Union Bank of Bavaria) New York Branch
430, Park Avenue
New York, N.Y. 10022
Telephone: (212) 758-4664
Telex: 12 67 45 ubb nykb

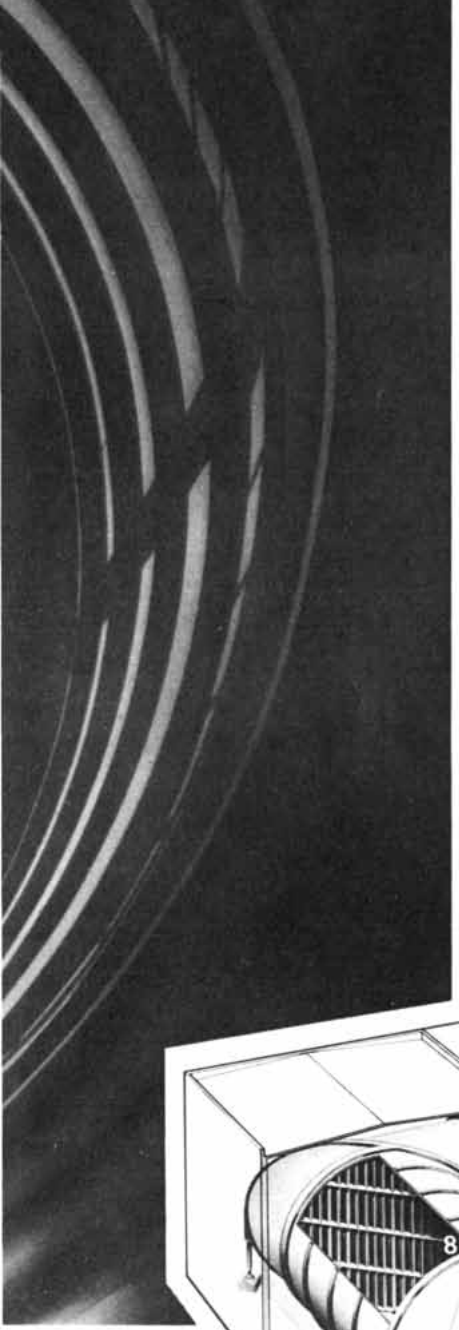
BAYERISCHE VEREINSBANK
INCORPORATING BAYERISCHE STAATSBANK AG

Bayerische Vereinsbank Head Office Munich International Division
Kardinal-Faulhaber-Strasse 1
D-8000 München 2
Telephone: (089) 2132-1
Telex: 52 99 21 bvmd
SWIFT: BVBE DE MM



"WHAT WE FOUND WITH THIS WIND TUNNEL WAS AS GOOD AS FINDING AN OIL WELL."

— Professor Ernst Fiala,
Director of Research and Development
Volkswagenwerk AG.



"In developing the Rabbit, we were able to reduce its aerodynamic drag considerably. This stretched its travel range by 1.2 miles per gallon. We've sold a million Rabbits. So millions of gallons of gasoline are being saved each year by this one research effort in the Wolfsburg climatic wind tunnel."

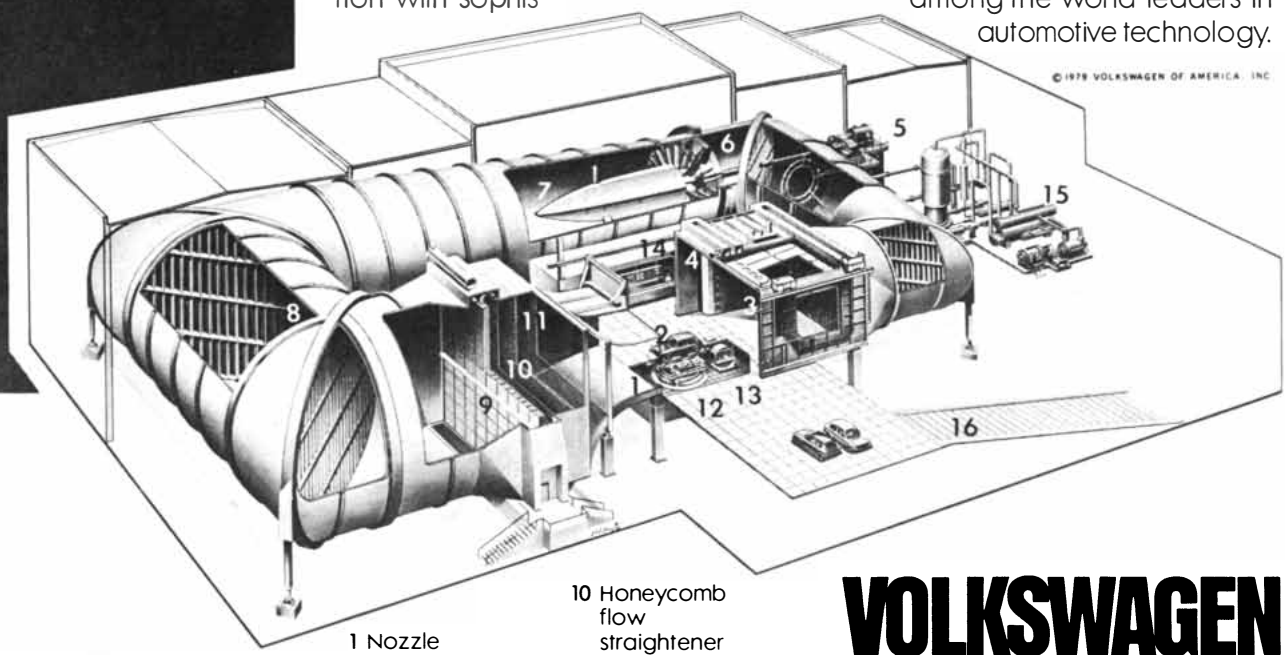
The benefits don't stop at gas savings. This is one of the few wind tunnels in the world capable of handling full-size vehicles and simulating an enormous range of weather conditions. With it, we can test the aerodynamics of engines and radiators as installed in the car. We can study interior climatization under dynamic conditions. And we can get a true picture of overall air flows and forces.

With this tunnel, we can heat the air stream to 113° F or cool it to -22° F. We can create wind velocities to 112 mph, rainstorms, 95% humidity, blazing sunlight. All this in conjunction with sophisticated measurement and control systems.

Since 1965, we have developed all our cars with the aid of this wind tunnel. Among other results, we achieved a 13% improvement in the drag coefficient of the Dasher resulting in a 4 to 5% fuel saving. By proper shaping and placement of the outside mirrors in the Audi 5000 and other cars, we can direct the air stream so the side window will stay clean in dirty weather.

In view of the fact that cars may spend half their power output at 40 mph pushing air aside and that power needs go up with the cube of the velocity, it's easy to see the importance of full-scale aerodynamic studies. They will surely have an important effect on the world supply of fuel.

The Wolfsburg wind tunnel is contributing to the evolution of fuel-efficient cars and continues to assure the place of Volkswagen among the world leaders in automotive technology.



1 Nozzle
2 Test section
3 Collector
4 Movable insulating cover
5 Electric drive with transmission
6 Fan
7 Diffuser
8 Corner vanes
9 Air cooler
10 Honeycomb flow straightener
11 Turbulence screens
12 Aerodynamic balance
13 Dynamometer test bench
14 Control center, process computer
15 Refrigeration plant
16 Access ramp to test section

VOLKSWAGEN DOES IT AGAIN

S 29

making. Members of the group include MAN which has spent DM 1 billion in the last ten years on R & D – its name appears more than once in this report – and its interests include some of the vital export areas such as engineering and transport, and energy which is just as essential to Germany's economic future.

"We're forced to export to survive," says Dr. Manfred Lennings, chief executive of the group. He appreciates that technology transfer to customer countries may follow, which is why GHH has to keep at least a step or two ahead in automation, mechanization and capacity utilization.

Other metals

Ores – copper, zinc, lead – are still being mined in Germany, but the mining companies have all become processors, refiners and appliers of the metals. These activities and the techniques developed as a result have inevitably led, as we shall see, to a considerable diversification of the companies concerned.

For example Metallgesellschaft (MG) has diversified into products such as plastics and rubber which compete directly with non-ferrous metals. This in turn has led to an expertise in chemicals on the one hand; to metallurgy and to engineering on the other.

Metallurgical expertise has brought custom and co-operation (as chief executive Herr Karl Gustav Ratjen points out) from the leading automobile companies among others. It has also overspilled into what may be called the group's "technological flagship," Lurgi, which is not only an internationally accepted authority in coal technology but also a considerable plant and engineering contractor. "MG earns its living with process knowhow," says Herr Ratjen.

The other important German mining and non-ferrous metal group is Preussag which has diversified into the processing of rare metals and substances for the electronics industry; into offshore oil and gas exploration engineering, tooling, services and prospecting; into firefighting materials and environmetal protection systems in general. A combination of some of these activities has produced participation in major international projects such as KBB in which Preussag has joined forces with the German steel and engineering group of Salzgitter and US partners to plan, construct and operate underground storage for strategic oil reserves in Louisiana and Texas.

A third major German company dealing in metals, in this case precious metals, is Degussa which has, among others, developed into a significant chemicals and pharmaceuticals producer as well as an originator of materials, processes and techniques for environmental protection.

"We think of ourselves as producers of specialties to industry rather than manufacturers of mass products," says chief executive Herr Gert Becker. "In our R & D activities we try to apply our extensive knowledge and experience in chemicals and metals partly to the development of our own technologies but also to their use by our customers."

One example of this knowledge and experience is Degussa's contribution to the purification of vehicle fuels and exhaust gases by catalysts, mainly to comply with strict legislation in the US and Japan. Other processes of air, water and waste purification and noise abatement are also among the group's technological achievements.

Degussa prides itself on its expertise in the recycling of precious metals, resulting in winning back as much as 20 per cent of used gold, 30 per cent of silver and 70 per cent of platinum.

The group's activities in the pharmaceutical and detergent sectors are referred to in the next sections of this report.

The high cost of new drugs

Research in the pharmaceuticals industry costs substantially more than fixed assets or production plant, says Dr. H. Gareis, of Hoechst. A recent survey carried out by Medizinisch Pharmazeutische Studiengesellschaft (MPS), an association set up by seven leading German companies, indicates that in the ten years to 1975, these companies spent DM 3.8 billion on research, averaging 15 to 20 per cent of their total sales.

The 43 pharmaceutical substances produced during the same period consequently cost almost DM 90 million *each* The seven, who include Germany's Big Three chemical groups – Hoechst, Bayer and BASF (through its subsidiary Knoll) – as well as C.H. Boehringer Son (Ingelheim) Boehringer Mannheim, E. Merck and Schering, represented two-thirds of the industry's total investment in research of DM 1 billion in 1975, but only 31 per cent of its domestic sales. This, in the view of Professor Dr. Jan Thesing, a director of Merck, "establishes the critical importance of exports in marketing the results of their research."

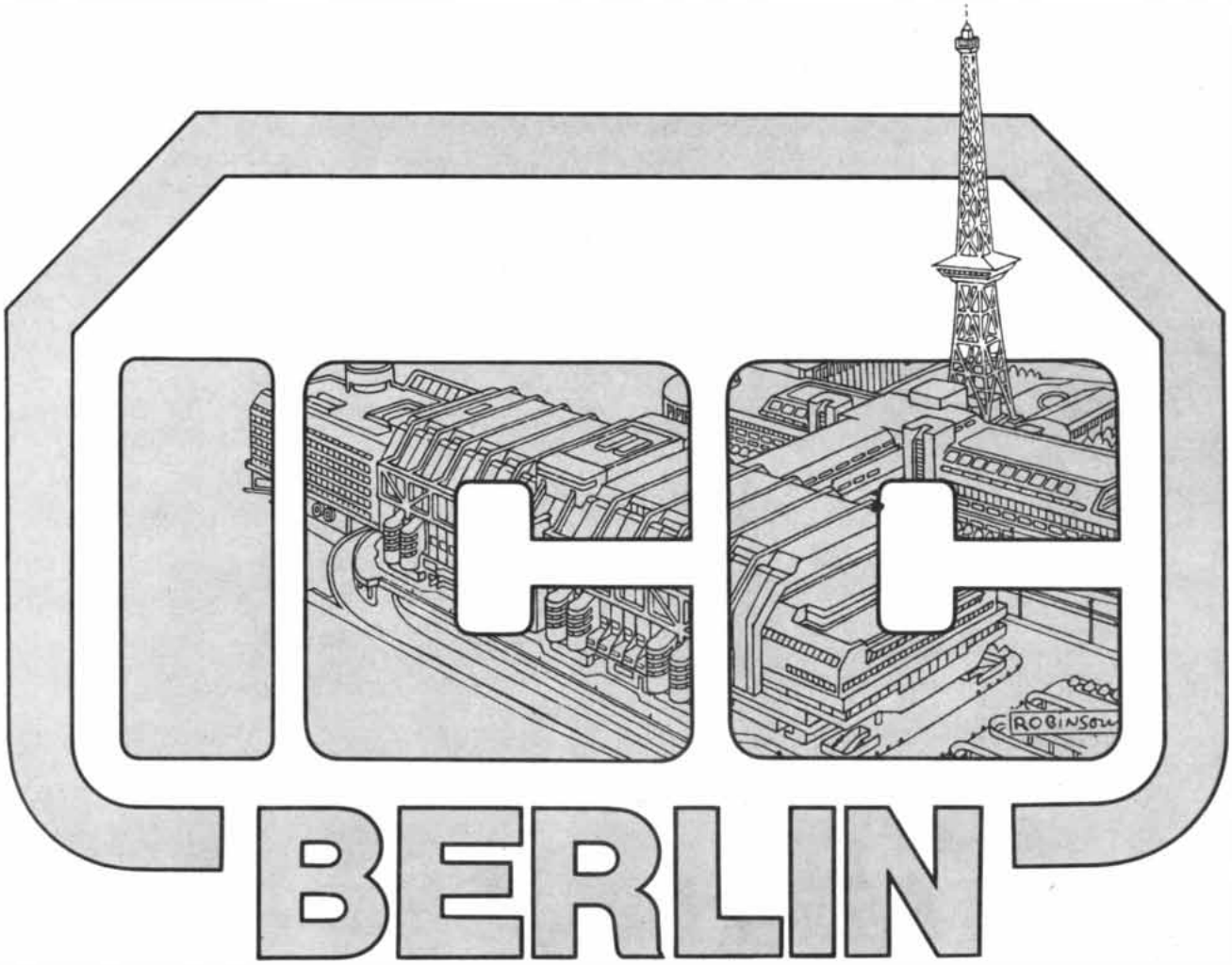
The survey analyzes one year's (1975) research expenditure which shows that the search for new substances amounted to only 31 per cent of the total, while effectiveness and safety absorbs fully half of it. This seems to point to a changing attitude to drugs following what Dr. Gareis believes to have been "the peak period of progress," 1930 to 1960.

"This attitude imposes one of the very big limitations on pharmaceuticals research," he says. "We are not prepared to run even the slightest risk when introducing a new drug: yet we know that our level of scientific knowledge does not allow us to predict a risk exactly . . . An increasing part of our research budget is devoted to work that makes no contribution to progress. Rather it is aimed at providing security against risks with inadequate resources."

Red tape, too, has grown enormously. Professor Dr. Karl Heinz Buechel, Bayer's research chief, complains that while ten years ago approval for a new drug required documentation running to 200–500 pages, documentation has now soared to a "mountain four meters high."

New drugs still trickle onto a highly competitive and congested market. The treatment and diagnosis of heart and circulatory ailments seems one major area of interest to all the leading companies but also to diverse groups such as Degussa which, as noted, is active mainly in the precious metals and chemicals field. But, as Professor Buechel points out, new applications for existing drugs is also gaining in importance. So is the search to reduce the risk of infection in say accidents or surgery: in this area Merck claims a significant advance.

One notable feature of the industry in Germany is the tendency for companies outside the Big Three chemical concerns to diversify and so spread the risks of long-term and costly research, only one new drug emerging from up to 8,000 substances tried. Virtually all the



More than just an exciting destination for your (inter)national meeting.

This is where things are happening. Through our contacts you can reach Science and Technology, Trade and Industry, Theater, Art and Business too. Give us a call or write. We'll send you the facts or brief you in person.

ICC BERLIN

International Congress Center Berlin

AMK Berlin
Company for Exhibitions, Fairs and Congresses Ltd.

Our Representative. Dieter H. von Schramm

International Congress Center Berlin
Congress Hall Berlin
Exhibition Grounds Berlin
Deutschlandhalle/Ice Palace Berlin

Box 19 17 40, Messedamm 22,
D-1000 Berlin 19

1625 Eye Street, N.W., Suite 1002,
Washington, D.C. 20006
Telephone (202) 659-0440, Telex ttm 64134

drug companies are now active in other areas: for instance, about half the sales of Schering, the world's largest single producer of the contraceptive pill, are now generated in other fields, including industrial and agrochemicals. It is to strengthen its industrial chemicals base, particularly in the US, that Schering recently acquired the chemicals product division of the US company Ashland Oil.

Merck, too, has diversified with a growing stake in the production of liquid crystals (in which it claims to have captured a large share of the world market) and in chromatography

A feedstock problem for chemicals

Germany's Big Three, mentioned in the previous section, are among the world's top five chemicals concerns (du Pont de Nemours of the US and ICI of the UK are the other members of the top group, their placings depending on the momentary state of the dollar and the pound sterling respectively). The German industry's expenditure in research and development, almost entirely self-financed, is expected to reach nearly DM 5 billion this year, well over a quarter of all industrial R & D in the Federal Republic.

In addition, the industry also plans to spend DM 5.5 billion on new plant and buildings. And despite its growing investment in overseas locations (Hoechst is now investing more abroad than on their German locations), the industry still exported more than 40 per cent of its output last year.

One reason for increased overseas investment is the slow-down in growth in Germany since the oil crisis of 1973/74. Overall chemical sales increased by only 2.7 per cent last year and the increase disguises continuing decline in certain important sectors such as synthetic fibers. Raw material supplies, too, particularly oil, is a source of growing concern for manufacturers, as is the downward trend of the dollar in which much of the industry's exports are calculated.

Professor Dr. Matthias Seefelder, chief executive of BASF, as president of the industry's association articulated its concern thus: price increases for petrochemicals averaged 30 per cent last year; those for naphtha (which competes with gasoline and of which the German chemical concerns used 14 million tons last year) even by 35 per cent. Naphtha supplies have also been affected by the increased octane rating for gasoline in the US made necessary by the reduction in lead. This year's oil prices are affected not only by the OPEC rises but also by the surcharge imposed by Iran.

As we shall see, securing the petrochemicals base played an important role in the recent VEBA-BP deal.

Much of the industry's R & D, says Professor Seefelder, is "defensive", aimed at ensuring the quality and applicability of existing products. "This naturally reduces the opportunity for 'offensive' research which now amounts to only 25–30 per cent of all R & D expenditure."

Of the rest, about half of total investment goes into ensuring the effectiveness and safety of products and further 25–30 per cent towards conforming to official regulations, including environmental protection. The latter cost the industry DM 2.7 billion in 1977 and the trend continues to point upwards.

Price and immediate availability of oil are not, of course, the chemical industry's sole worries. All indications point to growing scarcity towards the end of the century or even before: and, according to Professor Seefelder, the time-scale of gradually changing the feedstock for chemicals from oil to other raw materials is shortening.

At present 95 per cent of primary chemical products are based on oil or gas. In a recent statement, Professor Dr. Herbert Gruenewald, chief executive of Bayer, envisages no fundamental short-term change, at least until the mid-1980's. However, by the end of that decade, the trend towards coal-based chemicals may intensify: and this may be followed by increased utilization of oil shale and sands.

The large-scale commercial exploitation of "biomasses" – organic waste materials – may not occur before the next century.

In the meantime, the oil chiefs plead for a more rational allocation of oil and gas supplies: the chemical industry, after all, consumes only 6 per cent of the world's crude oil. One way of meeting this objective could be an acceleration of the use of alternative fuels for motor vehicles: as noted earlier, coal-based methanol or ethanol from sugar cane and other crops are being tested as feasible alternatives to gasoline.

Professor Gruenewald refers to the "reactivation" of coal-based chemistry – of which more later – and forecasts ammonia, methanol, methane and possibly ethylene glycol as the first chemical products derived from coal or coal-based synthesis gas.

He warns that the recycling of biomasses could entail tough competition with food crops, especially in underdeveloped countries. And he advocates an increased allocation of funds generated by 'convenient' and profitable industrial technologies towards R & D aimed at ensuring future supplies of raw materials.

In the meantime, economy in the use of raw materials as well as energy is a major objective of chemical companies. Professor Dr. Rolf Sammet, chief executive of Hoechst, has given spectacular examples of such savings: for instance, the amount of energy used for the production of ethylene powder at Hoechst has been reduced to a mere one-seventh of the 1963 levels. On the productivity front, the output of trevira filament per man hour increased from 9 kg in 1970 to 54 kg in 1978 and should reach 75 kg by 1980.

On the other hand, Professor Sammet adds, research costs have also risen dramatically. In the 20 years from 1950/54–1970/74 research costs per chemical substance rose 11.5 times and the development period for plant protection substances lengthened from three years to seven years.

The worldwide expansion of German chemical companies in terms of rising foreign investment has already been noted. Major acquisitions in recent years include the Vyandotte Corporation (US) by BASF, the French pharmaceuticals company Roussel-Uclaf by Hoechst and Miles Laboratories (also US) by Bayer. The last one enlarges Bayer's foothold in the US not only in its own specialist chemicals, medical and diagnostic fields but also in consumer products.

Bayer has made good productive use of the sale of its share in Chemische Werke Huels, an important German chemicals manufacturer, to VEBA which has thus acquired a majority holding in that company. VEBA itself has financed its increased holding in Huels from the DM 800 million generated by a deal with the German associate of British Petroleum (BP). This has enabled a streamlining of VEBA and the amalgamation of its chemical interests with Huels: and,

SIEMENS

Siemens quality means: advanced technology.

For example, office communications as swift as a voice can travel, as silent as a machine can print.

Siemens advanced technology in office communications covers everything from all-electronic data terminals, teleprinters and TWX terminals, with print speeds up to 270 cps, computer-oriented laser printers at 21,000 lines per minute, to the only all-electronic private business telephone systems that can be programmed in plain English.

This uncompromising commitment to quality by Siemens is the basis of our growth in a broad range of products and systems. In telecommunications, quality means advanced technology. In the medical area, exceptional reliability. In electronic components, thorough testing. In power engineering systems, material and energy savings.

With U.S. sales of over $300 million, a major capital investment in manufacturing plants, joint ventures, and field offices across the country, Siemens is in a strong position to make constructive contributions to the American industrial community. By sharing in the benefits of a worldwide R & D investment of $4 million each working day, Siemens and its customers can look forward to mutually beneficial growth.

If quality and advanced technology are important to you, talk to Siemens.

Siemens all-electronic teleprinter with mini-disk storage and editing unit.

Siemens... constructive contributions.

Siemens Corporation
Department 0400
186 Wood Avenue South,
Iselin, New Jersey 08830

equally important, the guarantee of 3 million tons of oil per year up to the year 2000 by BP strengthens as well as secures supplies for the group's petrochemical activities.

As VEBA chief executive Herr Rudolf Benningsen-Foerder points out, the deal has created an integrated oil company that operates at all levels, from drilling for oil (through the exploration company Deminex) and its processing to gasoline stations (through the distribution company ARAL) and trading (through Raab Karcher).

As for the enlarged Huels, Herr von Bennigsen's objective is to reduce the present 50 per cent share of mass-produced organic and synthetic chemicals in favor of more highly specialized and priced products. R & D expenditure is to be increased by 25–30 per cent to achieve this aim which, Herr von Bennigsen says, still has to be worked out in detail.

Specialization, sophistication, more efficient conversion rates, higher technology input are, of course, objectives shared by the entire industry. It applies to every sector: for example, as Professor Gruenewald points out, plastics are being improved all the time as substitutes for metals, not so much because there is less metal available but because plastics can be purpose-made, lighter, more resistant to corrosion. (Plastics sales by German manufacturers rose by over 8 per cent last year).

There is no alternative to the artificial fertilizer, he says, but, like drugs, they must have no harmful side-effects: "the balance of nature" must be preserved. (Sales of fertilizers declined slightly last year). On the other hand, sales of plant protection substances – fungicides, herbicides, insecticides – a strong line of business for the Big Three as well as other chemical manufacturers, rose by over 9 per cent: and the trend certainly seems to justify the decision by other companies, including Schering (see previous section) to develop it as a major outlet.

Detergents

One important sector of the chemicals industry is detergents which is the speciality of the Henkel group. It is Europe's second largest manufacturer of detergents and, in terms of sales (after

the Big Three), Germany's fourth largest chemicals group. It is also Europe's biggest producer of adhesives.

Henkel has been marketing some of the world's best known detergents and, in recent years, it has, in co-operation with Degussa, introduced a new product in which the phosphate content is halved without reducing the value of the detergent.

Apart from the fluctuating prices, imported natural phosphates are also a considerable polluter of water. Hence the search for a substitute ingredient which, says Dr. Bruno Werdelmann, the technology chief of Henkel, has led to "this new concept" in detergent technology.

Coal technology

Coal, both hard and soft, is Germany's largest natural resource not only for energy, as we shall note in the next section, but also as a raw material for the steel industry, for chemicals, and, indirectly, a possible substitute for oil, gasoline and natural gas.

The Ministry for Research and Technology estimates that German coal reserves of both varieties are likely to last for anything between 100 and 300 years. This lies behind the already considerable volume of research which has gone into coal technology this century, partly to supply the all-important aviation fuel and gasoline needed by the German forces during two world wars.

One such process with a long history, Fischer-Tropsch, is still used extensively and, to some degree for strategic reasons, in South Africa, a country with large coal but no oil or gas resources.

Coal technology languished in Germany during the years of cheap and plentiful oil. It has revived in the aftermath of the Middle East War and has gained further impetus recently by the Iran crisis, oil shortages in the US and Ireland, the nuclear stalemate in Germany and elsewhere and, as Professor Gruenewald of Bayer, has emphasized, by the threat to the raw material supplies for the massive German chemicals industry.

Among the world leaders for upgrading, converting and processing coal is Lurgi, a member of the Metallgesellschaft (MG) group. Lurgi's work, as summed up by chief executive Dr. Dietrich Natus, is fashioned by Germany's own historical experience: to adapt coal to new uses where coal is the dominant or sole natural resource (as in South Africa) or where oil and gas resources show signs of drying up.

Thus Lurgi is upgrading low-grade steam and soft coal to be used as coking coal – an increasingly scarce and expensive type of coal – in steel furnaces or as a substitute for natural gas in the direct reduction of iron ore. Or coal is gasified first to be used as a steel reduction gas or (as already mentioned) converted into methanol to be used in combination with gasoline in automobiles.

Lurgi as well as Krupp have designed plants for the conversion of coal into gasoline, diesel oil and petrochemical feedstock. Sasol, the South African company, which already uses the Fischer-Tropsch process in an existing plant is incorporating the Lurgi process in what is expected to be the world's largest coal gasification plant converting 9 million tons of low-grade high-ash coal into 2 million tons of liquids, mainly gasoline and diesel oil. In the US several companies are considering the production of synthetic natural gas (SNG) from coal.

Several other German companies, including those directly involved in coal utilization (such as STEAG, a subsidiary of Germany's largest coal producer Ruhrkohle), or gas production (Ruhrgas), are also experimenting, with Government participation, on new methods of upgrading or converting coal. Dr. Natus of Lurgi thinks scarcity and rising prices of oil and natural gas will, in the 1990's at the latest, force the world, particularly countries with ample coal reserves such as Germany, to make the best of those large but finite coal resources. "But we have to think well ahead in high-cost, energy importing Germany to survive," he adds.

Energy: the nuclear dilemma

If there is one major concern common to Government, public authorities and private enterprise in Germany, it is the question of energy. Adequate and reliable supply of energy is a condition for economic growth: technology may help towards better utilization and conservation but it cannot by itself create fresh resources.

At the heart of Germany's energy program for the rest of this century and some decades beyond is nuclear power. However, the expansion of nuclear electricity which was to have been speeded up after the 1973/74 oil crisis has come to a virtual standstill as a result of a national debate over safety and even the need for nuclear power.

In 1974 it had been hoped that by 1985 nuclear power would provide about 35 per cent of generated electricity and meet around 15 per cent of all energy requirements. Last year, the figures were 11 per cent and 3 per cent respectively: but as no nuclear power station has been sanctioned for four years and it takes 6 to 8 years from starting construction to commissioning reactors, the 1985 targets now seem unrealistic.

Germany imports most of its oil and increasing quantities of natural gas: its only major indigenous source of fuel is coal. Table VI shows the country's vulnerability to external events.

Although expansion in energy consumption has slowed down since the pre-Middle East War years, from 6–7 per cent a year to nearer 4 per cent, if forecasts of sustained economic growth (the prediction for this year is 4 per cent) come true, demand for energy will increase again. The events in Iran have shown, if proof were necessary, the very tight balance and sensitivity of energy supply and

TABLE VI **Structure of Primary Energy Consumption: 1970–78**

Year	Crude Oil %	Hard Coal %	Soft Coal %	Natural Gas %	Nuclear Electricity %	Hydro Electricity %
1970	53.1	28.7	9.1	5.4	0.6	2.5
1974*	51.5	22.6	9.6	12.7	1.1	2.0
1978**	52.5	17.7	9.1	15.4	3.0	1.7

* *Year after the Middle East War*
** *0.6 per cent is classified as "other sources"*

demand in the western world, including Germany.

The utmost importance of the nuclear issue is underlined by the creation of a nuclear Cabinet committee under the chairmanship of Helmut Schmidt, the Federal Chancellor. And recently Research and Technology Minister Volker Hauff presented his own energy "scenario" of a non-nuclear future to the Social Democratic Party of which he is a member. To replace the contribution of nuclear power with indigenous coal, he estimates, would require an annual output of 230 million tons of extra hard coal by the year 2000, or almost three times the present domestic production. Coal production in Germany has been declining for the past 20 years because of increasingly difficult mining conditions and high wage costs. That over 60 per cent of Germany's electricity is based on hard and soft coal is only made possible by a high level of subsidy coupled with high tariffs. Minister Hauff's scenario, even if the required additional 150,000 miners could be recruited, would obviously lead to even higher ones.

If the goal of reduced dependance on imported fuel were to be attained, there seems no alternative to nuclear power. (Indeed most of the top industrialists interviewed for this report emphasized this view. Security of energy supplies has been a recurring theme).

The most directly hard-hit by the four-year standstill in nuclear power station approvals have been the constructors, such as Kraftwerk Union (KWU) and Brown Boveri. KWU, now wholly owned subsidiary of the Siemens group, has six nuclear power station contracts of a combined value of DM 7.1 billion, or 29 per cent of its total, nuclear and non-nuclear power station order book paralyzed as a result.

What concerns Herr Klaus Barthelt, chief executive of KWU, equally is the possibility of a "brain drain" of highly qualified scientists, technologists and engineers. "We live ten years into the future," he says, referring to the time which elapses between the blueprint and commissioning of nuclear reactors. "There is a danger of our technology becoming out of date if we cannot put it into practice on our home ground."

That danger threatens not only KWU, Brown Boveri and other constructors. There are about 700 suppliers of equipment, nearly three-quarters of them small or medium-sized companies.

Dr. Hans Goehringer, chief executive of Brown Boveri, sees the problem from yet another angle. He puts Germany's reserve generating capacity at around 10–12 per cent of the total which, he says, could be used up by the mid-1980's even if demand rose at a below-average annual rate of 3.5 to 4 per cent which, in turn, assumes an even lower rate of economic growth.

Meanwhile, KWU, Brown Boveri and several other companies involved in Germany's energy program for the future are continuing with their efforts towards long-term, though admittedly minor solutions. For instance, MAN's energy division is involved, in addition to delivering gas centrifuges for uranium enrichment for the three-nation Urenco plant in the Netherlands, in a number of projects concerned partly in better utilization of fuels and partly in renewable energy sources such as the sun and the wind.

An example of the former is the utilization of waste heat in small diesel generators as an alternative to district heating whose extensive use, in the view of Dr. Friedrich Laussermair, MAN's research chief, is limited by very high costs and political difficulties. About a dozen of these small, fully automatic plants, delivering around 1 megawatt (MW) of electrical energy and 1.5 MW of heat, have already been installed in Germany. "Typical customers" are swimming pools, office and industrial buildings and residential areas.

MAN has designed a wind-powered unit to be built on the German North Sea coast which could, depending on wind conditions, deliver 3 MW of electricity or more. The company has also begun to construct, with a Spanish partner, a solar farm energy plant near Madrid: and in another experiment (in Western Australia) to couple solar energy with the waste heat of a diesel unit and so (according to Dr. Laussermair) "to raise the efficiency of a solar farm energy plant considerably."

KWU and Brown Boveri are also involved in solar energy experiments. Brown Boveri has delivered several solar units, the latest for the Federal Post Office to operate in Germany's highest peak, the Zugspitze.

These and other projects and experiments enjoy federal Government support in the form of direct subsidies, contracts and, when appropriate, tax allowances. Some, particularly in the nuclear field, are carried out in co-operation with other members of the European Community. An example is the SNR300 280 MW prototype fast reactor being built in co-operation with Belgium and the Netherlands.

A 300 MW high temperature reactor project (in which Brown Boveri is among those participating) has, in the longer term, the dual purpose of supplying electricity as well as process heat for, say, gasifying hard or soft coal. It may be coupled with a helium gas turbine (a prototype has already been built by GHH Sterkrade) which operates at fuel efficiencies of 80 per cent or over, nearly three times the present average for power stations.

However, the federal Government's growing anxiety over delays in the nuclear power program is reflected in the already mentioned increase in the proportion of aid to non-nuclear energy projects and encouragement to reinforce the base for conventional fuels.

In the short term, the deal between the UK-based oil concern BP and VEBA (in which the federal Government has a 44 per cent stake) will, as VEBA chief executive Herr von Bennigsen points out, help secure oil supplies (BP's guarantee to deliver 3 million tons of oil until 2000 has already been referred to), enable VEBA to improve the capacity utilization of its remaining refineries, strengthen the group's downstream activities (including the Huels chemical company, already mentioned) and their more efficient and profitable integration.

VEBA, one of Germany's leading public utilities, should thus be in a better position to compete with the strongly entrenched multinational energy concerns.

But this, as said, is in the shorter term. In the long run, Germany – like the US after the Harrisburg, Pennsylvania accident – will have to make a democratic decision on the future of energy supplies. Herr Barthelt, of KWU, does not think that "the lights will go out" altogether in the mid-1980's, even if no further nuclear power stations are sanctioned in the meantime. On the other hand, housewives may be asked not to use their electric appliances on Mondays and there could be general power cuts; but by that time it may be too late to make up for lost ground.

The address for challenges and innovation

Pipelines and facilities. Solving construction and plant engineering problems necessitates a comprehensive, well balanced spectrum of services and the employment of sophisticated organizational forms – And, of course, experience.

For example, one of Europe's largest gas distribution systems, as well as gathering, oil and products supply systems in numerous countries the world over. This ideal combination of decades of operating experience plus innovative engineering has given Pipeline Engineering the know-how and capabilities it needs to meet the challenge of ever-changing markets. Our portfolio of services permits solutions to be custom tailored to suit your particular needs. For entire projects or individual segments: Feasibility and viability studies, masterplanning, design engineering, purchasing, construction and assembly, commissioning, personnel training, maintenance, project management, logistics, budgeting, financing, cost control, project engineering, materials management and site supervision. For such applications as plant and off-site engineering, cathodic protection, pipeline transport systems, machine and process technology, electrical engineering and telecommunications.

Pipeline Engineering GmbH
A Subsidiary of Ruhrgas AG

Moltkestrasse 76, D-4300 Essen 1,
West Germany,
Phone (201) 184-1, Tx 8579741 ple d

Financing innovation and technology

It has been noted that last year business spent DM 14.2 billion on industrial research and development, or 77 per cent of industry's total expenditure for that purpose. There are some who believe that neither it nor the Government's contribution was quite sufficient to safeguard Germany's continued technological excellence.

Dr. Guth, of Deutsche Bank, referred to the Government's task in the statement cited earlier in this report when he said that in the long run substantial promotion of research and development is the "most productive cost" that exists. Yet, Dr. Guth emphasizes that any such promotion must be "on a general basis." The only exceptions could be high-risk, long-term projects such as the three-nation Airbus: otherwise any Government intervention, selecting companies and projects for R & D support was clearly incompatible with the free market economy.

Large companies have, as we have seen, little trouble in financing their own R & D efforts. This is not always so in the case of smaller companies.

There is, of course, (apart from the Government) a certain amount of institutional support available for innovative small and medium-sized companies. Four years ago Deutsche Wagnisfinanzierungsgesellschaft (Association for Financing Enterprise) was set up by the German banking community with an initial capital of DM 10 million, since then increased to DM 30 million, to advance equity capital and loans ranging from DM 200,000 to DM 2 million. The Government on its part has undertaken to underwrite 75 per cent of the losses suffered by the Association, up to a maximum of 15 years. By mid-1979 the Association was expected to have supported 30 companies up to within DM 2 million of its total capital resources. Nevertheless the banks themselves are the first to admit that this is merely scratching the surface. Small business still has problems in finding venture capital.

Medium-sized companies have another route to long-term risk capital (which innovation inevitably involves). They are called Beteiligungsgesellschaften (Partnership associations) in which financial and other institutions take up large minority holdings and, depending on the credit-worthiness of the new enterprise, the banks may provide overdraft facilities. This form of support, however, still presupposes substantial capital outlay by the innovator, though it has the advantage, as Dr. Guth points out, of being able to part with the partnership once the enterprise is on its feet.

Though top bankers recognize the possibility of an "investment gap", they dispute its role in any "technological gap" between Germany and other western nations, including the US. If there were such a gap, says Dr. Friderichs of Dresdner Bank, it is due more to the upward trend in the value of the D-mark, the higher wage costs, the environmental debate which not only hinders the building of nuclear power stations but also the export of nuclear knowhow. "Political considerations such as the non-proliferation of nuclear weapon technology" – a clear hint to objection by the US authorities to Kraftwerk Union's fuel reprocessing contracts in Latin America – also hinder progress.

Dr. Friderichs admits that financing innovation, including its attendant risks, could be met by the banks only to "a limited degree." He emphasizes – as does Herr Juergen Reimnitz, management board member of Commerzbank – that any bank credit to innovators must be backed by adequate security which some such projects obviously do not possess. "We set up the Association for Financing Enterprise to finance enterprise, *not* adventure," says Dr. Friderichs.

The problem, he adds, is often not finance but the quality of management: "We prefer a second-class idea by a first-class manager to a first-class idea by a second-class manager."

The question of the banks using technical consultants for assessing the chances of innovative projects is not excluded by Dr. Friderichs; nor is the possibility of establishing technical departments as have some leading US banks. All top German banks have, after all, introduced economic departments, a post-war development. Herr Reimnitz concurs: "We have no closed minds on the subject."

Nevertheless all top banks insist on a "substantial capital base" as a condition for credit, whether for innovative enterprise or any other one. Any easing of the credit situation would, according to Dr. Friderichs, entail a change in banking legislation coupled with an approval by the banks' supervisory authority to create "internal risk funds."

For foreign contracts incorporating high-risk long-term technical projects the State-backed Hermes Insurance Company provides an up to 80–85 per cent guarantee of the total price for the bank or banks financing those contracts. (Its usefulness to the contractors was amply demonstrated after the revolution in Iran, which put contracts by German companies an estimated combined value of DM 8 billion at risk).

Could a similar guarantee apply to the small innovative firm undertaking long-term product research and development?

According to Dr. Guth any direct aid of this kind would run against the principles of the free market. The only way to support such projects, he believes, is the extension of tax benefits already being offered by the Government to small and medium-sized firms.

Exhibitions, too, promote technology

The promotion of technology features in many exhibitions staged in Germany, though the methods of promotion vary somewhat.

The "multi-branch" concept is favored at one of the best known world fairs, the annual Hanover Fair held in April. Herr Hubert Lange, a member of the management board of Deutsche Messe- und Ausstellungs AG – the exhibition company – justifies the concept which, in his view, combines several complementary specialist fairs within a single show by claiming that technical buyers may, within a single visit, find solutions not only to their specific queries but to a whole series of complex problems which may arise during the visit.

For example, they may inspect not only certain electronic components relevant to their own business but also a range of finished products and their applications to data processing in the office, in information systems, in production control in engineering, in conveyor or warehouse systems and so broaden their horizons.

The multi-branch concept, says Herr G. A. Voment, Hanover Fair general manager, saves the buyer time as well as ex-

pense and at the same time enables him to establish contacts for which he would have had to visit several fairs. Similarly, it offers the exhibitor the advantage of showing his products to an audience much wider than he would if he had spread his resources over a series of specialist fairs.

Nevertheless, Hanover organizes specialist exhibitions as well: both concepts will be combined when Hanover stages fairs on behalf of entire countries, Brazil next year and Greece in 1981 (to coincide with that country's entry into the European Community).

Brazil already boasts of associates and branches as well as joint ventures with several leading German companies. The 1980 exhibition, says Herr Voment, is designed to show the extent to which Brazil has already become a supplier of components and parts to Germany, one of the outstanding examples of transfer technology from Europe and the US.

As an economically decentralized country, like the US among others, Germany has a string of cities and towns running exhibitions, some of which dispute the validity of Hanover's preference for the multi-branch concept. In this year's program for science and technology-based fairs, one finds Munich concentrating on single themes such as surgery, office equipment, laser optics, computer systems and their applications, electronic manufacturing. Duesseldorf shows, among others, information systems, educational aids, hospital equipment, the world of medicine, iron and steel technology, industrial heating systems, plastics and synthetic rubber. Dortmund offers consumer electronics and office data processing; Essen the role of chemicals in power stations and at a separate fair, the associated pollution problems; Nuernberg shows energy concepts; Frankfurt (a major exhibitor of consumer goods) chemical plant and, of course, stages the annual auto show.

According to Dr. Walter Marzin, general manager of the Munich fair, while composite exhibitions have a certain value to visitors from faraway countries and to State purchasers (such as the East Europeans), specialist fairs have the attraction of being up-to-date and more concentrated. They are normally staged only every two years or even less frequently, are extremely well attended and also have the advantage of developing from the "explosive growth" of certain science-based industries such as electronics.

Hence, says Dr. Marzin, data processing in the office is deliberately separated in Munich from electronics in manufacturing. Systems shows – pure software, dealing with computer applications – "are organized 100 per cent by users' groups"; medicine, banks, public authorities. Similarly laser optics are shown with the strictly practical objective of application rather than a theoretical exercise.

Duesseldorf, in the heart of the Ruhr, Germany's industrial powerhouse, is equally "user-orientated", with certain exhibitions – such as this year's IMPRINTA, an information systems fair showing computer-controlled printing plant and composition – spaced as far apart in time as five years.

Some of these specialist exhibitions attract up to 300,000 visitors, 40 per cent of them foreigners which, in the opinion of Herr Karl-Heinz Wismer, one of the managers, justifies specialization as well as the time lag. "It is difficult to produce a series of innovations at annual intervals," says Herr Wismer.

West Berlin, Greater Germany's capital but now a sadly divided city, is also a prominent exhibition city. To a greater extent than any of its German rivals, Berlin shows not only industries and their products but also sells itself as a prime industrial and technological location. It is to serve this twin objective that the new vast International Congress Center, one of the largest and finest in Europe, opened its doors in April this year.

As a western outpost deep in Communist East Germany, Berlin is heavily subsidized by the federal Government. It enjoys special tax rates and allowances, investment grants and special funds from both German and European institutional sources.

The city already boasts of around 160 centers of learning and research in the form of universities, higher education and research institutions as well as a substantial presence of high-technology industries. Between 25 and 30 per cent of Berlin's manufacturing output comes from the electrical and electronics industry, the plants including those of Siemens, AEG-Telefunken, Osram and Bosch. The capital equipment industry is the next largest.

As we have seen, one of Germany's leading pharmaceuticals and chemical companies, Schering, has both its headquarters and main production facility in Berlin. The company has recently completed a DM 600 million project comprising new and extended office buildings and laboratories, thus creating what amounts to a "new industrial town" in the city's Wedding district.

Foreign companies have a fair representation in Berlin, with more than 5 per cent of its manufacturing workforce and 6 per cent of its sales originating from subsidiaries and associates of US corporations. (These include IBM, Gillette, RJ Reynolds Industries, ITE Imperial and Litton Industries).

Just under a year ago, supported by the Federal and city authorities, the local chamber of commerce and the banks, Berlin set up its own promotion agency, Wirtschaftsfoerderung Berlin GmbH. It has, according to a spokesman, the advantage of being a private organization which is able "to cut corners" in its dealings with companies interested in a Berlin location. In other words, it can act more quickly, flexibly and decisively than a public authority: an advantage which this "island city" badly needs to halt depopulation and avert the threat of economic decline.

Our best seller since 1890

... and still going strong: the seamless steel tube. Springboard for one of the biggest tube manufacturers in the world. But Mannesmann makes more than just tubes: machinery and equipment for the steel industry, material handling systems, process compressors, hydraulics, machinery for the construction and plastics industry, turnkey industrial and utility plants. Mannesmann markets its own as well as complementary products the world over. Mannesmann is amongst the leading companies in the Federal Republic of Germany with more than 100,000 employees and over 180,000 stockholders. And an annual turnover in excess of $6 billion.

Ask the man from Mannesmann

Inquiries in United States may be adressed to:

Mannesmann Pipe & Steel Corporation
Houston/Texas 77056
Post Oak Central Building –
Suite 2300
2000 South Post Oak Road.

Three Lively Introductions to the Excitement of Science and Mathematics

New for College Courses in Basic Physics

Introductory Physics

Mashuri L. Warren, San Francisco State University and Lawrence Berkeley Laboratory
Consulting Editor: Edwin F. Taylor, Massachusetts Institute of Technology

Teachers and students will welcome this volume's concrete approach. *Introductory Physics* comprehensively covers the essential topics in basic physics, with an organization easily adapted to a wide range of courses for beginning physics students, students in the biomedical sciences and other students in the natural sciences.

Basic concepts in *Introductory Physics* are introduced with careful definitions, worked examples, and applications. Extended examples challenge the better student, and each chapter features a review of important principles and sources for further reading. Recognizing students' lack of full familiarity with mathematics, Mashuri L. Warren introduces mathematical concepts gradually, reintroducing techniques as needed.

Dr. Warren's approach emphasizes experimental physics: the basis of measurement, units of measurement, the evidence underlying important ideas. *Introductory Physics* explains the important interplay between theory and experiment, drawing on examples from engineering, medicine, industry, biology, transportation, and sports.

1979. 683 pages, 628 illustrations, hardbound:$18.00

New for Precalculus Courses

Algebra and Trigonometry with Analytic Geometry

Arthur B. Simon, California State University, Hayward

"This book has three purposes: to prepare students for calculus, to illustrate the strong relationship between algebra and geometry, and to present some interesting applications of mathematics."
—from the Preface

Arthur B. Simon achieves the goals he sets forth by making sure students get intense, active practice. Included in this book are more than 2,000 exercises, graded in difficulty, many with answers provided.

Students will appreciate the book's straightforward tone and helpful organization. Each chapter begins with an outline alerting students to important points, and most sections conclude with a convenient review. Each new concept is introduced intuitively and illustrated with several examples before an accurate formal definition is presented.

Teachers will appreciate the flexibility of this text, which is designed to fit the structure of a variety of courses. The Preface suggests outlines for both semester and quarter courses of different emphasis and duration.

1979, 533 pages, 460 illustrations, hardbound:$15.95

New in High School and Liberal Arts Mathematics

Elementary Algebra

Harold R. Jacobs

Elementary Algebra is a complete, day-by-day algebra course. Each lesson is suitable for one or two days' work and includes four sets of exercises that can be tailored to fit students' needs. Set I reviews earlier ideas and techniques; Sets II and III offer parallel exercises, with the answers to Set II in the back of the book; and Set IV provides more challenging exercises.

Elementary Algebra offers a complete instructional package, including a detailed Teacher's Guide, Test Masters, and Transparency Masters. Like Harold Jacobs' other acclaimed texts, *Elementary Algebra* not only explains ideas with extraordinary clarity but also captures the interest and imagination of students, who will find themselves drawn into mathematics even if they have resisted it in the past.

1979, 876 pages, 690 illustrations, hardbound:$12.00
Teacher's Guide, Transparency Masters, and Test Masters.

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104

Hewlett-Packard computer advances

The HP 300. A new generation puts high-powered, flexible computing into a small, compact console.

No larger than a free-standing terminal, the HP 300 is a new, low-cost business system that achieves a new level of computing concentration. Once programmed, it is friendly enough for people with no computer experience to use, yet packs a breadth of capabilities that will surpass normal expectations of a sophisticated applications designer.

The HP 300 computer is designed to simplify the development and control of dedicated on-line business applications. It incorporates HP's proprietary silicon-on-sapphire (SOS) integrated circuit technology: Exhaustively tested and refined, SOS semiconductor chips give the HP 300 the related benefits of smaller size, greater performance, and lower cost than computers using traditional technologies. The HP 300 is designed for office environments and plugs into a standard electrical outlet.

The HP 300 can handle up to 16 local or remote terminals in transaction processing environments, and two HP 300 Workstations with complete integrated display system (IDS) capability. Its "virtual memory" allows the use of extremely large programs and data sets without being confined to physical memory size. A powerful new operating system, Amigo/300, in conjunction with IDS, greatly simplifies user interface, and enables the HP 300 to perform both as a multiprogramming and multitasking system. For example, the HP 300 can print reports at the same time that higher-priority data entry or data base management inquiry operations are taking place at the system's terminals.

Eight programmable soft keys on the side of the video display screen can be captioned to give the operator a series of choices in performing a job. The display itself can be divided into multiple "windows" to view and control different parts of an application. These windows can even be scrolled individually, both vertically and horizontally, to a width of 160 columns and a length of several thousand lines.

Two widely used business languages, RPG II and Business BASIC, are available, as are System Language SL/300, IMAGE/300 data base management with an English-like inquiry capability, and TYPIST/300 text editing. There's even a HELP key that gives the user direct access to an on-line reference manual for the system. Full control of all job priorities and applications is maintained at the HP 300's central console.

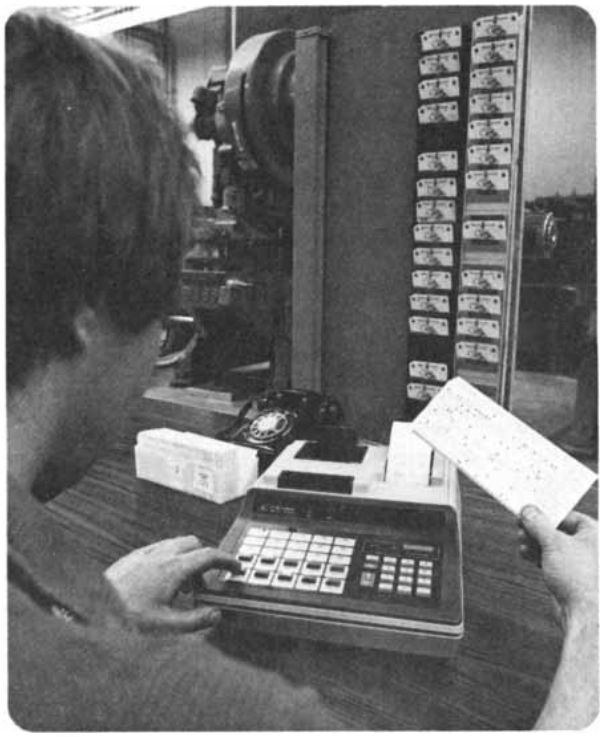
The price of the system is $36,500* with 256k bytes of main memory, 1 megabyte of mass storage on one flexible disc drive, and one 12-megabyte fixed disc. This can be expanded to 1 megabyte of main memory and 490 megabytes of disc storage. The HP 300 offers the capabilities, flexibility, and growth potential of computers costing two or three times its price.

If you'd like to know more about this new generation computer, the June and July issues of the Hewlett-Packard Journal explore its possibilities and technology in some detail. Just mail the coupon, and we'll be pleased to send you both issues.

extend your possibilities.

Factory data collection made economically viable.

With DATACAP/1000 software and HP data capture terminals, the HP 1000 computer can provide manufacturers with immediate, accurate information about production quality, inventory levels, or work-in-process.



HP—no slouch when it comes to manufacturing—has developed a factory data collection system that meets our standards of practicality and accuracy, as well as economics. We've been using it ourselves for several months and, based on results, think it will make its mark in the outside world.

The system is installed at one of HP's manufacturing divisions, in the plastic and metal fabrication shop. DATACAP/1000 was designed to increase the effectiveness of operations management. It is a general purpose, real time data capture software package that can be tailored to the user's application and installed without programming.

The user decides what information is wanted, and what data must be captured to provide this information. Then, through a simple, interactive process on a display terminal, the user creates specifications for data entry transactions. These include entry sequence, method (i.e. keyboard or card/badge readers), validation, and storage. Data can be stored on a disk file, an IMAGE/1000 data base, magnetic tape, or a combination of these. Transaction logging can be selected to provide an audit trail.

Data is captured and validated at the source using HP's new 3075, 3076, and 3077 data capture terminals. The terminals are simple to operate. HP employees typically learn to use them in a few hours, and are comfortable with them in a matter of days. Available in benchtop or wall-mounted versions, they come with more than 50 combinations of options to suit each work center. A variety of data entry methods can be used, such as keyboard, or badge or card reading.

Up to 56 data capture terminals can be linked to the HP 1000 and managed by DATACAP/1000 by using one or more simple, low-cost communication links. Each link can be up to five miles long, and uses a single, shielded, twisted pair cable.

Instead of filling out a card, the worker now walks to a nearby data capture terminal and pushes the attention key. When the terminal indicates that it is ready, the worker keys in a transaction specification number. The terminal then runs through a series of prompting lights which guide the worker through the entry of data: shift number, name, quantities, run hours, setup hours, etc. All keyed entries are subjected to validation procedures, so errors are caught and corrected on the spot.

As a result, the HP division's supervisors and managers now have access to timely and accurate information on which to base operational decisions.

HP DATACAP/1000 software is priced at $3000.* HP 3075A, 3076A and 3077A data capture terminals are priced between $2090* and $4230*, depending on the selected options.

HEWLETT hp PACKARD

1502 Page Mill Road, Palo Alto, California 94304

For assistance call: East (301) 258-2000, West (213) 877-1282,
Midwest (312) 255-9800, South (404) 955-1500, Canada (416) 678-9430

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

( ) HP 300 computer system
( ) HP DATACAP/1000 and data capture terminals

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

*Domestic U.S. prices only.

00947

Corals and Coral Reefs

Tiny coral polyps, living in symbiosis with photosynthetic algae, build huge limestone reefs that harbor more plant and animal species than any other ecosystem on the earth

by Thomas F. Goreau, Nora I. Goreau and Thomas J. Goreau

Man's ability to alter the surface of the earth is rivaled among biological organisms only by colonies of tiny coral polyps, which over aeons of geologic time accrete massive reefs of limestone. True reef corals are limited in geographical distribution to the clear, warm, sunlit waters of the tropical oceans; they are found in the great reef tracts of the Indo-Pacific and the western Atlantic. Reefs are important land builders in tropical areas, forming entire chains of islands and altering the shoreline of continents.

There are three major types of coral reef. Fringing reefs grow in shallow water and border a coast closely or are separated from it by a narrow stretch of water. Barrier reefs also parallel a coast but are farther away from it, are larger and are continuous for greater distances; the best-known is the Great Barrier Reef off the northeastern coast of Australia, which forms an underwater rampart more than 2,000 kilometers long, as much as 145 kilometers wide and as much as 120 meters high. Atolls are rings of coral islands enclosing a central lagoon, and hundreds of them dot the South Pacific. Consisting of reefs several thousand meters across, many of them are formed on ancient volcanic cones that have subsided, with the rate of growth of the coral matching the rate of subsidence. This explanation of atolls was proposed by Charles Darwin during the voyage of the *Beagle* and was confirmed in the 1950's by Harry S. Ladd and Joshua I. Tracey of the U.S. Geological Survey when their extensive drilling programs on Pacific atolls hit volcanic rock hundreds of meters down.

Although tropical ocean waters are impoverished in nutrients, having low concentrations of dissolved nitrates, ammonia and phosphates, coral-reef environments have among the highest rates of photosynthetic carbon fixation, nitrogen fixation and limestone deposition of any ecosystem. The reef ecosystem also probably supports a larger number of animal and plant species than any other. The key to this prodigious productivity is the unique biology of corals, which plays a vital role in the structure, ecology and nutrient cycling of the reef community.

The Biology of Corals

Because corals are sessile they were for a long time thought to be plants. In Ovid's *Metamorphoses* he refers to the coral as an organism that is soft under water but hardens on contact with air. (What he was actually seeing was the death of the living tissue, which exposed the hard skeleton.) In 1723 the naturalist Jean André Peyssonel proposed to the French Academy of Sciences that corals are animals. His view was derided, and he subsequently abandoned scientific work. Since then, of course, he has been proved right. Corals belong to the large and varied phylum of coelenterates, which are simple multicellular animals. The phylum's name is from the Greek *koilos,* hollow, and *enteron,* gut, because the main body cavity of its members is the digestive cavity.

The closest relatives of the true corals are the sea anemones, which corals resemble in basic body structure and overall appearance. The soft coral polyp consists of three layers of cells and is basically a contractile sac crowned with a ring of six tentacles (or a multiple of six) surrounding a mouthlike opening. The tentacles have the specialized stinging cells called nematocysts, which discharge an arrowlike barb and a toxin that stuns animal prey such as microscopic crustaceans. From the mouth of the polyp the short muscular gullet descends into the stomach cavity and is connected to the body wall by six partitions (or a multiple of six), increasing the area of the digestive surface. The free edges of the partitions are extended into mesenterial filaments: convoluted tubes that can be extruded through the mouth or the body wall.

The size of the polyps is highly variable, from about one millimeter in diameter in some species to more than 20 centimeters in others. Each polyp can give rise to a large colony by asexual division, or budding. Corals also reproduce sexually, producing free-swimming larvae that settle and establish new colonies. The most striking feature of coral colonies is their ability to form a massive calcareous skeleton. Individual coral colonies weighing several hundred tons and large enough to fill a living room are common in many reefs. In most species the polyps are in individual skeletal cups, some extending their tentacles to feed by night and some partially withdrawing into the cups by day. In the contracted condition the polyps can resist drying or mechanical injury at low tide, when some of the colonies may be exposed. The skeletal cups consist of fan-shaped clusters of calcium carbonate crystals, which are arranged in patterns that are characteristic of each coral species.

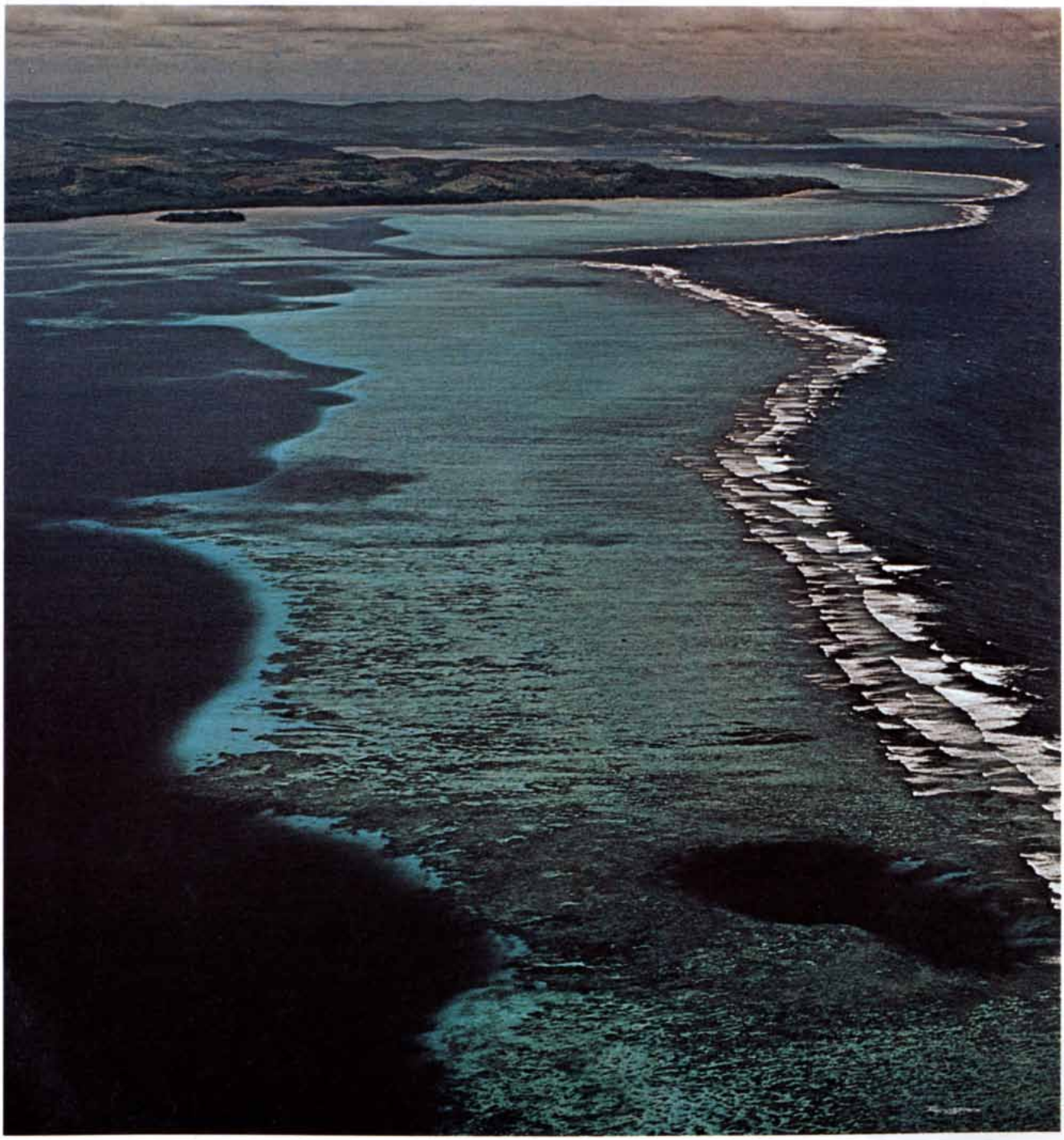
A remarkable feature of all reef-building corals is their symbiosis with the unicellular algae known as zooxanthellae. The coral polyps contain large numbers of these algae within cells in the lining of their gut. The zooxanthellae are yellow-brown marine algae of the family *Dinophyceae,* to which many of the free-living dinoflagellate algae also belong. The algae live, conduct photosynthesis and divide within the cells of their coral host, and on this symbiosis rests the entire biological productivity of the coral-reef ecosystem.

Since the zooxanthellae of reef-building corals need light for photosynthesis, such corals grow only in ocean waters less than 100 meters deep. The corals also require warm waters (above 20 degrees Celsius) and do not tolerate low salinity or high turbidity. Where deeper colonies are shaded by a dense overgrowth of shallower ones, the deeper colonies maximize their light-gathering capacity by growing in ramifications like the branches of forest trees. In shallow water, where light is abundant but wave stress is high, the colonies deposit robust branching skeletons; in deeper water, where light is scarce, the colonies form horizontal platelike structures in which each polyp may harbor an increased number of zooxanthellae. Under highly adverse conditions such as

prolonged darkness or freshwater flooding it is no longer advantageous for the coral polyps to maintain their zooxanthellae and they expel them from their tissues. Since the skeletal-growth rate of corals is dependent on their algal partners, true reef-building corals are almost never found outside the range of stable symbiosis.

Some coral species harbor no zooxanthellae; some of these species are found in crevices under the large structures erected by reef-building corals. Many of them are solitary cup corals such as *Astrangia,* which encrusts shells and rocks as far north as Cape Cod. Such corals can tolerate lower salinities, lower temperatures and greater depths: up to 6,000 meters in the deep sea. Even the deep, cold waters of the Norwegian fjords harbor great banks of *Lophohelia,* a colonial branching coral. Although these nonsymbiotic corals are distributed worldwide, their rate of growth is much lower than that of their symbiotic relatives, and they do not form massive reefs. In isolated instances colonies of these corals do contain symbiotic algae, but the algae do not appear to contribute significantly to the nutrition of their host.

The zooxanthellae are stored within individual membrane-bound cavities inside each of the cells in the stomach wall of the coral polyp. The feedback mechanism whereby the host regulates the number of its algal cells has not been determined, but there is little evidence that the corals "farm" and digest their



DIMES REEF off the Palau Islands in the western Pacific is shown in this aerial view. The shallow crest of the barrier reef is marked by the waves breaking over it. The Palau Islands, which are part of the Western Caroline group, are situated 1,060 miles southeast of Manila.

MOTOROLA HELPS GET ENERGY FROM

In northern California, the natural heat of the earth is being harnessed to serve the energy needs of man. Natural steam, heated by the molten rock below the earth's crust, is being tapped, extracted and piped to power some of the electrical generators that help light up San Francisco, 90 miles to the south.

The place is called The Geysers, and it's part of a pioneering effort by Union Oil Company of California to make geothermal energy a practical alternative to expensive imported oil.

The Geysers may be an unorthodox power source, but it has one thing in common with every other branch of the energy industries: the need to maintain good, reliable communications, no matter how remote or primitive the site.

At The Geysers, as in many places around the world, the solution to that problem is Motorola communications.

ELECTRONIC PROBLEM-SOLVING.

One problem at The Geysers, for instance, was the rolling terrain and steep, narrow canyons, among which conventional high-frequency radio signals could get diffused and lost. Motorola solved that one with ingenious simplicity: a low-band two-way radio system that, as one engineer put it, "gets into the nooks and crannies."

ELECTRONICS EVERYWHERE.

But this is merely the tip of the iceberg of Motorola's experience in energy-industry communications.

One of America's largest oil refineries has 35 Motorola

THE EARTH, THE SEA AND THE SUN.

systems and subsystems, among which are pagers that tell a man he's wanted on the phone; closed-circuit video monitoring systems; and alarm and control systems that not only tell when something is going wrong, but also when everything is working right.

In the North Sea, a Motorola microwave system will provide a data and voice-communication link that will help one person control six unmanned oil-production platforms. He'll be able to check pressures and flow rates, regulate meters, pumps and motors, all by touching a few buttons.

Some of the Motorola equipment on the Alaska Pipeline is so sophisticated that a hard-hat worker in the field can talk directly to an executive in an office a thousand miles away.

In Canada, specially designed Motorola equipment is in use at an oil mine, an extraordinary strip-mining process for extracting petroleum from tar sands.

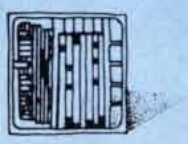
Motorola has made the apparently impossible happen by taking radio communication underground into deep-shaft coal mines.

And in solar energy, Motorola has gone beyond communications to actual energy development. Our engineers are producing photovoltaic systems that convert sunlight into electricity.

ELECTRONICS AND PEOPLE.

Motorola's preeminence in energy-industry communications is as much a matter of people as of technology. We made an early and total commitment to solving energy-industry communications problems, not merely as suppliers but as participants.

Microelectronics is at the heart of the matter, as it is in many of the things we do today. But if there are similarities among communications devices, there are none at all among the communications systems that the energy industries need in all their activities.



A microcomputer, drawn larger than life.

In designing these systems, Motorola brings to bear a combination of expertise and enthusiasm that helps us keep expanding the limits of what's possible in electronics.

MOTOROLA

Making electronics history since 1928.

For further information, write Public Affairs Office, Corporate Offices, Motorola, Inc., 1303 E. Algonquin Road, Schaumburg, Illinois 60196.

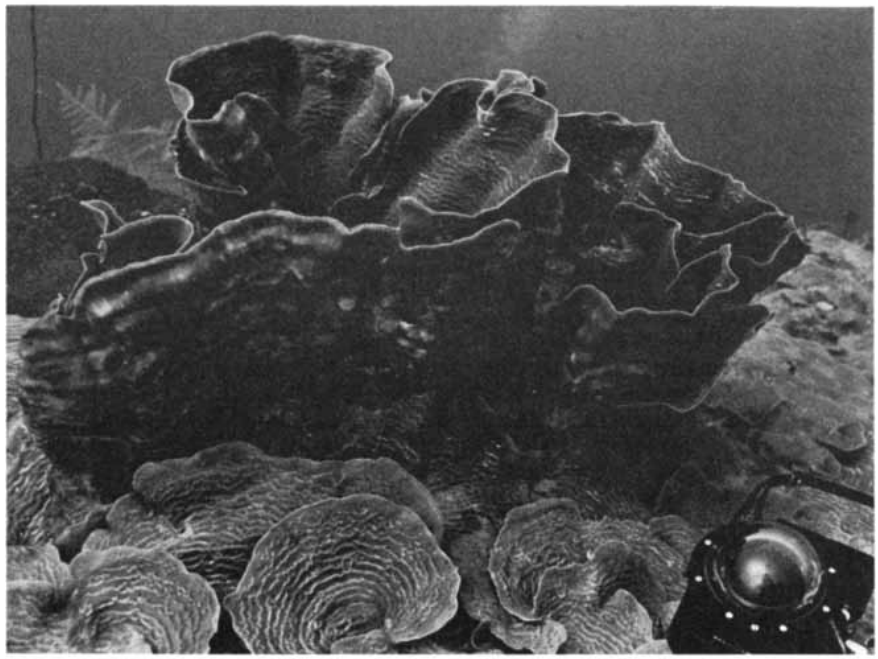
Motorola and Ⓜ are registered trademarks of Motorola, Inc.

STAGHORN CORAL is a species commonly found in the shallow, sunlit waters of tropical coral reefs. Along the reef crest, where the corals must withstand the mechanical forces of waves, the staghorn colonies are robust and have short branches. In the sheltered waters behind the reef crest the colonies grow taller and have longer, slenderer branches, as is seen here. Corals always grow toward the light.

LIVING CORAL POLYPS were photographed at sunset, when they emerge to feed. By day they retract into their skeletal cups and so are able to withstand drying if the colony is exposed at low tide. The polyps, which reproduce sexually and by asexual division, coat the entire surface of the coral skeleton. They feed on small plankton animals, which they stun with stinging cells (nematocysts) on their tentacles.

algae. Instead the coral polyps seem to control the population of zooxanthellae by extruding the older and less metabolically active algae.

Robert K. Trench and his colleagues at the University of California at Santa Barbara have shown that specific strains of zooxanthellae are adapted to specific coral species. Some strains can live successfully in several different corals, and some corals are not discriminating about the lineage of their symbiotic algae. The fascinating problems presented by the symbiotic selectivity of corals are only beginning to be explored, and corals provide a valuable experimental system for the study of cellular interactions in general.



TWO SPECIES of the coral *Agaricia* growing side by side differ strikingly in shape and size. One type has whorled fronds, whereas the other has shinglelike plates. Such complex morphological differences are produced by subtle environmental gradients, such as the decline of ambient-light intensity with depth. The corals are at a depth of 43 meters off Jamaican coast.

The Physiology of Coral Symbiosis

The modern study of the physiology of coral symbiosis began with a series of elegant experiments done by C. M. Yonge on the Great Barrier Reef Expedition of 1929. Yonge showed that symbiotic corals take up phosphates and ammonia from the surrounding seawater by day and release them at night. In order to study this phenomenon in greater detail two of us (Thomas F. Goreau and Nora I. Goreau) supplied carbon in the form of the radioactive isotope carbon 14 to reef corals. During the daylight hours the zooxanthellae assimilated the radioactively labeled carbon and photosynthetically fixed it into organic matter at a rate that was dependent on the intensity of the light. Some of this organic matter was then "leaked" from the algae to the coral host. Subsequent work by Trench and Leonard Muscatine of the University of California at Los Angeles and by David Smith of the University of Oxford showed that the leaked compounds include simple nutrients such as glycerol, glucose and amino acids. These compounds are utilized by the coral polyps in energy-yielding metabolic pathways or as building blocks in the manufacture of proteins, fats and carbohydrates.

It has long been known that the rates of metabolic reactions are strictly limited by the rates at which waste products are removed from the immediate environment. In higher animals the task is accomplished by specialized circulatory and excretory systems. These systems are absent in the anatomically simple coelenterates, which rely largely on the slow process of diffusion to remove soluble inorganic waste products such as carbon dioxide, phosphates, nitrates, sulfates and ammonia. The zooxanthellae, however, need for photosynthesis the very substances the coral polyp must get rid of, and they are believed to actively take them up from their host.

The photosynthetic demands of the zooxanthellae therefore result in the cycling of the coral's waste products into new organic matter. During the daylight hours the symbiotic algae produce more oxygen than the coral polyp can utilize for its respiration, and some of the carbon dioxide produced by the respiratory process is refixed by the algae into new organic matter. In order to estimate the efficiency of the internal carbon cycling in corals one of us (Thomas J. Goreau) determined the abundance in the coral tissue and skeleton of carbon 13, a rare but nonradioactive natural isotope, with respect to the abundance of the common natural isotope carbon 12.

For reasons it is not necessary to explain here, photosynthesis takes up carbon 12 slightly faster than it does carbon 13. Hence the organic matter synthesized by the zooxanthellae will have a relative preponderance of carbon 12, and a pool of carbon compounds enriched in carbon 13 will be left behind. It is from the compounds in this pool that the calcium carbonate coral skeleton is built. By determining the relative amounts of the two isotopes with a mass spectrometer it was estimated that about two-thirds of the carbon taken up in photosynthesis and calcification is recycled from the respiratory carbon dioxide of the coral polyp, with the rest being taken up from the seawater.

Organic matter leaked by zooxanthellae is only one of the three major sources of coral nutrition. Corals are efficient carnivores, immobilizing animal plankton with the stinging cells of their tentacles or trapping them on filaments of mucus that are then reingested. A polyp can detect a potential food item chemically, and it responds by extending its tentacles, by opening its mouth or by extruding its mesenterial filaments. James Porter of the University of Georgia has analyzed the content of the coral stomach and found that the polyps feed mostly on tiny crustaceans and wormlike plankton that hide in the interstices of the reef by day and emerge at sunrise and sunset.

Studies with radioactively labeled compounds have also shown that corals are able to take up dissolved organic matter across their body wall. Since corals actively feed on plankton, take up nutrients from seawater and absorb chemicals released by their zooxanthellae, they fill several ecological roles simultaneously: primary producer, primary consumer, detritus feeder and carnivore. This complex food web reduces their dependence on any single food source, which might be subject to random variation as environmental conditions change.

Calcification in Corals

Growth in corals is achieved by an increase in the mass of the calcareous skeleton and the overlying living tissue. The skeleton of corals is composed entirely of aragonite, a fibrous crystalline form of calcium carbonate ($CaCO_3$); calcite, the commoner crystalline form of calcium carbonate, is absent. In the reef many algae also deposit aragonite or a more soluble form of calcite with a high magnesium content. Working in Bermuda, Heinz A. Lowenstam of the California Institute of Technology showed that some calcareous organisms tend to deposit the less soluble calcite in

the cold seasons and the more soluble aragonite in the warm seasons, but the mechanisms by which organisms regulate the mineralogy of their skeleton are still unknown.

Coral polyps absorb calcium ions from seawater and transfer them by diffusion and by an active pumping mechanism to the site of calcification. Calcium ions are a major biochemical regulator of cell metabolism and must be kept at extremely low levels if the cells of a tissue are to function. Although coral tissues have a total calcium concentration similar to that of seawater, the concentration of free ions in them is much lower because most of the calcium is bound to membranes or to organic molecules. Lothar Böhm, working in our laboratory at the University of the West Indies in Jamaica, has shown that the calcium bound in these organic complexes turns over rapidly.

One of us (Nora I. Goreau), working in collaboration with Raymond Hayes of the Morehouse College School of Medicine in Atlanta, recently made detailed electron-microscope studies of coral polyps. In the course of these studies minute calcium carbonate crystals enclosed within membrane-bound vesicles were observed in the outer cell layer of the polyp. The crystals are extruded through the membrane to the coral skeleton, where they act as nuclei for continued crystal growth. This work may serve to clarify basic mechanisms of calcification in the cells of a variety of organisms, particularly because corals lack the hormonal controls over calcification that complicate these mechanisms in more advanced organisms.

The major obstacle to the study of the physiology of calcification in corals has been the difficulty of keeping corals alive and healthy in laboratory aquariums long enough to make accurate measurements of the calcium uptake. One of us (Thomas F. Goreau) circumvented the problem by measuring calcification in situ in the living coral reef. This was done by providing the coral with calcium in the form of the radioactive isotope calcium 45 and measuring the uptake of the radioactive calcium into the coral skeleton. The method is so sensitive that growth can be detected in corals that have been exposed to the radioactive calcium for only a few hours, which is what makes field studies practicable.

Such studies have shown that although reef-building corals grow under fairly uniform conditions of temperature, illumination and water circulation, there are very large differences in the growth rates of different species. The highest rates are invariably found in the branching corals, such as the West Indian elkhorn and staghorn corals. *Millepora* ("fire coral") is a close second, with the *Poritidae* ("finger corals") third. The massive corals grow more slowly. In the branching corals most of the growth takes place at the tips of the branches, and new branches develop almost anywhere on the older parts of the colony.

Factors Influencing Calcification

A crucial factor influencing the rate of calcification is the conversion of respiratory carbon dioxide (CO_2) into carbonic acid (H_2CO_3), which is in turn converted into bicarbonate (HCO_3^-) and carbonate (CO_3^{--}) ions. The enzyme responsible for the addition of water to carbon dioxide to form carbonic acid is carbonic anhydrase, which is present in high concentrations in corals. The subsequent formation of bicarbonate and carbonate ions is rapid and does not require catalysis by an enzyme. Drugs that inhibit carbonic anhydrase bring about a dramatic decline in the calcification rate.

The growth of the coral skeleton is on the average 14 times faster in sunlight than it is in darkness, and it can be decreased by drugs that block photosynthesis. Even daily variations in light intensity have a measurable effect on the

WORLD'S CORAL REEFS (*color*) can be divided into three basic types: atolls, barrier reefs and fringing reefs. Reefs of the West Indies are primarily fringing ones. Reef-building corals are found only in sunlit tropical waters (*broken lines*) because their ability to rapidly

calcification rate: the uptake of calcium is fastest at noon on a clear, sunny day, is reduced by 50 percent on a cloudy day and by nearly 90 percent in total darkness. The intensity of the ambient light also decreases with depth: the flux of light at a depth of 60 meters is only 4 percent the flux at the surface. As a result the rate at which calcium is deposited into the coral skeleton probably decreases rapidly with increasing depth.

The striking dependence of the growth of coral on the intensity of the ambient light is observed only when the zooxanthellae are present. If the symbiotic algae are removed (by keeping the coral colony in darkness for several months), the rate of calcification is low and is no longer affected by changes in light intensity, as is normally the case with nonsymbiotic corals. How do the zooxanthellae enhance the calcification rate? The answer seems to be that the fixation of carbon dioxide by the algae gives rise to an increase in the concentration of carbonate ions in the cells of the coral polyp through a series of linked chemical reactions, raising the *p*H of the fluid in the cells so that it is more alkaline. By precipitating its excess carbonate ions in the form of insoluble calcium carbonate the polyp is able to restore its *p*H to the normal level and at the same time build up its limestone skeleton. The zooxanthellae may also stimulate calcification indirectly by increasing the amount of free energy available for the active transport of calcium ions to the site of calcification. The algae therefore work synergistically with carbonic anhydrase to enhance the formation of calcium carbonate. Calcification can proceed in the absence of algal photosynthesis but only at a greatly reduced rate.

accrete limestone skeletons depends on their symbiosis with algae known as zooxanthellae.

The fact that calcification in corals is biologically controlled is further indicated by seasonal variations in the growth rate. These variations are reflected in measurements made by one of us (Thomas J. Goreau) of the concentration of the trace metal magnesium and of the heavy and light isotopes of carbon and oxygen in seasonal growth bands. Once the environmental and physiological influences that affect the growth of coral are better understood the variations in the composition of coral skeletons will provide a detailed chemical record of past environments much as tree-ring records do.

The synergistic effect of zooxanthellae on the calcification rate was clearly a decisive factor in the evolution of coral reefs. The development of enormous coral communities in the face of battering by heavy seas became possible only when the processes of calcium carbonate deposition became efficient enough for the rate of deposition to exceed the rate of loss through physical and biological attrition.

Reef Architecture

Coral polyps may not dominate the biomass (the total mass of living matter), the biological productivity or even the calcification in all parts of a coral reef. Nevertheless, the existence of many of the animal and plant communities of the reef is based on the ability of coral to build a massive wave-resistant structure. The dynamic interactions of the geological and biological processes that control the growth of coral reefs are well illustrated in the 150-mile fringing reef along the northern coast of Jamaica, which we have studied for the past 28 years.

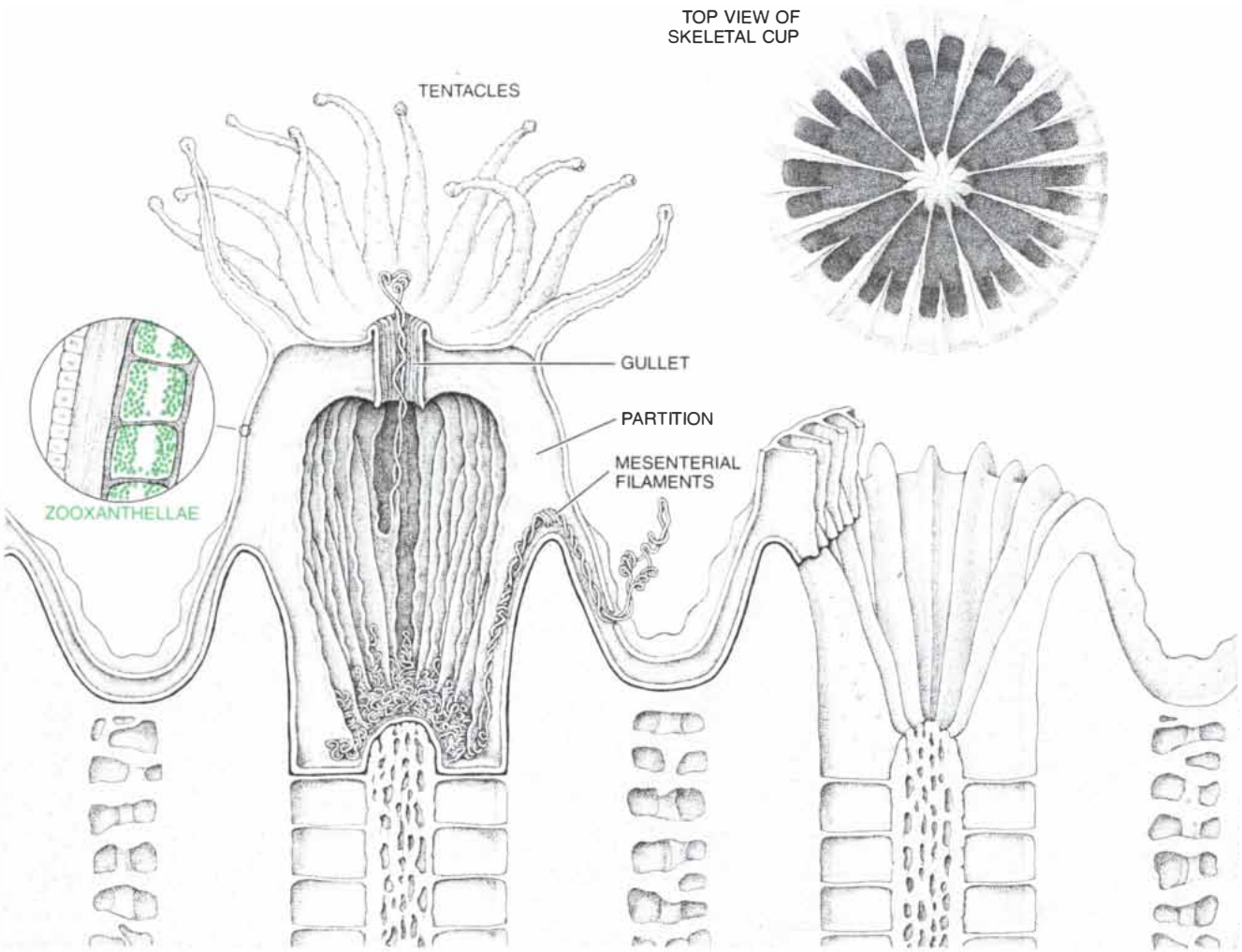
The major structural feature of the living reef is a coral rampart that reaches almost to the surface of the water. It is made up of massive rounded coral heads and robust branching corals, which build a rigid, cavernous palisade of intergrown coral skeletons. Living on this framework are smaller and more fragile corals and large quantities of green and red calcareous algae. The biomass of these algae is small compared with that of corals, but their productivity and turnover are so high that the sand consisting of their skeletal remains makes up the bulk of the calcium carbonate deposited in the reef.

Hundreds of species of encrusting organisms live on top of the coral framework, binding the coral branches together with their thin growths. Innumerable fishes and invertebrates also hide in the nooks and crannies of the reef, some of them emerging only at night. In addition sessile organisms cover virtually all the available space on the underside of coral plates and on dead coral skeletons.

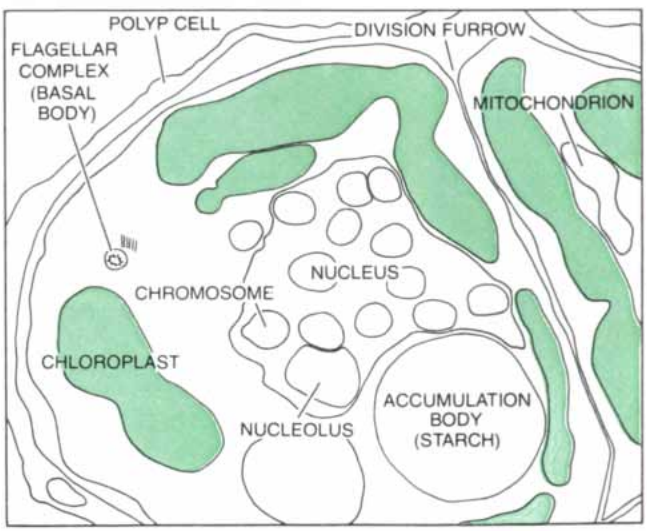
The crest of the reef runs parallel to the coast, in some places touching the shore and in others enclosing a sandy lagoon about five meters deep and up to a few hundred meters wide. This area is protected from the surf and is dotted with isolated coral heads. The lagoon is dominated by patches of calcareous algae and a community of bottom-living animals, notably sea urchins and sea cucumbers, which earn their keep by filtering organic matter out of the sediments or the overlying water. Many of these organisms graze on filamentous algae; if the grazers are removed from an area of the lagoon, a dense mat of algae forms after only a few days. The burrowing and churning activities of the grazers are important because they release nutrients created by the bacterial decomposition of organic matter buried in the sediments. Dense "lawns" of the sea grass *Thalassia* form special habitats harboring their own community of sea urchins, conchs and many other species.

Seaward of the reef crest is the fore reef, where corals blanket nearly the entire sea floor. The corals form massive buttresses separated by narrow sandy channels, down which passes a steady flow of fine sediment originating with the disintegration of dead corals, calcareous algae and other organisms. The channels resemble narrow winding canyons with vertical walls of solid coral growth. They may be as much as 10 meters deep, and some are completely roofed over with coral. This dramatic interdigitation of buttresses and channels dissipates wave energy and at the same time allows the free flow of sediments that would otherwise choke the growth of the coral.

Down from the buttress zone is a coral terrace, a slope of sand with isolated coral pinnacles, then another terrace and finally an almost vertical wall dropping into the darkness of the greater depths. The distribution of coral species and other animal communities in the reef is zoned by depth, a feature that enables paleontologists studying a section of an ancient reef now on dry land to accurately estimate the original depth of that section from the fossil animals associated with it. In water deeper than 100 meters few algae or symbiotic corals grow well because of the low light levels, and the fauna is dominated by animals that catch or filter the organic detritus sifting down from the reef above. The detritus feeders include the true sponges, the antipatharians ("precious corals") and the gorgonians (sea



ANATOMY OF THE CORAL POLYP is simple: the animal is basically a contractile sac made up of three tissue layers. The cylindrical body is topped by a central mouth surrounded by tentacles. From the mouth a muscular gullet descends into the central digestive cavity, which is connected to the body by a series of vertical partitions. The free edges of the partitions extend into mesenterial filaments. Cells in the lining of the digestive cavity harbor the symbiotic algae, which live, photosynthesize and divide within the host cells. The polyps sit in protective limestone cups consisting of a radial array of vertical plates, which interdigitate with the partitions of the polyp. Each polyp deposits new floors under itself as it grows upward. In the Tropics corals grow from one to 10 centimeters a year, depending on species.



DIVIDING ALGAL CELL, or zooxanthella, is enlarged 13,250 diameters in the electron micrograph at the left. The striated sacs within the cells are sections through a single large chloroplast, where photosynthesis takes place; the other cell organelles are indicated on the map at the right. The zooxanthellae greatly increase the metabolic efficiency of the coral host by absorbing the waste products of coral respiration and recycling some of them into new organic matter. They also "leak" essential nutrients to the coral polyps and enhance the rate of calcification. The electron micrograph was provided by Robert K. Trench of the University of California at Santa Barbara.

fans). Also common here are the sclerosponges, an ancient group that were major reef builders in the geological past but were long thought to have become extinct hundreds of millions of years ago. Our diving studies of the deep reefs of Jamaica showed them to be alive and well but displaced to deeper habitats by the faster-growing corals, which evolved later.

Reef Growth

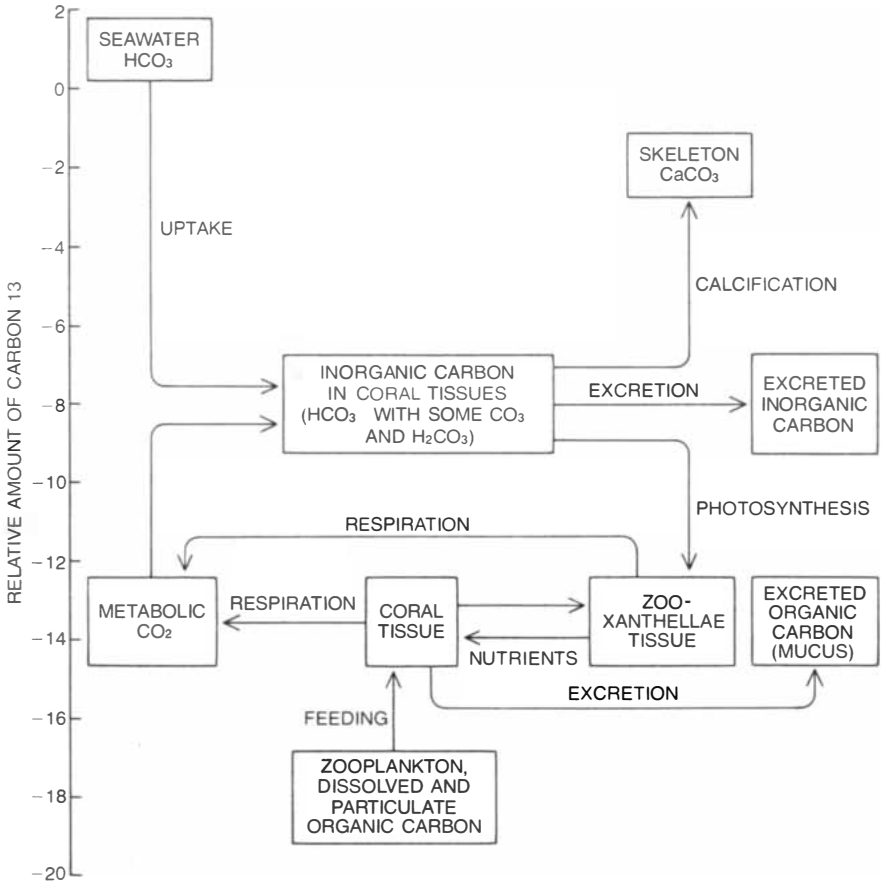
The growth of the reef is the result of a dynamic relation between the upward extension of the coral framework and the flushing away of a much larger volume of fine-grained detritus. The export of sediment from the reef is largely accomplished by gravitational flow and creep, either into the lagoon or down the channels of the buttress zone into deep water. Unstable piles of coral may also grow until they topple under their own weight and slide away. When the lower Jamaican reef was explored in the research submarine *Nekton Gamma II* at depths of more than 200 meters, enormous piles of sediment and huge blocks of solid reef were observed at the base of the drop-off; they may have been dislodged by earthquakes. Such dislocations create fresh substrates for encrusting organisms and help to establish coral communities on the steep lower slopes, particularly the platelike whorled colonies of *Agaricia*.

Two other major processes influence the growth of the reef: biological erosion and submarine lithification. Many species of filamentous algae, fungi, sponges, sea worms, crustaceans and mollusks bore into coral skeletons, excavating holes by mechanical rasping or chemical dissolution. The commonest is the boring sponge *Cliona,* which saws out tiny chips of calcium carbonate; the chips are a major component of the fine sediments. *Cliona* can riddle a coral skeleton with holes without damaging the living coral polyps. In the deeper waters many corals grow in flat, thin sheets to maximize their light-gathering area and hence are quite susceptible to erosion by borers, which can cause the corals to break off and fall downslope. In some places, however, the coral is so overgrown with encrusting organisms that it remains in place even though it is no longer directly attached to the reef.

Counteracting the effects of biological erosion is submarine lithification: the deposition of a fine-grained carbonate cement in the pores and cavities of the coral skeleton. Sediments trapped in the reef framework are rapidly bound together by encrusting organisms and the calcareous cement. The origin of the cement is not yet clear; it may be an inorganic precipitate manufactured by bacteria that live in the crevices of the reef. Studies at the Discovery Bay Marine Laboratory in Jamaica done in conjunction with Lynton S. Land of the University of Texas at Austin showed that once the cement has hardened it is in turn bored and refilled; the filled holes are apparent when thin sections of the aggregate are examined under the microscope. Submarine lithification results in the outward accretion of the fore reef and stabilizes the steep profile of the drop-off wall. The growth of reefs is therefore the product of a dynamic balance among framework growth, sediment transport, bioerosion by borers, mechanical destruction and submarine lithification, with the relative importance of these factors varying from reef to reef.

The living reef is basically a veneer growing a few millimeters a year on top of a complex topography of superposed ancestral reefs. In Jamaica as much as nine meters of reef has built up since the present sea level stabilized some 5,000 years ago. The ancient reefs remain, providing a record of changes in sea level and of the uplift of land by the movements of tectonic plates.

The rise and fall of the sea level over the past few million years has been caused by changes in the volume of water tied up in land glaciers and ice sheets during the Pleistocene ice ages. When ice sheets grew in the Northern Hemisphere, the sea level dropped and coral reefs were stranded above the waterline. Today fossil ridges and wave-cut notches mark the ancient sea level. A succession of stranded reefs are found in Jamaica, Barbados, New Guinea and on other coral coasts; these reefs were formed 80,000, 105,000, 125,000 and 200,000 years ago, when the climate was warmer and the sea level higher than it is today. Conversely, in Jamaica a series of drowned and overgrown ridges can be seen at 25, 40 and 60 meters below the present sea level. These drowned reefs were formed during periods of intensive glaciation 8,000, 11,000 and 14,000 years ago, when the sea level was considerably lower than it is today. The ancient reef is therefore a dimly visible palimpsest under the living reef, like a medieval manuscript that has been repeatedly erased and written over but shows faint traces of its history. Such



CYCLING OF CARBON among the zooxanthellae, the coral host and the environment is outlined. The various carbon pools are plotted on a vertical axis according to the ratio of the two stable isotopes of carbon: carbon 12 and carbon 13. The position of each pool is therefore an indication of the relative importance of the processes by which each pool gains and loses carbon and the extent to which these processes utilize one of the two isotopes preferentially. For example, because photosynthesis takes up carbon 12 faster than carbon 13 it leaves behind a pool enriched in carbon 13, from which the calcium carbonate of the coral skeleton is formed. About two-thirds of the carbon utilized in photosynthesis and calcification is recycled from respiratory carbon dioxide. Level of carbon 13 in coral tissue reflects the composition of its food sources.

features help in establishing the chronology of the Pleistocene ice ages and the volume of water added to the oceans by the melting of the ice.

Reef Ecology

The history of the modern Jamaican reef since the sea stabilized at its present level 5,000 years ago has not been long enough to establish a climax community: an ecosystem in equilibrium. This fact is evident from the almost haphazard development of reefs along any coral coast: some areas have well-developed reefs and others have only isolated patches of coral. Often there are no obvious environmental influences or catastrophic factors (such as earthquakes or tidal waves) that would explain such differences in development. It seems rather that chance variations in the settlement of free-swimming coral larvae and growth play a major role in determining the formation of reefs, and that there simply has not been enough time for corals to occupy all favorable habitats.

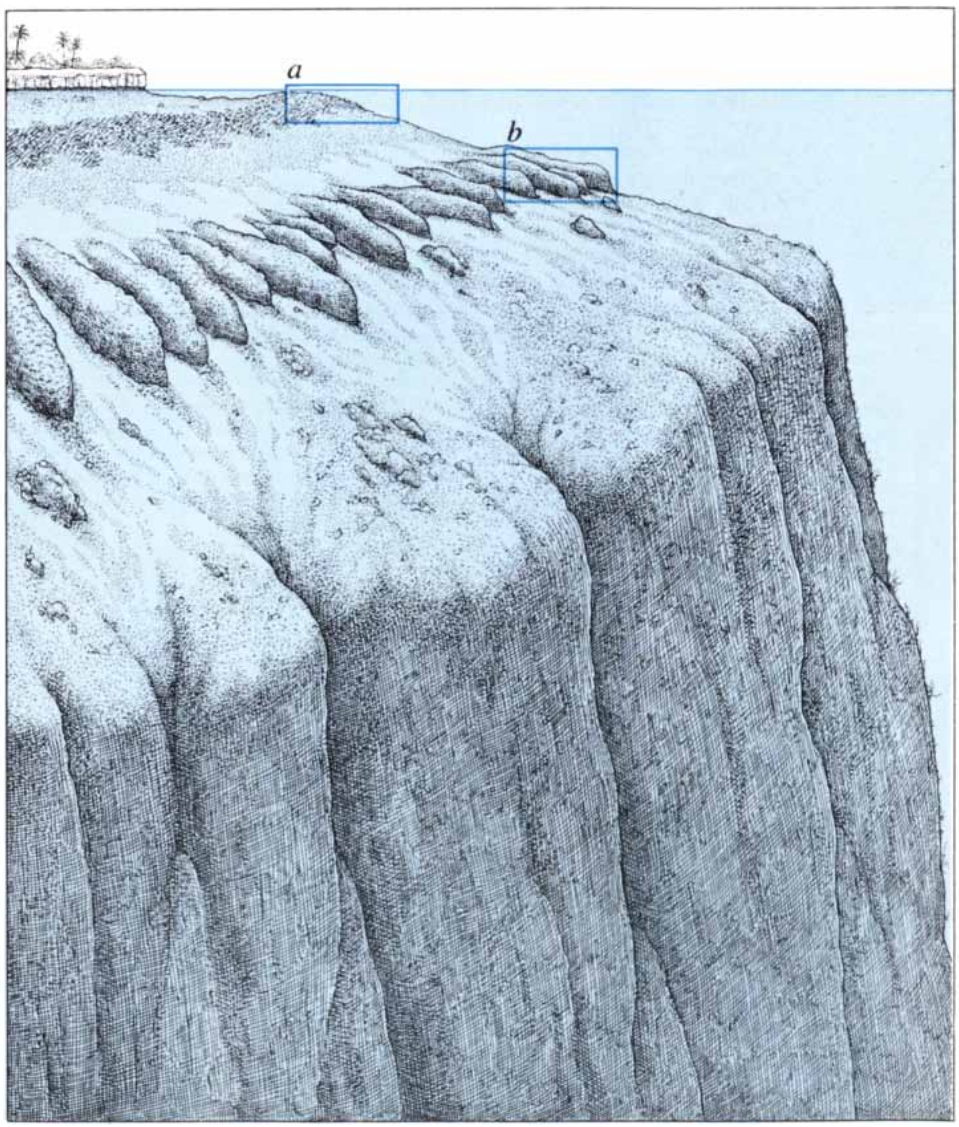
The role of chance in coral settlement is also reflected in the variability of the major species that fill the same structural roles in any reef. In some Jamaican reefs the dominant coral is the branching coral *Montastrea annularis,* but in similar habitats the same role is filled by the different species *Agaricia tenuifolia,* which forms colonies of identical shape, size and orientation. Hence in the creation of diversity in a coral reef historical variation is in many reefs just as significant as the approach to an ecological equilibrium where many specialized organisms coexist.

The many localized habitats and species in the reef give the reef community a wealth of interactions within and among species whose complexity can only be dimly grasped. An intuitive understanding of the major interactions can be gained only after years of field experience. Even then one may focus on so few components of the community that it is easy to miss the significant roles played by many obscure, unexamined or unknown organisms.

The intense competition for food and space in the reef habitat has given rise to a wide variety of survival strategies. For example, Jeremy Jackson and his students at Johns Hopkins University have shown that many encrusting organisms possess specific toxins for defensive or offensive purposes. Corals growing close together compete for space, and some species are able to extrude mesenterial filaments from the gut to kill the polyps of adjacent colonies. Judith Lang, working at Discovery Bay, has shown that among coral species there is a hierarchy of aggression such that slow-growing but aggressive corals can avoid being overgrown by faster-growing but less aggressive ones. This process may lead to an increased diversity of species. In some instances, however, the result is precisely the opposite: James Porter has found that in the reefs on the Pacific coast of Panama the overwhelmingly dominant coral, *Pocillopora damicornis,* is both the fastest-growing and the most aggressive.

Grazing on algal and coral tissues by fishes, sea urchins and other animals has two important effects. Selective grazing may keep a few dominant species of algae from crowding out the more marginal species, so that a diversity of species are able to exist. Experiments in which grazers are excluded from an area of the reef usually result in choking densities of a few dominant algal species, which is rare under ordinary circumstances. Grazers that scrape tissues off hard substrates also create fresh surfaces where new algae can grow and the larvae of sessile organisms can settle. Leslie S. Kaufman of Johns Hopkins has found that some fish species systematically kill patches of coral tissue so that "farms" of algae can grow on the bare coral skeleton. The fishes, which graze on the algae, chase any intruders on their territory, including much larger fishes and even human divers. How much damage to the reef is done by such biological space clearing compared with that done by slumping and storms is not known.

Much also remains to be learned about the nutrient and energy cycles of reefs. The richness of reef biological processes in the face of the poverty of dissolved nutrients in tropical surface waters is evidence that there is an efficient internal cycling of nutrients within the reef ecosystem, but the matter

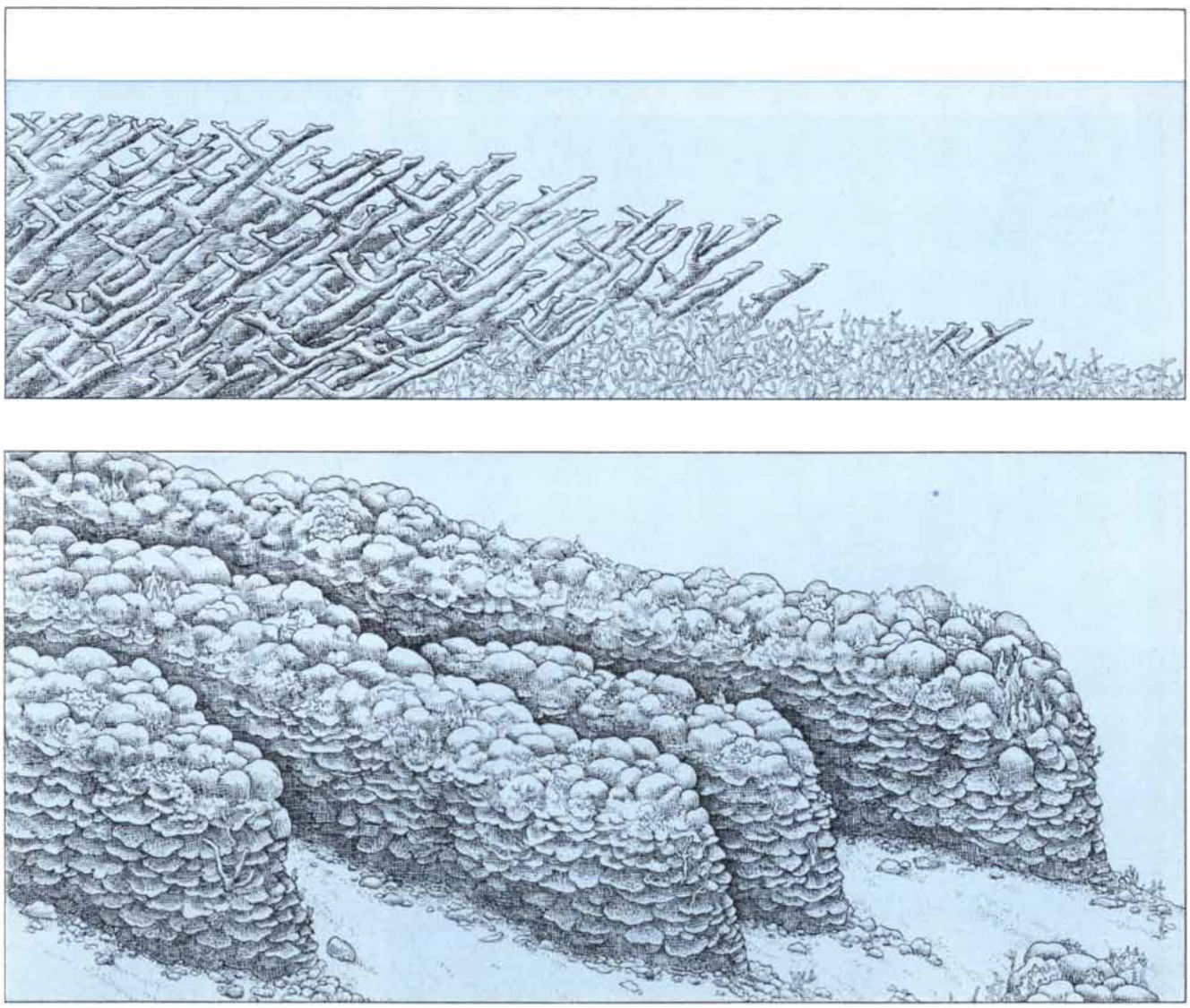


ARCHITECTURE OF THE FRINGING REEF along the northern coast of Jamaica is depicted in these three-dimensional views. Several zones can be distinguished on the basis of reef structure, depth and the associated animal and plant communities. The reef crest extends to a depth of about 15 meters and comprises the shallow coral rampart and the surf zone (*a*). The fore reef extends from 15 meters to 30. This region is a medium-energy environment, with an ambient-light intensity about 25 percent that at the surface. The buttress zone (*b*), where coral

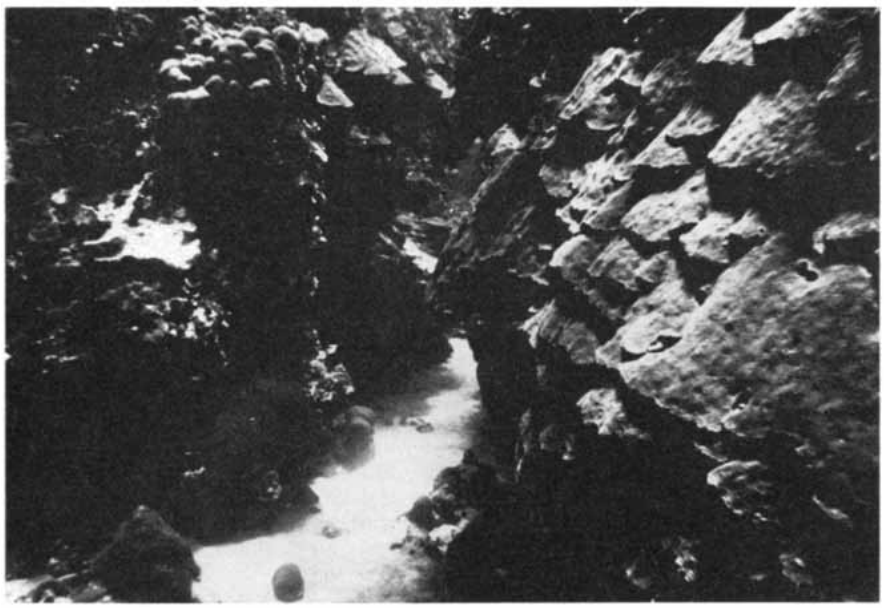
has yet to be investigated in enough detail. The major limiting nutrient in the oceans is generally thought to be nitrogen, and in coral reefs large amounts of the atmospheric nitrogen dissolved in the seawater are fixed in utilizable forms by filamentous blue-green algae. Another source of nitrates is the oxidation of ammonia by bacteria in the course of the decomposition of organic matter in the sediments of the reef lagoon. Recent work indicates that the oxidation of ammonia to nitrate is particularly intense in the fine-grained organic sediments trapped by the roots of sea-grass beds.

The coral reefs of the Atlantic, the Caribbean and the Indo-Pacific do not differ fundamentally in their structural forms, their habitats and the interactions of their species, even though the organisms occupying specific ecological roles vary greatly between oceans and even between individual reefs. Between the Pacific and the Caribbean, however, there is one significant difference: in the Pacific the active growth of coral goes down only to 60 meters, and in the Caribbean it goes down to 100 meters. The reduced range in depth of the Pacific corals may be due in part to periodic infestations by the crown-of-thorns starfish (*Acanthaster planci*), which feeds on coral by turning its stomach inside out, spreading it over the coral and digesting the coral tissues. Before the recent well-publicized outbreak of *Acanthaster* the organism was limited to deeper water and was rarely seen. Then an unexplained population explosion gave rise to a food shortage that forced the starfishes to move up to shallower water, where their destructive effects were readily apparent. The lower limit of reef growth in the Pacific may therefore be affected by periodic starfish grazing. Much remains to be done to prove the hypothesis, however, not least because many Pacific reefs also show signs of being more intensively eroded mechanically than Caribbean reefs.

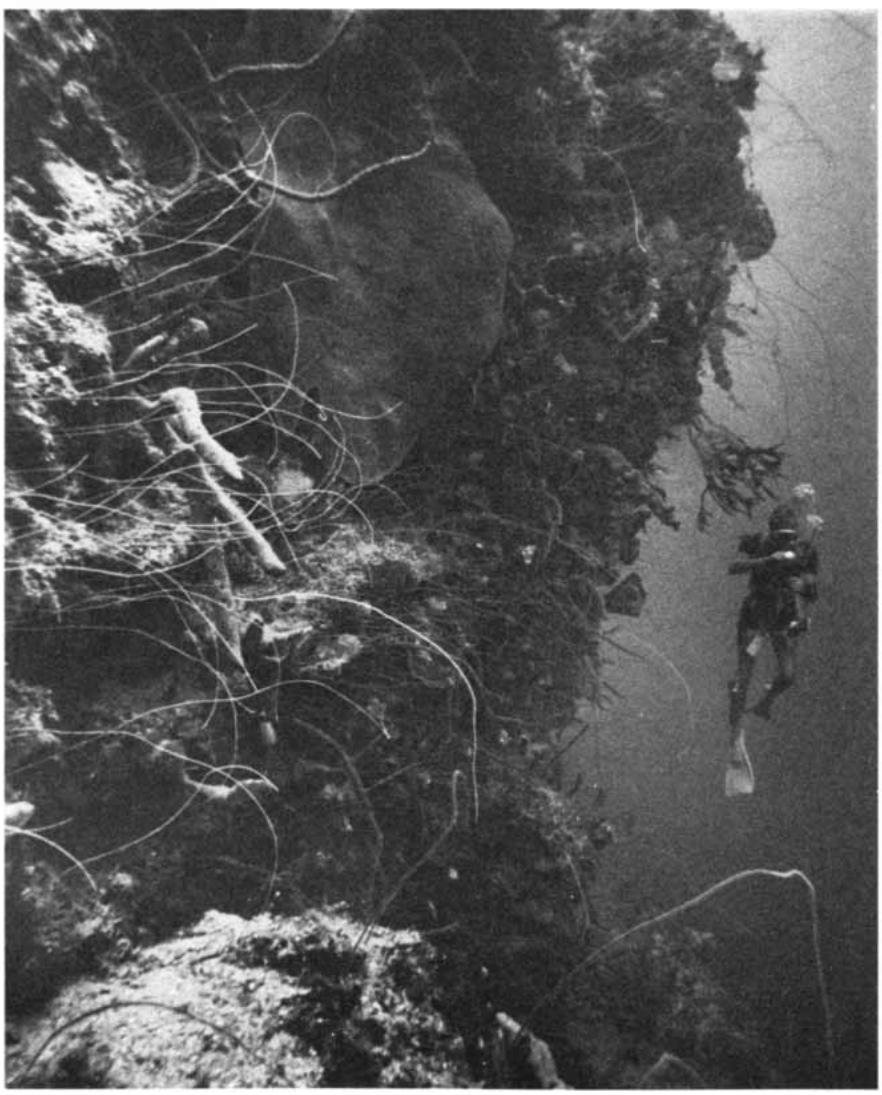
These points illustrate some of the handicaps ecologists face in attempting to predict the stability of reef populations in response to environmental changes or the sensitivity of reef food networks to alterations in the abundance of particular species. Since coral reefs are localized centers of high biological productivity and their colorful fishes are a major source of food in tropical areas, many marine biologists view with alarm the spread of tourist resorts along coral coasts in many parts



buttresses alternate with sandy canyons, serves to dissipate the mechanical energy of the waves and allows the flow down the reef of fine sediments, which would otherwise choke coral growth. The coral colonies are still varied but smaller in size, and much of the available space is occupied by sand-producing calcareous algae, sponges and large gorgonians (sea fans). The deep fore reef extends from 30 meters to 70. This zone has a steep topography and is poorly illuminated, with a light flux about 5 percent that at the surface. At depths below 30 meters coral growth becomes patchy, with a progressive reduction in number of species and size and density of colonies. There is also extensive transport of sediment from the shallow zones above. Beyond the deep fore reef the vertical wall drops off into darkness.



BUTTRESS CANYON between two walls of coral was photographed at a depth of about 12 meters off the northern coast of Jamaica. The wall at the right is covered by colonies of the coral *Monastrea annularia*. Shape of the colonies serves to maximize their light-gathering area.



DROP-OFF WALL of the Jamaican fringing reef is shown at a depth of about 40 meters. The steep fore-reef slope is covered by a dense growth of sponges, gorgonians and whip corals.

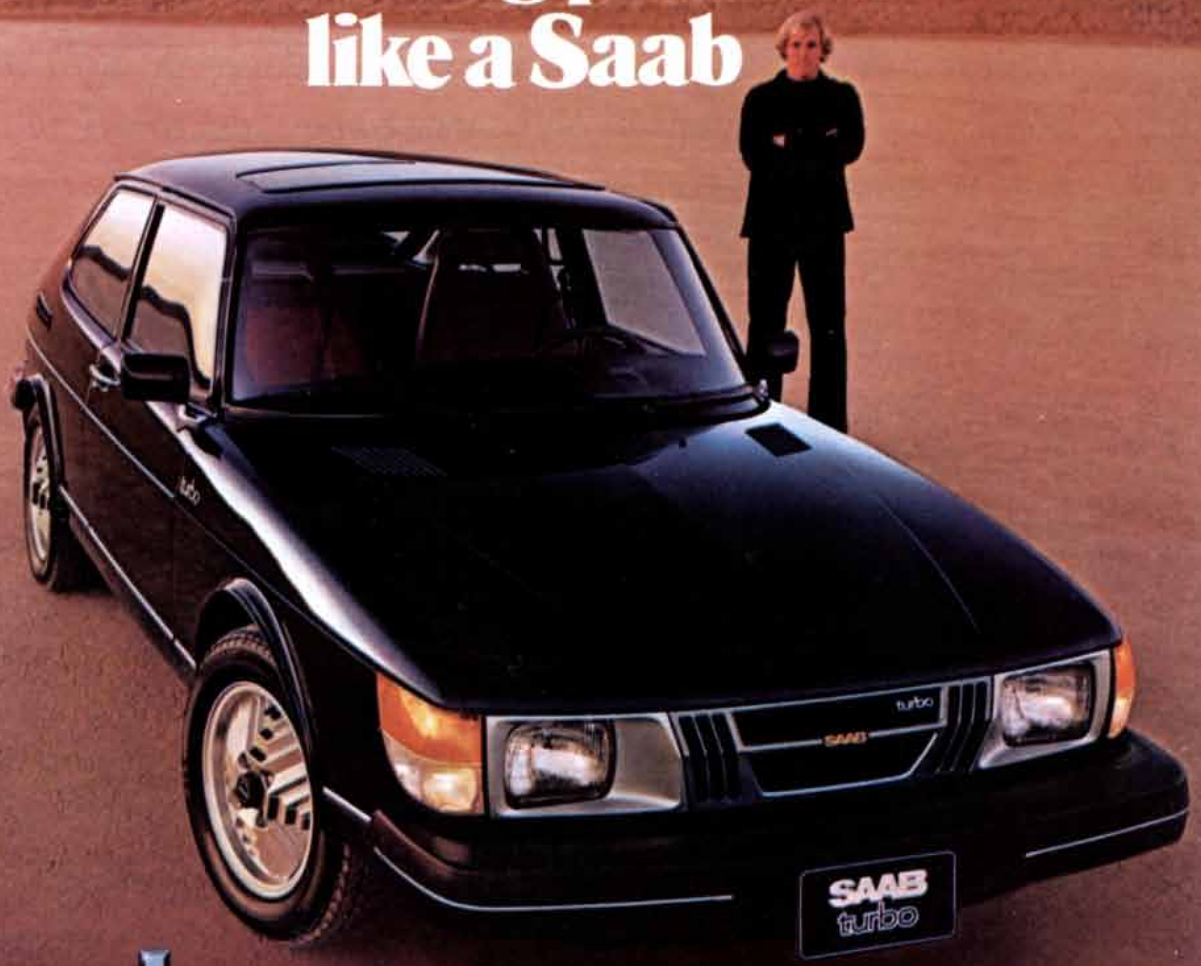
of the world. Such developments are almost always accompanied by increased dumping of sewage, by overfishing, by physical damage to the reef resulting from construction, dredging, dumping and landfills, and by destruction of the reef on a large scale to provide tourists with souvenirs and coffee-table curios. In many areas (such as Bermuda, the U.S. Virgin Islands and Hawaii) development and sewage outfalls have led to extensive eutrophication: the overgrowth and killing of the reef by thick mats of filamentous algae, which in turn support the growth of oxygen-consuming bacteria. The results, which are being intensively studied by Stephen V. Smith and his colleagues at the University of Hawaii, include an increased sensitivity of corals to bacterial diseases, the death of living coral and the resulting erosion of the reef, and the generation of foul-smelling hydrogen sulfide.

Breaching a Barrier

The proposal for digging a new canal at sea level across the Isthmus of Panama arouses further concerns about the viability of coral reefs and their intricately interwoven physical and biological resources. The large range of the tides on the Pacific side of the isthmus and the smaller range of the tides on the Caribbean side, together with the higher mean sea level on the Pacific side, would result in the effective movement through the canal of Pacific marine species into the Caribbean and the Atlantic. Since the reefs of the Caribbean and the Pacific have been evolutionarily isolated for millions of years, such a large-scale incursion of species into new habitats could enable certain species to multiply and spread unchecked, with ecological consequences similar to the explosive multiplication of the English rabbits introduced into Australia. For example, the crown-of-thorns starfish is common on the Pacific side of the isthmus but is not present on the Caribbean side, and its spread through the sea-level canal could decimate the corals of the Caribbean and the Atlantic. In addition poisonous sea snakes, unknown in the Atlantic, are common on the Pacific side of the isthmus. Peter Glynn and Ira Rubinoff of the Smithsonian Tropical Research Institute in Panama have warned that the sea-level canal could cause a greater perturbation in the natural environment than any previous engineering work.

The proposed sea-level canal illustrates not only the concerns of coral-reef biologists but also their ignorance, since it remains difficult to predict the deleterious effects of human activities in an environment as complex as the reef ecosystem. All the same it does not seem unduly alarmist to caution against taking the stability and productivity of the reef community for granted.

Nothing performs like a Saab



turbo

The world is filled with ordinary cars. But there's still a chance to own a thoroughly unique automobile. The new Saab 900 Turbo.

A feel for the road.

At its heart is the Saab turbo engine. Saab engineers took the kind of power that dominates the big race tracks like Indianapolis and LeMans and harnessed it to work at everyday driving speeds. They did this by designing a turbocharging system that puts out more torque at lower engine speeds. So when you have to pass, or merge into fast moving traffic, there's an incredible surge of power at your disposal.

The performance car perfected.

In addition to having extraordinary power, the new Saab Turbo is uniquely equipped to control it. A longer wheelbase, new suspension geometry and front wheel drive give the car cat-like agility – even on rain, snow and ice. Bilstein gas shocks and rack and pinion steering help you cling to the road. Power assisted 4-wheel disc brakes help the low profile Pirelli P6 tires come to smooth, sure stops.

Sports car excitement/Comfort for 5.

The 5 passenger interior is tasteful, restrained and superbly organized. The relationship between the steering wheel, windshield and the hood provide an ideal field of vision.

Upholstery is soft velour, front seats are fully reclining and the driver's seat is heated. A sliding sunroof made of steel is standard.

Engineered in Sweden, this is one of the world's most unique cars. For sheer comfort, excitement and driving confidence, nothing performs like a Saab Turbo.

SAAB 900

Supernovas and Star Formation

The explosive death of a massive star may trigger the birth of other stars. Support for this hypothesis comes from theory and from observations of old supernova remnants and meteorites

by William Herbst and George E. Assousa

A supernova, the catastrophic explosion of a massive star at the end of its life cycle, hurls a shell of gas into the interstellar medium. The explosion simultaneously spews out a tremendous amount of energy: about 10^{51} joules (one joule being the energy needed to lift one kilogram 10 centimeters). Some 25 years ago Ernst J. Öpik of the Armagh Observatory in Northern Ireland suggested that the explosive death of a massive star might trigger the birth of other stars. He argued that the shock wave from a supernova could effectively compress a diffuse cloud of interstellar gas and dust to a density high enough for gravity to pull the material together into a star (or stars).

Until the past half decade observational evidence of supernova-induced star formation was lacking because the objects that had been definitively identified as supernova remnants were simply too young to be associated with the birth of stars. Since 1972 the situation has changed in three ways. First, expanding shells of gas surrounding the remnants of ancient supernovas have been identified. Some of these shells coincide with swarms of young stars. Second, recent developments in molecular-line and infrared astronomy have enabled astrophysicists to observe the early stages of star formation: the compression and heating of the interstellar medium. Third, the discovery of anomalous concentrations of certain isotopes in meteorites indicates that a supernova may have presided at the birth of the solar system.

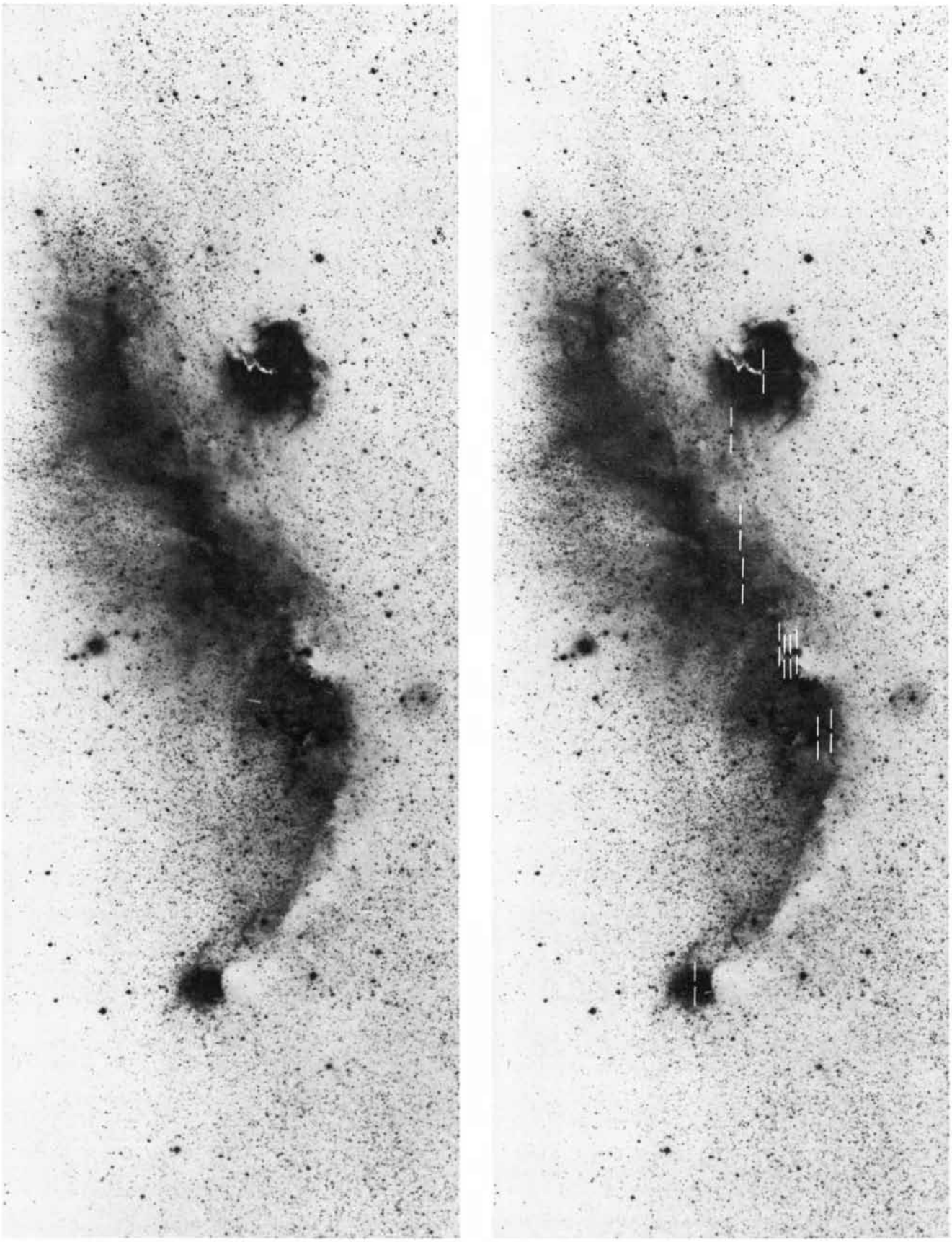
Whether or not a star becomes a supernova depends on its mass. When a cloud has achieved a sufficiently high density, the mutual gravitational attraction of its particles causes it to collapse into a star. As the material collapses the gravitational forces get stronger, causing the star to contract even more. The ever increasing gravitational force ensures that the star will continue to collapse. As the density of the star increases so does the temperature of its interior. At sufficiently high temperatures thermonuclear reactions deep within the star transmute hydrogen into helium. From this point on the course of the star's life cycle depends on the balance between the gravitational forces and the energy released by fusion reactions, with the star's collapse in effect being halted only temporarily by the thermonuclear reactions.

The exact mechanism by which a star becomes a supernova is not yet known and may even differ for different stars. One possible mechanism is the following. After most of the hydrogen in the core is converted into helium a massive star collapses until its internal temperature is high enough to transmute the helium into carbon. When the helium is exhausted, the star may collapse again, and its temperature rises steadily until the carbon starts to be transmuted into heavier elements. At this point either a supernova arises as a result of the explosive carbon burning or the heavier elements may be fused to form still heavier ones. Each thermonuclear reaction yields less and less energy, and so the gravitational forces eventually win out in bringing about the star's collapse.

What happens when the nuclear fuel is exhausted? If the star is less massive than the sun, it becomes a white dwarf: a star with a core consisting of a highly compressed gas of atomic nuclei (mostly helium nuclei) and the electrons stripped from them. If the star is substantially more massive than the sun, the gravitational forces draw the material together into a superdense state where electrons and protons are pushed into each other to form neutrons. The neutrons then are pushed together so that they constitute a nuclear fluid. If such a neutron star is sufficiently massive, even the extremely high density of the tightly packed neutrons is unable to halt the next gravitational collapse. It is thought that as the final collapse begins the neutron star becomes unstable and blows off its outer layers in a supernova explosion. The core of the star may collapse still further until it becomes a black hole, an object whose gravitational field is so strong that neither matter nor light can escape from it.

In our galaxy there is probably one supernova every 50 years or so, although most supernovas are so far away that the interstellar dust between them and the solar system makes them invisible. A supernova has such a large effect on the interstellar medium in its immediate vicinity that the remnants of the explosion can be detected for at least 100,000 years. Observations of supernova remnants and computer simulations of them provide a general picture of a supernova explosion. The ejected matter, which moves initially with a velocity of about 10,000 kilometers per second, collides with the rarefied matter of the interstellar medium and sweeps some of it up into an expanding but decelerating shell. After 1,000 years or so the matter has traveled roughly four light-years and is expanding at a rate of about 1,000 kilometers per second. After 50,000 years the old remnant of the supernova consists of a shell of gas some 150 light-years in diameter expanding at a rate of about 100 kilometers per second. The oldest remnants that can be detected began their outward journey some 100,000 years ago; their diameter is about 200 light-years and their expansion velocity is about 50 kilometers per second.

To be sure, the actual evolution of a particular remnant is governed by many factors, including the density and homogeneity of the ambient interstellar medium and the precise amount of energy released by the supernova. The general picture we have outlined is nonetheless a good one. Remnants that are older than 100,000 years expand with a velocity of at most 50 kilometers per second. Such remnants lose their distinguishing characteristics of filamentary optical emission (electromagnetic radiation emitted by filaments of gas in the remnants) and nonthermal radio emission (with the intensity of the radiation increasing with wavelength). For a time the remnants can still be detected as expanding shells of neutral (un-ionized) hydrogen. Near



ARC-SHAPED CLOUD of dust and fluorescent hydrogen nearly 100 light-years from top to bottom appears in these two negative prints of a photograph of a region in Canis Major made with the 48-inch Schmidt telescope on Palomar Mountain. The arc is the edge of an expanding spherical shell of gas. The energy of expansion suggests that the shell is the remnant of an old supernova. An association of infant stars of intermediate mass called Canis Major R1 lies along the right-hand side of the shell. The shell may have triggered the formation of Canis Major R1 by compressing gas and dust to a sufficiently high density for gravity to pull the material together into stars. The light reflected from the young stars gives rise to nebulas. The positions of some of the young stars are marked on the print at the right.

the central plane of the galaxy, however, hydrogen is already abundant, and so the shells are hard to find because they blend into the interstellar medium. Above and below the galactic plane the background hydrogen is scarcer. Carl E. Heiles of the University of California at Berkeley has shown that much of the neutral hydrogen there is in the form of shells.

It has long been known that supernovas had a seminal role in the origin of the solar system quite apart from the expansion of their shells. The heavier elements in the clouds of dust and gas that came together to form the sun and the planets were ejected by supernovas over a period of several billion years. (The lighter elements in the clouds, namely hydrogen and helium, were largely primordial.) The content of the ejecta probably differed from one supernova to another, but by the time the solar system began to form, the distinctive debris of aeons of supernovas had blended into clouds of homogeneous composition. In particular the isotopic composition (that is, the relative abundance of the nuclear species of the elements) of the clouds was homogeneous. The recent discovery of meteorites with an isotopic composition differing from the solar-system average means there was at least one supernova that exploded so close to the time of the birth of the sun and the planets that its ejecta could not thoroughly blend with the ejecta from earlier supernovas. Perhaps the shock wave from this supernova triggered the formation of the solar system [see "Did a Supernova Trigger the Formation of the Solar System?" by David N. Schramm and Robert N. Clayton; SCIENTIFIC AMERICAN, October, 1978].

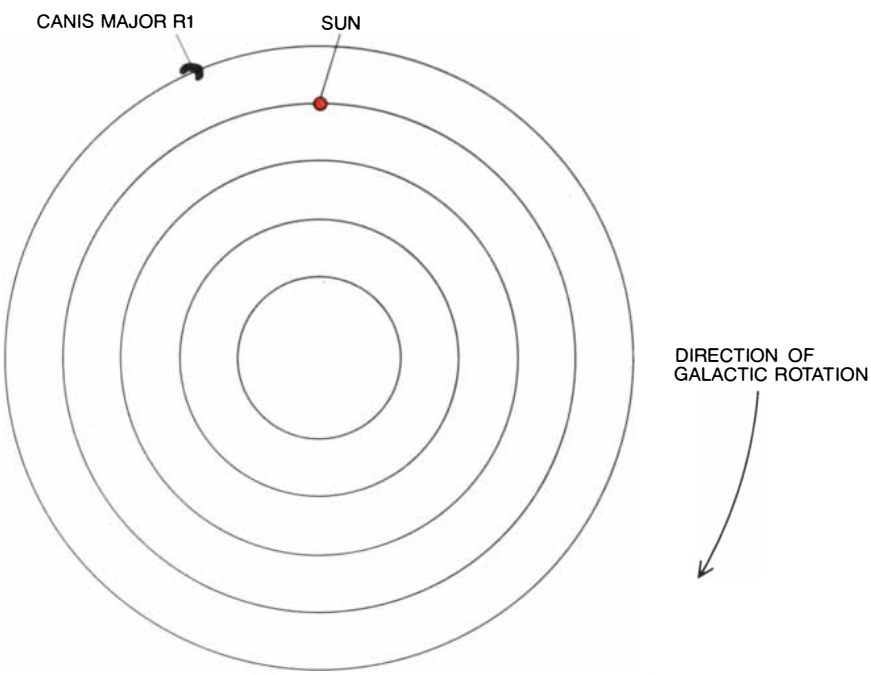
The solar system is only about a third the age of the galaxy, and so it is reasonable to assume that the conditions conducive to star formation at the time the sun formed might still be in effect today. There is no question that in the neighborhood of the sun stars have recently been born and are continuing to be born. For example, the Great Nebula in Orion is an aggregate of massive luminous stars at a distance of about 1,500 light-years. These massive stars are hot bluish-white stars of the spectral types O and B. They have surface temperatures between 16,000 and 45,000 degrees Kelvin, compared with the sun's 6,000 degrees, and they are between 800 and 500,000 times more luminous than the sun. Such aggregates, called OB associations, consist of between five and 50 stars each with a mass of roughly between 10 and 30 times the mass of the sun. Type O and Type B stars emit energy at such a prodigious rate that they cannot survive for more than 10 million years (perhaps a thousandth the age of the galaxy), and so they must have formed relatively recently.

Sprinkled among the stars of the OB association in the Orion nebula are less luminous stars that are between 100,000 and 10 million years old. Such stars, with masses less than three times the mass of the sun, are commoner than the O and B spectral types. Stars with these relatively low masses take longer to form (since the gravitational forces that draw them together are weaker), and they have not yet evolved to the stable hydrogen-burning stage where the synthesis of heavier elements begins. They are called pre-main-sequence stars. Those that exhibit emission lines in their spectra are known as T Tauri stars, after their prototype in the constellation Taurus. Many T associations (aggregates of T Tauri stars) are situated within the boundaries of OB associations, although some are found in regions where massive stars do not exist.

In 1966 Sidney van den Bergh of the University of Toronto discovered a new kind of star aggregate: the R associations, many of which serve as birthplaces for stars of intermediate mass (three to 10 solar masses). He found R associations by examining photographs of the Milky Way taken as part of the Palomar Observatory Sky Survey. He looked for stars that were so close to interstellar dust clouds that the light reflected from the clouds was visible as nebulas. Stars whose mass is less than three solar masses are probably also forming in R associations, but they are too faint to create reflection nebulas, and so they must be identified by other means. Many R associations exist within the boundaries of OB associations.

An R association of considerable interest is situated in the constellation Canis Major only a few degrees from Sirius. The association, Canis Major R1, consists of about 30 stars in reflection nebulas that lie mainly in an arc-shaped dust cloud with an extent of some 100 light-years. Here there is also an OB association, Canis Major OB1, whose stars extend beyond the dust cloud. The distance between the solar system and Canis Major R1 is about 3,700 light-years.

The R association intrigues us for two reasons. First, it includes stars of intermediate mass that still seem to be in the pre-main-sequence phase of their life cycle. According to a study conducted by one of us (Herbst), René Racine of the University of Montreal and John W. Warner of the University of Minnesota, such stars are probably about 300,000



DIFFERENTIAL RATE OF ROTATION OF THE GALAXY is responsible for the sun's moving away from Canis Major R1. The inner regions of the galaxy complete a circuit of the center much faster than the outer regions; hence the sun has an orbital period of 250 million years, whereas Canis Major R1 has a period of 270 million years. The abundant hydrogen between Canis Major R1 and the sun, and hence between Canis Major R1 and the solar system, appears to be receding because of the differential rate of rotation. The frequency at which radiation emitted by the hydrogen (1,420.4057 megahertz) is received on the earth has been Doppler-shifted by the motion of the hydrogen with respect to the receiver. Since the hydrogen is receding, the radiation is shifted to lower frequencies. Near the focus of the arc of Canis Major R1, however, radiation is being emitted that has been Doppler-shifted to higher frequencies. This indicates that the hydrogen there, which constitutes the expanding shell of the remnant of a supernova, is racing toward the earth with velocities of at most 30 kilometers per second.

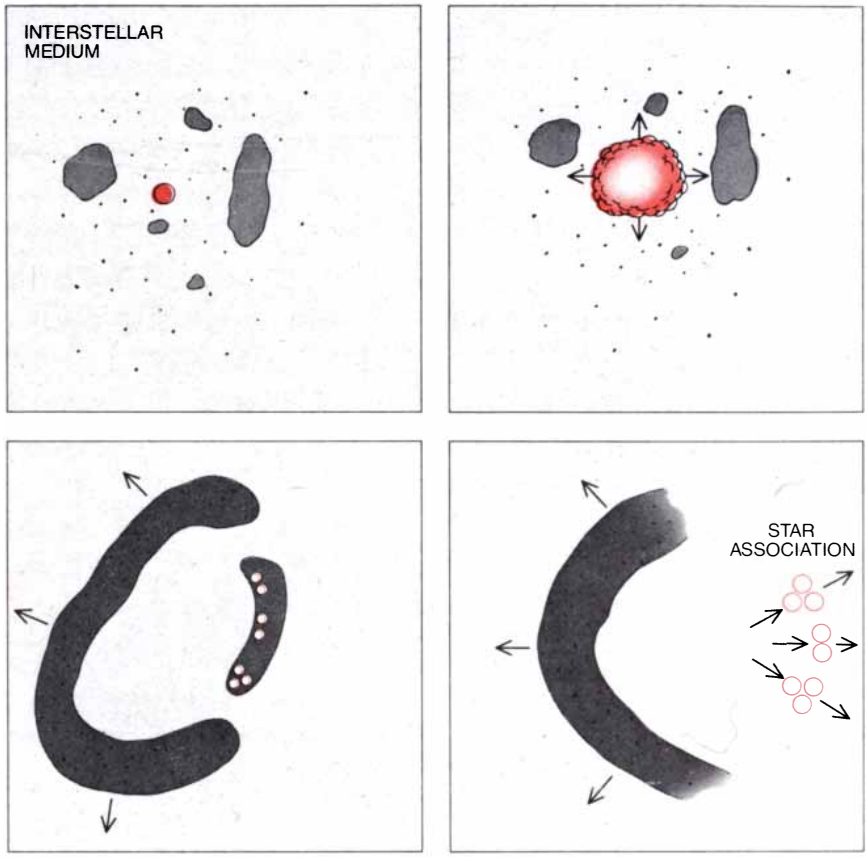
years old, which makes them among the youngest-known stellar objects. Second, the arc-shaped dust cloud in which the stars are forming lies along one edge of a giant expanding shell of gas.

The shell was first identified from 21-centimeter radio waves emitted by the neutral hydrogen in the shell. A hydrogen atom consists of a proton and an electron, which can spin either in the same direction or in opposite directions. If the spins are in the same direction, the electron may flip so that its spin is opposite to the spin of the proton. When the electron flips in this way, the hydrogen atom radiates at a frequency of 1,420.4057 megahertz, which corresponds to a radio wavelength of about 21 centimeters.

Although the amount of hydrogen in the shell is negligible compared with the amount of hydrogen along the line of sight between the solar system and Canis Major R1, it was possible to detect the shell because of its unusual velocity. Along the line of sight hydrogen atoms emit radio waves at the frequency of 1,420.4057 megahertz, but the frequency at which the radiation is received on the earth is Doppler-shifted by the motion of the atoms with respect to the receiver: the emission from hydrogen moving away from the observer has a lower frequency and that from hydrogen moving toward the observer has a higher frequency.

The galaxy is a differentially rotating body: its inner regions complete a circuit of the center much faster than its outer regions. The differential rate of rotation means that the frequency of the emitted radiation is Doppler-shifted according to its position in the galaxy. In the direction of Canis Major there is a fortunate circumstance: along the line of sight all the hydrogen happens to be receding (so that the radiation is shifted to longer wavelengths) except for a small region lying at the focus of the arc of Canis Major R1. There the neutral hydrogen races toward the earth with velocities of up to 30 kilometers per second. Presumably the far side of the shell, which is moving away from the earth, also consists of neutral hydrogen. It is impossible to detect the far side because along the line of sight all the background hydrogen also moves with recessional velocities.

Ronald J. Reynolds and Peter M. Ogden of the University of Wisconsin intensively studied the kinematics of the gas in the vicinity of Canis Major R1. Photographs taken as part of the Palomar Sky Survey had revealed a ring-like distribution of ionized gas adjacent to the R association. With a Fabry-Perot spectrometer Reynolds and Ogden looked at emission lines at visible wavelengths in an ionized region of the gas. Emission lines are the result of electrons in atoms radiating energy as they drop from higher energy orbits to lower ones. In the terrestrial laboratory transitions between certain orbits in certain atoms are "forbidden"; they are extremely improbable because it is much likelier that the atoms will lose excess energy in collisions with neighboring atoms. The density of matter in interstellar space is so low, however, that collisions between atoms are few and far between. Hence atoms sometimes do radiate in forbidden lines.



SUPERNOVA-INDUCED STAR FORMATION begins with a massive star (*color*) embedded in the cloud-laden interstellar medium (*top left*). As the star collapses under the gravitational attraction of its constituent particles, it becomes unstable and blows off its outer layers in a supernova explosion. After about 30,000 years the remnant is 100 light-years across and has swept up a shell of interstellar matter that is expanding with a velocity of at least 100 kilometers per second (*top right*). After 100,000 years the remnant has a diameter of 200 light-years and has slowed down to a speed of at most 50 kilometers per second (*bottom left*). The shock wave from the shell may have compressed clouds of gas and dust to a density high enough for gravity to begin to pull the material together into stars (*colored dots*). After about three million years the association is well developed (*bottom right*). By this time many stars have formed that cluster together in a group, which is surrounded by a large loop of slowly expanding hydrogen. The entire process can repeat itself if one of the infant stars develops into a supernova.

Reynolds and Ogden detected a forbidden nitrogen line toward the center of the ring of ionized gas. The line was split, that is, it consisted of two frequencies because the radiation was Doppler-shifted in two ways. This indicated that the gas at the center of the ring is divided into two regions whose velocities differ by 26 kilometers per second. Near the edge of the ring the line was not split. These observations confirm the presence of an expanding shell of gas adjacent to the R association; the regions of differing velocities constitute the front and back of the shell.

What is the origin of the expanding shell of gas? In our view asking this question amounts to asking what caused the stars to form in Canis Major R1, because it is extremely unlikely that the stars would have been created by chance along a 100-light-year arc of an expanding shell 200 light-years in diameter. And as we have seen, a supernova explosion is certainly energetic enough to produce such a shell.

On the basis of the assumption that the shell of Canis Major R1 is a supernova remnant, we calculated its age from its diameter and expansion velocity. We found the shell to be 800,000 years old. Since the stars in the R association are about 300,000 years old, the stars' age is consistent with our hypothesis that a supernova triggered their birth. The density of matter in the expanding shell

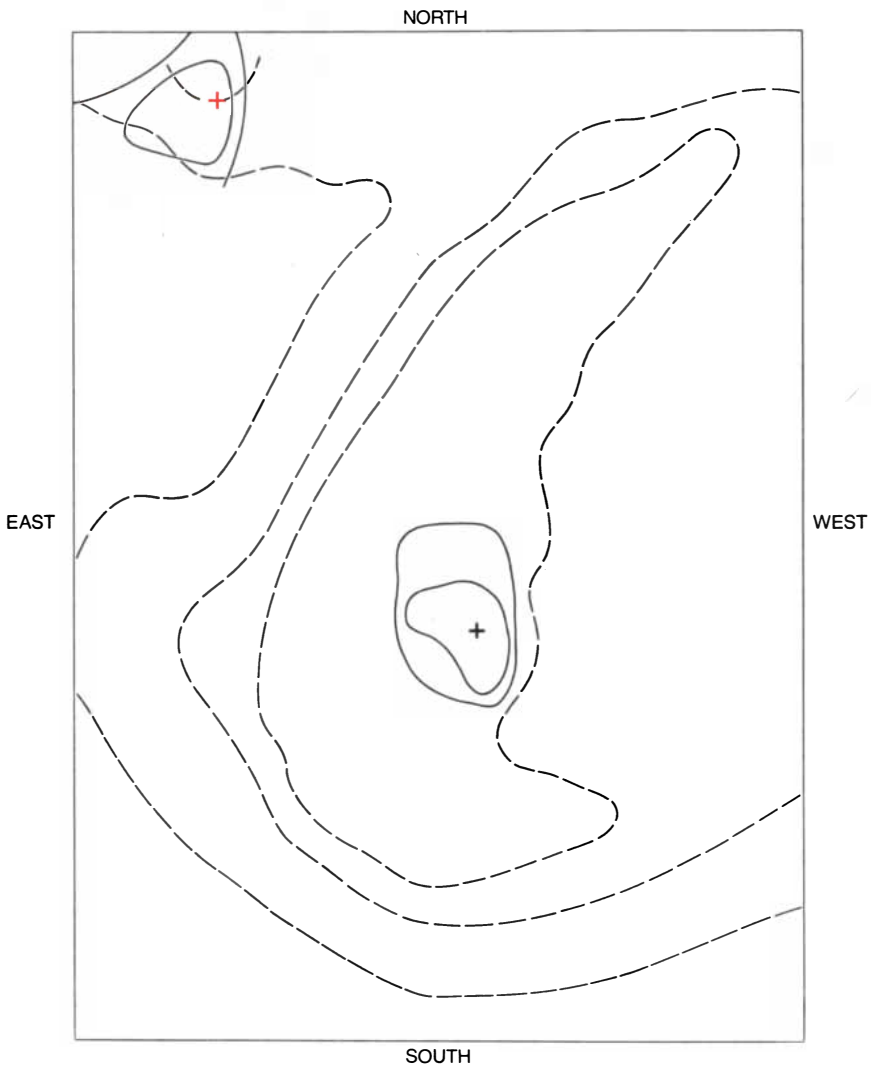
can be estimated from the amount of ionized hydrogen encircling three stars within the shell. From the density, the diameter and the expansion velocity we calculated the amount of energy needed to create the shell: roughly 10^{44} joules, comparable to the energy released by a supernova explosion.

Can any other phenomenon release that much energy? The answer is yes. A massive star could pump 10^{44} joules into a small region of space in the course of its several-million-year life cycle. It is not known, however, whether the ejecta of a massive star could effectively compress the material of the interstellar medium. Theoretical models suggest that in a homogeneous medium Type O stars would give rise to expanding shells that might trigger star formation. The actual interstellar medium, however, is by no means homogeneous, so that the models do not apply. Moreover, there is no luminous massive star near the center of the shell of Canis Major R1 that might plausibly have been serving as the energy source. The two most massive stars in the vicinity of the ring lie along its edge, not near its center. In the best photographs of Canis Major R1 faint and partial ringlike structures surround the two stars, but the structures are insignificant compared with the ring of Canis Major.

One of these two massive stars is particularly interesting because it is a runaway star: a star whose velocity of between 30 and 200 kilometers per second is quite large compared with the velocities of neighboring stars. Adriaan Blaauw of the University of Leiden has developed the most widely accepted explanation of runaway stars. Most stars are born as members of a binary system: two stars revolving about their common center of mass because of their mutual gravitational attraction. The more massive member of a binary system evolves faster and can eventually become a supernova. At that time a large amount of mass is instantaneously expelled from the binary system, and the unevolved star acquires a higher velocity with respect to neighboring stars.

The runaway near Canis Major R1 is racing from the sun with a velocity 30 kilometers per second faster than the velocities of the stars in its vicinity. Like most runaways, this one is hot and luminous and is traveling alone. Only the speed along the line of sight has been determined to date, so that it is not possible yet to tell whether the runaway was once at the center of the expanding shell. Nevertheless, if the speed tangential to the line of sight is comparable to the speed along it, then the star may well have been at the center of the shell some 800,000 years ago, when the shell was formed.

An additional fact relates the runaway to the formation of the shell. In their spectroscopic study Reynolds and Ogden found that an oxygen emission line is enhanced at velocities characteristic of the far side of the shell. This discovery is understandable if the runaway is currently inside or near the far edge of the shell, and that is exactly where it should be if it is the original companion of the supernova that produced the shell and if it acquired its recessional velocity of 30 kilometers per second at the time of the explosion.



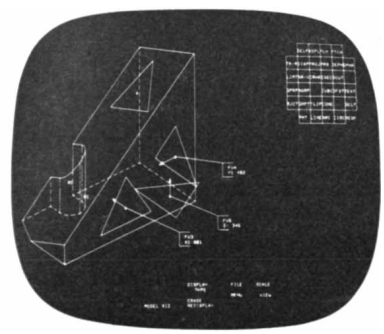
TWO SOURCES OF INFRARED RADIATION (*plus signs*) are embedded in molecular clouds (*solid lines*) associated with the old supernova remnant W44 (*broken lines*). Infrared waves are emitted by envelopes of dust that enshroud stars in their initial stage of evolution. The dust absorbs radiation from the stars, heats up and reradiates the energy at infrared wavelengths. One of the two infrared sources (*colored plus sign*) is definitely associated with W44, whereas it is possible that the other source (*black plus sign*) is only a background object. Both infrared sources were discovered by H. Alwyn Wootten of the University of Texas. The map covers about 30 minutes of arc in the sky, which is equivalent to the diameter of the full moon.

Canis Major R1 is a clear example of stars being formed at the edge of an expanding shell, but it is by no means the only one. Monoceros R1, an R association about 2,500 light-years from the sun, was examined recently by Marc L. Kutner, Robert L. Dickman and their co-workers at the Rensselaer Polytechnic Institute. The stars in Monoceros R1 resemble those in Canis Major R1 and are probably about the same age. Like Canis Major R1, Monoceros R1 consists of infant stars lying along a ringlike distribution of dense clouds of interstellar gas and dust. The diameter of the ring is about 25 light-years, 10 times less than that of Canis Major R1. This difference suggests that the initial density of the ring-forming material was greater in Monoceros R1. There is evidence of expanding neutral hydrogen near the center of this ring too, and the rate and energy of expansion are comparable to those of Canis Major R1. In Monoceros R1 there are no massive stars, and so there are no partial rings of luminous ionized gas like those in Canis Major. All the evidence points to the conclusion that a supernova exploded in Monoceros R1. The only other possible source of energy, an extremely massive star, is absent.

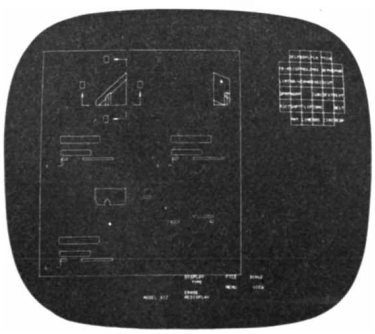
In two clouds at the edge of the Gum Nebula in the sky of the Southern Hemi-

Only from MCAUTO...

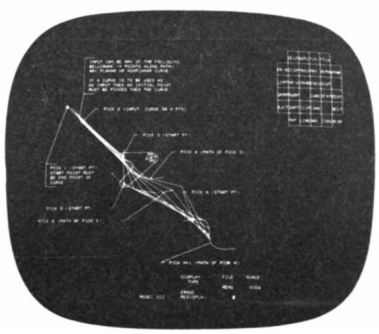
Computer-aided design and drafting with independent view and construct.



True 3-D Design Capability. Interactively create a fully-surfaced three-dimensional geometric model in a real-time environment.



Part Drawing System. Create a two-dimensional picture of either a fully-surfaced or wire-framed part—generating, arranging, and labeling section cuts as you require.



Kinematics Analysis. Analyze and display intermediate and terminal positions of bell cranks, rods and slides; critical positions and intermediate positions are displayed at increments you define.

Introducing CADD (we pronounce it "caddy"), from MCAUTO. A remarkable combination of applied mathematics, computer science, and proven terminal console procedures incorporated into a true three-dimensional, real-time graphics software system. A system which allows you to actually see your product develop—and change it on the spot if you wish—as it is conceived.

With CADD, you are creating, and have available for recall, an easily accessed base of geometrically accurate data which will provide the initial building block in your product development. By computerizing analytical geometry, by leaving all mathematical calculations to the computer, CADD preserves design-engineer time for *designing.*

You can expect better product designs, reduced lead times, improved productivity and job satisfaction, plus reduced design costs. And since CADD part design can be interfaced with analysis and NC programming, you save in the manufacturing process as well.

Send today for all the facts about CADD. Or phone us at (314) 232-8021.

MCAUTO®
MCDONNELL DOUGLAS
Setting the standard in data processing

W. R. Vickroy, MCAUTO, Dept. KI-500
Box 516, St. Louis, MO 63166
☐ Send me information on CADD.
☐ Have your representative phone me for an appointment.

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

sphere, Richard D. Schwartz of the University of Missouri discovered Herbig-Haro objects, semistellar nebulas that are associated in some poorly understood way with the early stages of star formation. He argued that the expansion of the Gum Nebula, which may owe its origin to at least one supernova, has triggered star formation in these clouds. There is circumstantial evidence supporting the hypothesis. Near the Gum Nebula, T. G. Hawarden and P. W. J. L. Brand found several "cometary globules": dust clouds with elongated tails that are compact, fairly opaque, isolated and regular in form. Without exception the elongated tails point away from the center of the Gum Nebula. The nebula NGC 5367 is in a cometary globule in Centaurus that coincides with a cluster of young low-mass stars, according to Howard Van Till and his co-workers at the University of Texas and P. Williams and his co-workers at the University of Edinburgh. Perhaps the cometary globules in the Gum Nebula are also the sites of star formation or could be such sites in the future.

We have limited the discussion so far to young but visible stars and to supernova remnants so old that only indirect evidence indicates they are actually supernova remnants and not something else. Evidence of supernova-induced star formation also comes from the discovery of molecular clouds that are evolving into stars next to known supernova remnants. For example, H. Alwyn Wootten of the University of Texas has mapped a compressed cloud at the edge of an old supernova remnant known as the Monoceros Loop and has found conditions conducive to star formation. In clouds associated with another remnant, W44, he discovered at least one source of infrared radiation and possibly two. Such infrared sources are thought to mark the presence of young stars. The radiation is emitted by the envelopes of dust that shroud stars in the initial stage of their evolution. The dust absorbs radiation from the stars, heats up and reradiates the energy at infrared wavelengths.

It is time to consider how widespread supernova-induced star formation may be. Most stars seem to have formed in associations, which are sometimes composed of separate subgroups of stars, the youngest stars being shrouded in gas and dust. Blaauw has suggested that Orion OB1 consists of four subgroups lined up in order of their age (from the oldest subgroup at one end of the association to the youngest at the other end), so that the association may have been created in successive bursts of star formation.

One model of the burstlike nature of star formation in an OB association does not involve supernovas. This detailed model, developed by Bruce G. Elmegreen and Charles J. Lada of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory, focuses on the copious amount of ultraviolet radiation emitted by infant stars that ionizes hydrogen in the ambient interstellar medium. The region of ionized hydrogen presses on a nearby molecular cloud, creating a shock wave. When the material at the edge of the cloud gets dense enough, it breaks into fragments that gravitationally collapse into the stars of a subgroup. Eventually these stars create a new region of ionized hydrogen, which generates a shock wave that triggers the birth of a new subgroup. The process repeats itself until the entire association is formed. This model does not, of course, explain how the initial subgroup of infant stars formed; a shock wave from another molecular cloud or from a supernova explosion could have provided the necessary pressure.

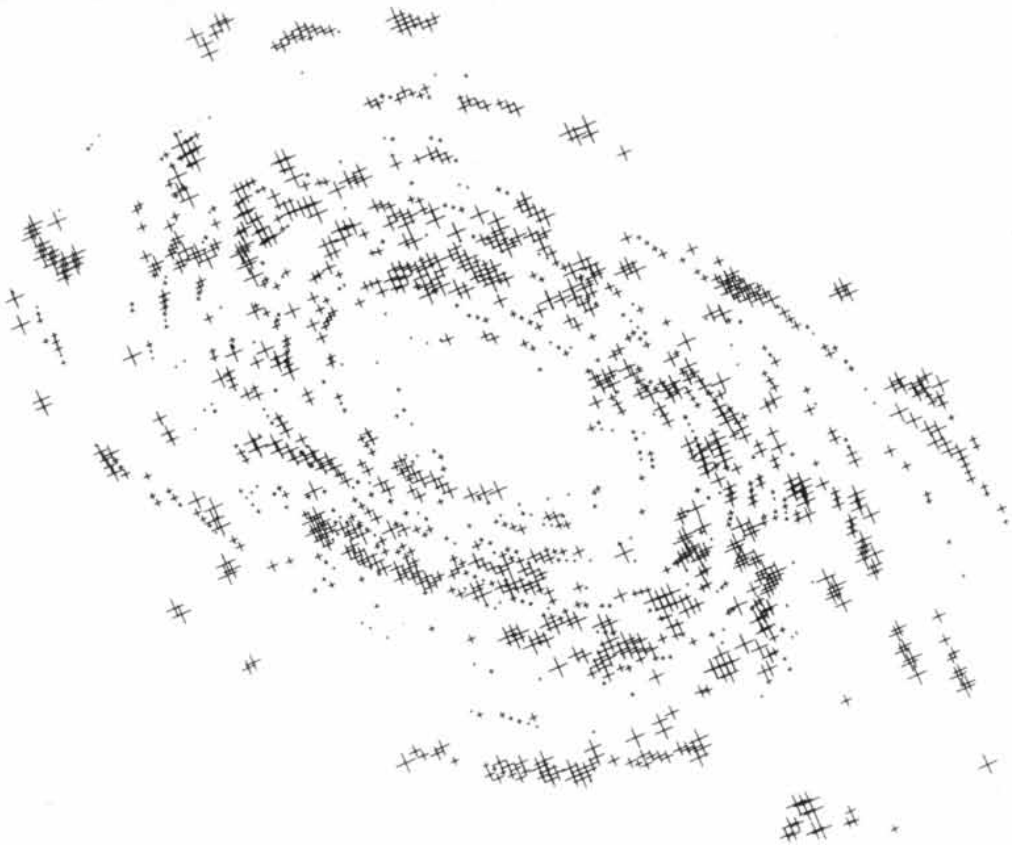
Although supernovas have not yet been incorporated into a detailed model of the burstlike nature of an OB association, they may play a significant role in the phenomenon. If a supernova triggered a burst of star formation in a dense primordial cloud, and if the most massive star to be formed from the cloud evolved into a supernova and triggered a new round of star formation, which in turn gave rise to another supernova and so on, then the observed distribution of stars in Orion OB1 would be accounted for. Moreover, the time between successive bursts would be several million years, identical with the observed age difference between the adjoining subgroups.

Regardless of whether or not supernovas are responsible for the subgroup structure of Orion OB1, there is no question that they are often found within associations. Blaauw was the first to recognize this fact when he proceeded to "trace back" three runaway stars to their origin in Orion OB1. He also noted that in the older subgroups in an association there are fewer extremely massive stars than there are in the younger subgroups. By counting up the number of these "missing" stars one finds that in the past 10 million years about 20 supernovas exploded in Orion OB1. The enormous shell these supernovas collectively have generated was studied by Reynolds and Ogden, who estimate that its energy content is 10^{45} joules. The last supernova in Orion probably goes back about 500,000 years, and it may be the source of a shock front that was observed recently by Lennox L. Cowie and his co-workers at Princeton University.

If a supernova triggered star formation that gave rise to another supernova, then a chain of star-forming regions would result. If many such chains were created in a differentially rotating galaxy, the distribution of stars would resemble the observed distribution in a spiral galaxy. That possibility is the basis of a new model of spiral-galaxy structure proposed by Huberto Gerola and Philip E. Seiden of the Thomas J. Watson Research Center of the International Business Machines Corporation. Computer simulation of self-propagating star formation in differentially rotating disks was able to reproduce the appearance of many spiral galaxies without requiring the presence of an underlying density wave. Such a density wave is the hallmark of the most widely accepted theory of the large-scale structure of spiral galaxies, developed by C. C. Lin of the Massachusetts Institute of Technology and Frank H. Shu of the University of California at Berkeley. The theory maintains that a density wave of spiral form sweeps through the central plane of a galaxy, compressing clouds of gas and dust to five or 10 times their original density. The clouds collapse into stars that form a spiral pattern. Work is now under way to precisely define the differences between the galactic structures predicted by the density-wave theory and those predicted by the supernova model. The predicted structures will then be compared with the observed ones.

It is possible that the universe is diverse enough for both models to be right. Some galaxies have two beautiful arms spiraling out symmetrically from the galactic nucleus. This kind of structure is not easily explained by the supernova model. Other galaxies appear to have no coherent large-scale spiral structure; they consist entirely of bits and pieces of spiral arms distributed in an apparently random way. Such galaxies are not easily explained by the density-wave model. Perhaps for a typical galaxy each of these mechanisms contributes to the spiral structure.

Our own galaxy may be a case in point. It has been known for 25 years that the sun lies within a long, narrow concentration of bright stars that resembles a spiral arm in other galaxies. The concentration is known as the Local Arm. Sections of two other spiral arms have been detected, but the entire structure of the galaxy cannot be mapped at visual wavelengths because the dust in the central plane of the galaxy obscures observations. There is some evidence that the two partially identified arms are density-wave arms, because they seem to exert a gravitational force on neighboring stars and gas. There is no evidence that the Local Arm exerts such a force, according to Lin and his coworkers. The density-wave theory predicts that the sun should lie almost exactly between spiral arms, but actually it lies in a region of active star formation. That can be explained if supernovas are chiefly responsible for star formation in the Local Arm.



SPIRAL GALAXY M81 (*top*) may have been created by supernovas. A theory of spiral-galaxy structure proposed by Huberto Gerola and Philip E. Seiden of the Thomas J. Watson Research Center of the International Business Machines Corporation maintains that if a supernova triggered star formation giving rise to another supernova, then a chain of star-forming regions could be produced. If many such chains were created in a differentially rotating galaxy, the distribution of stars would form a spiral structure. Computer simulation of this theory for the spiral M81 (*bottom*) gives rise to a distribution of bright infant stars (*crosses*) that looks like the actual distribution (*top*).

The Mathematics of Public-Key Cryptography

The search for privacy in an age of electronic communications has given rise to new methods of encryption. These methods are more practical than older ones and are mathematically more interesting

by Martin E. Hellman

The electronic communications systems that are proliferating throughout modern society offer speed, accuracy and ever diminishing cost. They also present serious problems of security. As the ordinary transactions conducted in person, on the telephone or by written correspondence have come increasingly to be conducted by new kinds of electronic systems the susceptibility of organizations and individuals to eavesdropping and forgery has grown dramatically. One way to prevent tampering with the new electronic systems and to protect the vast quantities of private information such as the credit records and medical histories now stored in computer data banks is to resort to cryptosystems: methods for encrypting, or transforming, information so that it is unintelligible and therefore useless to those who are not meant to have access to it.

Encryption is a special form of computation, and almost all modern cryptosystems depend on difficulty of computation for their security; they effect transformations of data so complicated that it is beyond the economic means of an eavesdropper to reverse the process. (Accounts of intelligence operations during World War II reveal that as recently as 35 years ago systems offering this type of security were not widely available. Since then the cost of computation has dropped by a factor of about a million, so that the equipment necessary for secure encryption is now reasonably priced.) Given unlimited computing power (an unrealistic assumption) such computationally secure systems could be broken, but in practice they appear to be unbreakable.

At present mathematicians lack the tools for proving systems to be computationally secure, and the history of cryptography demonstrates all too well that supposedly unbreakable systems often have hidden flaws. It is hoped that discoveries in complexity theory, a branch of mathematics that studies the difficulty (or cost) of computation, will eventually provide the tools needed to establish provably secure cryptosystems: computationally secure systems that can be guaranteed to be free of hidden flaws. In the meantime a group of mathematical problems characterized by a certain kind of computational intractability are serving as the basis of a new class of encryption procedures that are in many ways superior to current techniques. The proposed new systems, which were first put forward by Ralph Merkle, Whitfield Diffie and me at Stanford University, are termed public-key cryptosystems. To understand the significance of the term it is necessary to consider briefly how methods of encryption have developed historically.

A = 00000	B = 00001	C = 00010	D = 00011	E = 00100		
F = 00101	G = 00110	H = 00111	I = 01000	J = 01001		
K = 01010	L = 01011	M = 01100	N = 01101	O = 01110		
P = 01111	Q = 10000	R = 10001	S = 10010	T = 10011		
U = 10100	V = 10101	W = 10110	X = 10111	Y = 11000		
Z = 11001	= 11010	. = 11011	, = 11100	? = 11101		
; = 11110	: = 11111					
a = 00	b = 01	c = 02	d = 03	e = 04	f = 05	g = 06
h = 07	i = 08	j = 09	k = 10	l = 11	m = 12	n = 13
o = 14	p = 15	q = 16	r = 17	s = 18	t = 19	u = 20
v = 21	w = 22	x = 23	y = 24	z = 25	A = 26	B = 27
C = 28	D = 29	E = 30	F = 31	G = 32	H = 33	I = 34
J = 35	K = 36	L = 37	M = 38	N = 39	O = 40	P = 41
Q = 42	R = 43	S = 44	T = 45	U = 46	V = 47	W = 48
X = 49	Y = 50	Z = 51	0 = 52	1 = 53	2 = 54	3 = 55
4 = 56	5 = 57	6 = 58	7 = 59	8 = 60	9 = 61	= 62
. = 63	, = 64	; = 65	? = 66			

CRYPTOGRAPHIC SYSTEM is a mathematical system for encrypting, or transforming, information so that it is unintelligible and therefore useless to those who are not meant to have access to it. The encryption process generally begins with the conversion of the plaintext, or unenciphered message, into a string of numbers by means of a digital "alphabet" such as one of those shown here. In some cryptosystems it is more convenient to work with binary numbers, and so in the rather simple alphabet shown at the top five bits (binary digits) have been allocated to represent each letter, number or punctuation mark in the plaintext. Each bit can take two values (0 or 1), making a total of 2^5, or 32, characters in this alphabet. In other cryptosystems it is simpler to think in terms not of a binary (base-2) number system but a decimal (base-10) one. In alphabet shown at bottom two decimal digits have been allocated for each plaintext symbol, providing total of 10^2, or 100, characters. (Some of these may not be needed.)

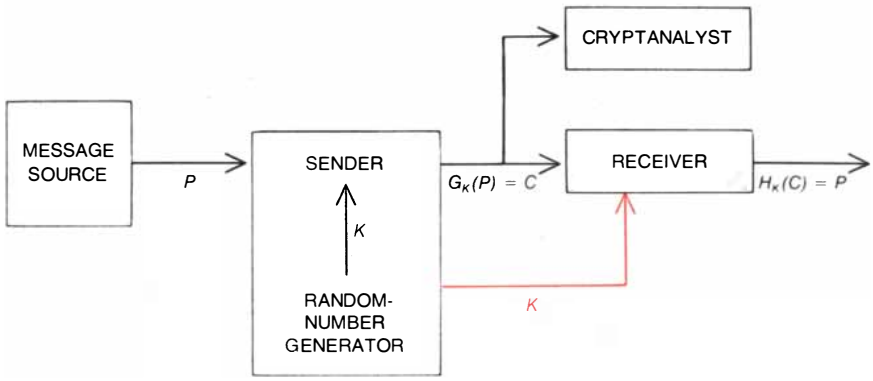
Any cryptographic technique, such as the substitution and transposition of symbols, that operates on a message without regard to its linguistic structure is called a cipher and is said to generate a ciphertext. (Codes, which I shall not discuss here, operate on larger linguistic units such as words or phrases.) More precisely, the basis of any cipher is an invertible function: an operation (performed by the sender of the message) that converts a plaintext, or unenci-

phered message, into a ciphertext and has an inverse operation (performed by the intended receiver of the message) that recovers the plaintext from the ciphertext.

Originally the security of ciphers depended on the secrecy of the entire encryption process, but eventually ciphers were developed for which the algorithm, or sequence of steps, of encryption could be revealed without compromising the security of a particular ciphertext. In such ciphers—the conventional cryptosystems of today—a set of specific parameters, called a key, is supplied along with the plaintext message as an input to the enciphering algorithm and along with the ciphertext message as an input to the deciphering algorithm. In other words, the specific transformations of the plaintext and the ciphertext depend on the key as well as on the enciphering and deciphering algorithms. In fact, the algorithms themselves can be made public, because the security of the ciphertext generated in such a system depends entirely on the secrecy of the key. In the new public-key cryptosystems not only the algorithms but also the key for implementing the enciphering algorithm can be revealed without compromising the security of the ciphertext.

To understand the advantages conferred by the public-key arrangement, consider a conventional cryptosystem employed for protecting information transmitted over an insecure communications channel such as radio. A system of this type can be viewed as a mathematical strongbox with a resettable combination lock. After the sender and the receiver agree on a sequence of numbers—the key—to serve as the combination of the lock, the sender places his message in the box, sets the combination and closes the lock. If the strongbox—the cryptosystem—is secure, no third party who intercepts the box while it is en route to the receiver will be able to get into it to read or alter the message. In other words, a conventional cryptosystem prevents eavesdroppers from extracting information from an insecure channel and prevents forgers from modifying information in the channel.

Until quite recently the principal users of cryptosystems were the military and diplomatic services of the world. The drawbacks of the conventional systems are particularly troubling, however, to the new commercial users of cryptography. To begin with, before any information can be enciphered and transmitted over an insecure channel the receiver and the sender must agree on a key. Since the security of the system depends exclusively on the secrecy of the key, the key must be transmitted by means of a secure channel such as a trusted courier, a system that is slow and costly. The distribution of keys is a particular problem in those instances when the individuals seeking privacy have had no prior communication or when privacy must be maintained over a large network, two situations that are often encountered in commercial dealings. Indeed, the cost and inconvenience of relying on couriers to distribute the amount of key information that is needed for any broad application of cryptography are virtually prohibitive.



IN A CONVENTIONAL CRYPTOSYSTEM someone who wants to send a private message is provided with an algorithm, or general enciphering procedure, *G* and obtains a key *K*, as is shown in this model of the flow of information in such a system. The key, which must be kept secret, is a set of parameters (typically a collection of large random numbers) for implementing the algorithm, which can be made public. In other words, the algorithm and key together specify the actual enciphering transformation G_K. The sender operates on the plaintext *P* with G_K to generate a ciphertext *C*, or $G_K(P)$. The ciphertext can then be transmitted over an insecure communications channel such as a radio channel. Another public algorithm effects the inverse operation G^{-1}, designated *H* for convenience. Knowing *H* and *K*, the receiver operates on the ciphertext *C* with H_K to recover the plaintext *P*, or $H_K(C)$. A cryptanalyst who knows *G* and *H* and intercepts *C* but does not know *K* will not be able to decipher the message. Hence security of such a system lies entirely in secrecy of its keys. As a result keys needed by users of system must be distributed by means of secure channel (*color*) such as trusted courier.

The requirement of key distribution is not the only drawback of the conventional cryptosystems currently in service. They also fail to meet fully the requirements of message authentication. Since a single key is shared between the sender and the receiver, there is nothing to prevent the receiver from sending himself messages that appear to come from the sender. Consider the difficulties such forgeries could cause in electronic mail or electronic banking systems. Conventional cryptosystems, then, cannot offer the same insurance against disputes over what message (if any) was sent that the exchange of signed documents can. The public-key systems, however, provide answers to both the problem of distributing keys and the problem of authentication.

In a public-key cryptosystem the sender and the receiver rather than agreeing on a single key each generate two distinct keys of their own: an enciphering key *E*, which serves to implement the system's enciphering algorithm, and a deciphering key *D*, which serves to implement the system's deciphering algorithm. The keys are related in the sense that they serve to implement inverse operations: operating on a plaintext message first with the transformation specified by *E* and then with the transformation specified by *D* reproduces the message, and in some (but not all) systems applying the transformations in the reverse order also reproduces the message. The trick is that it is computationally infeasible to derive *D* from *E*: the calculation would require a vast amount of computing time, perhaps thousands or even billions of years on the most powerful computer. Hence each user can publish his enciphering key in a public file such as a telephone book without compromising his deciphering key, which is kept secret. As in a conventional cryptosystem, the general procedures for enciphering and deciphering are public information. Therefore anyone who wants to transmit information to a particular person simply enciphers the information with that person's listed key *E* and sends the ciphertext over an insecure channel. Only the intended receiver, who knows the corresponding secret key *D*, will be able to decipher the transmitted message.

To return to the strongbox analogy, a public-key system provides a strongbox with a new kind of lock, which has two combinations: one to lock the box and one to unlock it. (The box does not lock automatically when it is closed.) The locking combinations of all such strongboxes are made public, so that anyone can lock information in a particular strongbox, but only the individual who owns the strongbox and has set the two combinations will be able to get the information out. With this kind of system there is obviously no need of a secure channel for the distribution of keys. Moreover, some of the public-key systems allow for the construction of a "digital signature" that prevents the forgery of messages by a receiver as well as

by a third party. In other words, these systems make it possible to dispense with the transporting of signed documents and to depend exclusively on the electronic transmission of information.

If an eavesdropper had unlimited computing resources, he could break a public-key system and recover a plaintext. The enciphering operation E is public and the number of possible plaintexts is immense but finite, and so E could be applied to each plaintext until the intercepted ciphertext was reproduced. Since such an attack requires an impossibly large amount of computing time, however, the public-key systems can still be computationally secure. There are also similar techniques for deriving the secret deciphering key D from the public enciphering key E, but once again the computational infeasibility of implementing those algorithms provides the systems with practical security. To put it another way, the systems are based on what are called trapdoor one-way functions. A one-way function is an easily computed function for which it is computationally infeasible to compute the inverse function. A trapdoor one-way function is an easily computed function for which it is computationally infeasible to compute the inverse function unless certain specific information that was employed in the design of the function is known. Hence like a trapdoor in the floor of a motion-picture haunted house, such functions are easy to go through in one direction, but unless one possesses the special trapdoor information (analogous in the haunted house to which brick to pull or which panel to push) the reverse process takes an impossibly long time.

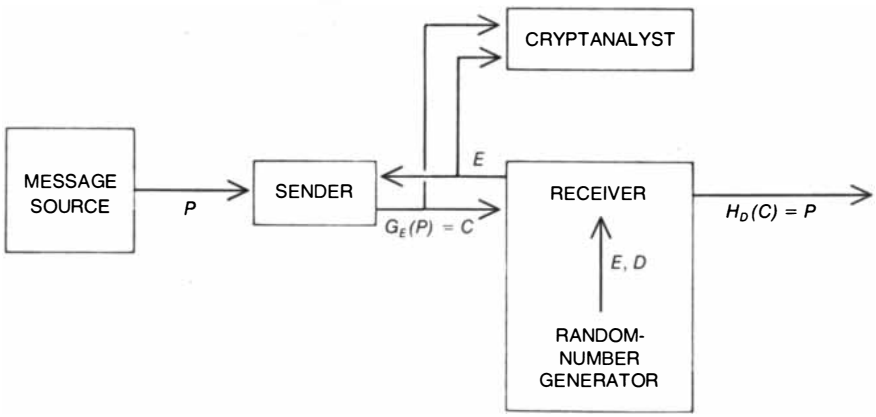
The search for trapdoor one-way functions on which to base public-key cryptosystems led naturally to the class of problems that complexity theory has identified as nondeterministic, polynomial-time problems, or NP problems. For the purposes of these cryptosystems the most important property of the NP problems is that at present all the algorithms that are known for finding general solutions to them call for rapidly increasing amounts of time, although a proposed solution can be quickly checked. In other words, as the size n of such a problem increases, the number of computational steps required to solve the problem increases in proportion to, say, an exponential function of n such as 2^n, whereas the number of steps required to check a possible solution increases in proportion to a polynomial function of n such as n^2. Exponential functions increase far more rapidly than polynomial ones, so that a method of solution that requires exponentially increasing amounts of computer time is impossible to implement for even moderate-size problems. For mathematicians concerned with cryptography the appeal of the NP problems resides in the fact that although it might take someone billions of years to find a solution to such a problem, once he found it he could convince the rest of the world of its validity in seconds. As a result these problems lend themselves readily to the construction of one-way functions. And for the NP problems on which public-key cryptosystems have been based it has been possible to build trapdoors into the functions as well.

I shall describe here two public-key cryptosystems based on NP problems: the trapdoor knapsack system, developed by Merkle and me, and the RSA system, developed by Ronald Rivest, Adi Shamir and Leonard Adleman at the Massachusetts Institute of Technology. The first of these cryptosystems is based on a well-known NP problem called the knapsack or subset sum problem: Given a knapsack of length C and a set of n rods all of the same diameter as the knapsack but of lengths $a_1, a_2, \ldots, a_n$, find a subset of the rods that completely fills the knapsack. To put it another way, given a set of numbers $a_1, \ldots, a_n$ and a sum C, determine which of the numbers add up to C.

The public-key cryptosystem based on this problem operates as follows. The sender begins by converting his message into a string of binary numbers. For example, five bits (binary digits) might be allocated for each letter, number or punctuation mark in the plaintext, providing an alphabet of 2^5, or 32, characters: A = 00000, B = 00001, C = 00010 and so on. Once the message is in binary form the sender consults a public directory of enciphering keys, which lists an ordered set of n numbers $a = (a_1, a_2, \ldots, a_n)$ for each user of the system. This set is called the user's trapdoor knapsack vector.

In mathematics an ordered set of n numbers is called an n-dimensional vector, and the "dot," or scalar, product of any two vectors of the same dimension is defined as follows: for vectors $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ the dot product $a \cdot b$ equals $a_1b_1 + a_2b_2 + \ldots + a_nb_n$. This form of vector multiplication is the basic operation of the enciphering algorithm in the trapdoor knapsack system. To encipher the string of binary numbers that represents his message the sender first breaks the string into blocks of n bits, and for each block $x = (x_1, x_2, \ldots, x_n)$ he forms the dot product $C = a \cdot x$ of that block with the public enciphering vector a, that is, $C = a_1x_1 + a_2x_2 + \ldots + a_nx_n$.

The sum C is the information the sender transmits over the insecure channel, so that any eavesdropper is confronted with the task of recovering x from C and the numbers $a_1, \ldots, a_n$. In what follows it will be convenient to refer to the elements of a vector x as the x_i's (or to the elements of a vector a as the a_i's), where the values of i are taken to be the integers from 1 to n. Since each x_i is equal to either 0 or 1, one can see that the problem of recovering x from C is equivalent to solving the knapsack or subset sum problem for these values of



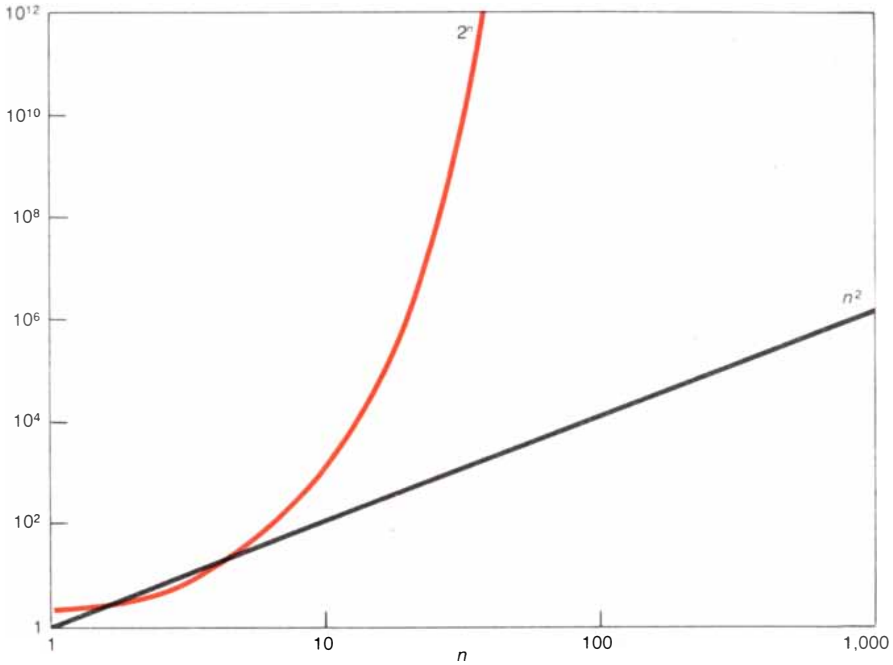
IN A PUBLIC-KEY CRYPTOSYSTEM there is no need of a secure channel for the distribution of keys. As is shown here, each receiver generates two distinct keys: a public key E for implementing the public enciphering procedure G and a secret key D for implementing the public deciphering procedure H, which is the inverse of G. The keys E and D are related in the sense that they serve to specify inverse transformations G_E and H_D, but given E it is computationally infeasible to derive D: computing D from E would require thousands or even billions of years on the largest computer imaginable. Hence the receiver may communicate his enciphering key E over an insecure channel, as is shown here, or even list it in a public directory without compromising his deciphering transformation. A person who wants to send the plaintext P to the receiver operates on it with the receiver's enciphering transformation G_E to generate a ciphertext C, or $G_E(P)$. This ciphertext is transmitted over an insecure channel, and the receiver operates on it with the deciphering transformation H_D to recover the plaintext P, or $H_D(C)$. As long as the deciphering key D is kept secret there is no way for an eavesdropper to decipher the transmitted message. The challenge in designing such a system is to find general procedures G and H for which pairs of inverse keys E and D are easily generated but for which it is computationally infeasible to compute D from E. A source of such pairs is a group of mathematical problems that are said to be in the class NP (*see illustration on opposite page*).

C and the a_i's. The receiver must solve the same knapsack problem, but to simplify the task he has additional information: his secret trapdoor parameters and deciphering key.

These steps should be made clear by a simple example [*see illustration on page 155*]. Consider a plaintext message in which the first word is HOW. In binary form the message begins 0011101-1101011011010. (This binary string, in which the last five-bit block represents the space between HOW and the next word in the message, is generated by the five-bit binary alphabet described above.) Now assume that the intended receiver's public enciphering key is a = (2,292, 1,089, 211, 1,625, 1,283, 599, 759, 315, 2,597, 2,463), or a_1 = 2,292, a_2 = 1,089 and so on. Here n equals 10, and the first block of information, which consists of the first n bits in the binary plaintext, is x = (0, 0, 1, 1, 1, 0, 1, 1, 1, 0). It is enciphered, then, as $C = a_1x_1 + \ldots + a_nx_n$, or C = (2,292 × 0) + (1,089 × 0) + (211 × 1) + (1,625 × 1) + (1,283 × 1) + (599 × 0) + (759 × 1) + (315 × 1) + (2,597 × 1) + (2,463 × 0). Therefore C equals 6,790, and to decipher the message it is necessary to determine which of the a_i's add up to 6,790. (If a_i is included in the sum, x_i is 1 and vice versa.)

None of the known methods for solving the knapsack problem is substantially less time-consuming than conducting an exhaustive search, that is, adding up all the 2^n possible subsets of the a_i's to see which subset yields C. In the example given above, where the number of elements n is equal to 10, this might be considered a workable approach. Someone intercepting C could try all the 2^{10}, or 1,024, possible combinations of the publicly listed a_i's and thereby recover the vector x. In this instance the number of elements in a is too small to provide real secrecy. The knapsack problem is an NP one, however, and therefore the computational difficulty of all known solution methods rapidly "blows up." When the number of elements n is, say, 1,000, the number of possible subsets $2^{1,000}$ is greater than the number of atoms in the known universe. Deciphering by checking $2^{1,000}$ different subsets is quite impossible, and so C effectively shields the secret information x. On the other hand, enciphering 1,000 bits of information in this system is quite efficient, requiring no more than 1,000 additions.

So far I have described what appears to be a one-way function: apparently no one, including the receiver, will be able to recover x. If the elements of vector a are chosen at random, this is exactly the type of system that results. Even in this simple example, however, a trapdoor has been built in. The vector a has been structured in such a way that with a small amount of additional information

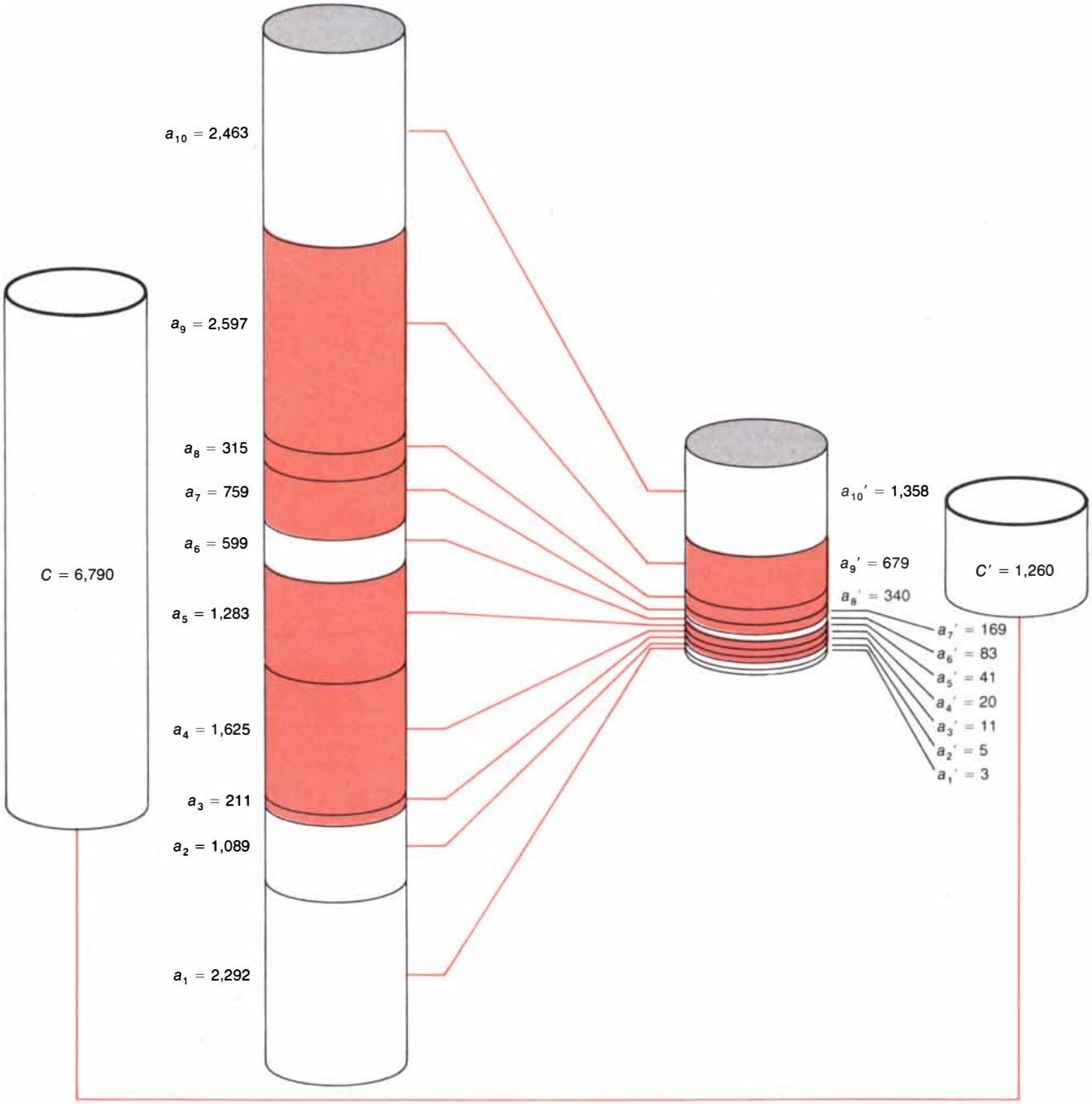


PROBLEMS IN THE CLASS NP (which stands for nondeterministic, polynomial time) are characterized by the fact that although it is easy to check a nondeterministic, or guessed, solution, it is hard to find a correct solution: As the size n of an NP problem increases, the number of computational steps and hence the time required to check a solution increase in proportion to a polynomial function of n such as n^2 (*black curve*), but all known methods of finding a solution increase in proportion to a more rapidly growing function of n, typically an exponential one such as 2^n (*colored curve*). Exponential functions increase far more rapidly, and when n is sufficiently large, NP problems become computationally infeasible. Hence they lend themselves readily to the design of one-way functions: easily computed functions whose inverses are infeasible to compute. In some cases such problems can be developed into trapdoor one-way functions: easily computed functions whose inverses are infeasible to compute unless certain facts employed in design of functions are known. Trapdoor one-way functions serve as basis of public-key cryptosystems: public key specifies easily computed function, which is infeasible to invert unless one knows secret key; that key specifies easily computed inverse function.

x can be derived from C much more rapidly than by an exhaustive search. As I have noted, not all NP problems lend themselves to the insertion of such a trapdoor. Here the trapdoor can be devised because there are certain vectors for which the knapsack problem is not difficult to solve. The receiver takes one of those special vectors a' and disguises it, publishing the resulting ordinary-looking vector a in the public file of enciphering keys. The trapdoor information enables him to move back and forth between a difficult knapsack problem involving a and the easy but equivalent knapsack problem involving a'.

To be more precise, in generating his public vector a the receiver begins by choosing a vector $a' = (a_1', \ldots, a_n')$ in which each element a_i' is larger than the sum of the preceding elements $a_1' + a_2' + \ldots + a_{i-1}'$. For example, if a' equals (3, 5, 11, 20, 41, 83, 169, 340, 679, 1,358), then a_2', which equals 5, is greater than a_1', which equals 3; a_3', which equals 11, is greater than $a_1' + a_2'$, which equals 3 + 5, or 8, and so on. Now, consider a ciphertext C' = 1,260 that was generated with this special vector a'. In other words, C' equals $a' \cdot x'$ for some binary vector $x' = (x_1', \ldots, x_n')$, that is, 1,260 equals $3x_1' + 5x_2' + 11x_3' + 20x_4' + 41x_5' + 83x_6' + 169x_7' + 340x_8' + 679x_9' + 1{,}358x_{10}'$.

Once again the problem of decipherment is equivalent to solving a knapsack problem, but in this instance because of the special property of the vector a' the solution x' is easily determined. To begin with, a_{10}', which equals 1,358, is greater than C', which equals 1,260, and so obviously cannot be part of the subset sum, that is, the rod is too long to fit into the knapsack. Hence x_{10}' must be 0. The next-largest element in the vector is a_9', or 679, which is less than C', or 1,260. As the special property of a' dictates, the sum of the eight remaining elements of a' must be less than 679, and so those elements alone cannot "fill" the knapsack of length 1,260. Therefore 679 must be part of the sum, and x_9' must be 1. Since x_9' equals 1 and x_{10}' equals 0, the equation $C' = a' \cdot x'$ can now be rewritten as $1{,}260 = 3x_1' + 5x_2' + 11x_3' + 20x_4' + 41x_5' + 83x_6' + 169x_7' + 340x_8' + 679 + 0$. Subtracting 679 from both sides of the equation reduces the problem to determining which of the elements $a_1', \ldots, a_8'$ add up to 1,260 − 679, or 581 (the length of the still empty part of the knapsack). Since a_8'



KNAPSACK PROBLEM is an NP problem from which a trapdoor one-way function can be derived. The cylinder and set of rods shown at the left illustrate the classic knapsack problem: Given a knapsack, or cylinder, of length C and a set of n rods all of the same diameter as the knapsack but of lengths $a_1, a_2, \ldots, a_n$, find a subset of rods that fills the knapsack completely. This problem is in the class NP because the best method known for solving it is not much more efficient than trying all 2^n possible subset sums to see which one equals C, and yet a guessed solution can be checked with no more than n additions. Even in the small 10-rod examples shown here, finding a solution (*color*) by this method requires the testing of 2^{10}, or 1,024, different subsets, and when n is, say, 100, the task becomes impossible. An ordered set of numbers such as $a = (a_1, \ldots, a_n)$ or $x = (x_1, \ldots, x_n)$ is called a vector, and the "dot" product of two vectors $a \cdot x$ is defined as the sum $a_1x_1 + \ldots + a_nx_n$. Given a fixed vector a, a function of the variable vector x can be defined as the dot product of x with the vector a, that is, $f(x) = a \cdot x$. If the elements $x_1, \ldots, x_n$ of x are all equal to 0 or 1, then inverting this function, or determining which value of x gives a particular sum $C = a \cdot x$, is equivalent to solving the knapsack problem for C and the given values of $a_1, \ldots, a_n$. The function is one-way because the knapsack problem is in the class NP. Moreover, a trapdoor can be built into the function, because for certain vectors, or sets of rods, $a' = (a_1', \ldots, a_n')$ the knapsack problem is easy to solve. In these sets, such as the one shown in the problem at the right, each element is greater than the sum of the preceding elements. To determine which subset fills the knapsack begin with the last, or largest, element a_n'. In this case a_{10}' equals 1,358, which is greater than 1,260, the length of the cylinder C'. Hence a_{10}' is not in the subset (that is, in the sum $C' = a_1'x_1 + \ldots + a_{10}'x_{10}$, x_{10} equals 0). But a_9', which equals 679, is smaller than 1,260, and since the remaining elements in the set add up to a number even smaller than a_9', it must be in the subset (that is, x_9 equals 1). The problem is now reduced to filling the remainder of the cylinder, whose length is $C' - a_9'$, or 581, with a subset of the remaining rods $a_1', \ldots, a_8'$, and so on. Continuing in the same way, the problem can be solved (or the function based on it can be inverted) with no more than 10 comparisons and 10 subtractions. As colored lines indicate, there is a way to move back and forth between the easy and hard knapsack problems. Parameters for effecting that transformation are secret trapdoor information for trapdoor one-way function based on knapsack problem (*see illustration on page 155*).

equals 340, which is less than 581, it is included in the sum. Thus x_8' is 1. Continuing in this manner, it can be determined that the x' is the original message block $x = (0, 0, 1, 1, 1, 0, 1, 1, 1, 0)$.

Constructing an easy knapsack vector such as a' is not difficult, but how does the receiver get from a' to a and back again? To accomplish that feat he chooses two large random numbers w and m and generates the vector a according to the equation $a_i = a_i'w$ modulo m, for each i from 1 to n. The expression "modulo m" indicates that a_i should be taken to be the remainder left when $a_i'w$ is divided by m. For example, if w equals 764 and m equals 2,731, consider the element a_4' of vector a'. Since a_4' equals 20, $a_4'w$ equals 15,280. Dividing 2,731 into 15,280 gives 5 with a remainder of 1,625, so that a_4, or $a_4'w$ modulo m, equals 1,625.

Modular arithmetic plays a large part in public-key cryptosystems, because it turns smooth, or continuous and continually increasing or decreasing, functions into discontinuous ones, introducing a large factor of confusion into the calculation of their inverses: the values of x that correspond to particular values of $f(x)$. Consider the simple function $f(x) = 4x$. As x increases, the value of $f(x)$ increases in a very orderly way; for example, $f(3)$ is 12, $f(4)$ is 16, $f(5)$ is 20, $f(6)$ is 24 and so on. As a result if one is given a specific number $y = f(x)$, it is not difficult to determine x by a process of guesswork and elimination without ever actually solving the equation $y = 4x$. In other words, if a function is smooth, then no matter how hard solving explicitly for x is, it may still be possible to determine the value of x for a particular $f(x)$ through trial and error. For example, if $f(x)$ equals 20, one might guess that x equals 3. Then $f(x)$ would equal 12, which is too small, so that the correct value of x must be greater than 3. If, however, $x = 6$ were tried, $f(x)$ would equal 24, which is too large, so that x must be less than 6, and so on. Such smooth functions present problems in public-key systems, which depend on functions for shielding numbers.

Consider what happens, however, if modularity is added. When $f(x)$ equals, say, $4x$ modulo 7, as x increases, the value of $f(x)$ jumps around in a quite haphazard way. For example, $f(1)$ is 4, $f(2)$ is 1, $f(3)$ is 5, $f(4)$ is 2, $f(5)$ is 6 and $f(6)$ is 3. Even in such a simple case it is clear that this function that includes modularity provides far better protection for the values of x than the one that does not include it. In the case of the trapdoor knapsack system, applying modularity in the generation of the difficult knapsack vector a prevents the recovery of a' for anyone who does not know the secret transformation parameters w and m.

For anyone who does know w and m,

however, the conversion back into a' would not be difficult at all. In fact, with those parameters it is quite easy to convert the difficult knapsack problem involving the vector a and the transmitted ciphertext message C into an easy knapsack problem involving the vector a' and a new sum C' and then to solve (or decipher) for x. To begin with, it is a simple mathematical exercise to calculate the inverse of w modulo m, that is, the number w^{-1} that when multiplied by w modulo m gives 1. There is a fast procedure for finding inverses in modular arithmetic (based on Euclid's algorithm for finding the greatest common divisor of two numbers) that makes this calculation efficient, even in a realistic system in which w and m are on the order of 50 digits long. (Incidentally, for this purpose w and m must be chosen to be relatively prime; if they had a common factor, or divisor, there would be no multiplicative inverse of w modulo m.)

To decipher the message C, then, the receiver first calculates $C' = Cw^{-1}$ modulo m. To see what this operation accomplishes remember that C equals $a_1x_1 + \ldots + a_nx_n$. In modular arithmetic, as in ordinary arithmetic, it is permissible to multiply both sides of an equation by the same quantity so that Cw^{-1} modulo m equals $a_1x_1w^{-1} + \ldots + a_nx_nw^{-1}$, or $a_1w^{-1}x_1 + \ldots + a_nw^{-1}x_n$, modulo m. The vector a was generated from the vector a', however, by computing $a_i = a_i'w$ modulo m for each i. Hence a_iw^{-1} equals a_i' modulo m for each i, that is, a_1w^{-1} equals a_1' modulo m, a_2w^{-1} equals a_2' modulo m and so on. Substituting these last results into the preceding equation, one discovers that C', or Cw^{-1} modulo m, equals $a_1'x_1 + \ldots + a_n'x_n$, or $a' \cdot x$.

In other words, calculating Cw^{-1} modulo m is all that is needed to convert the problem of deciphering C into an easy knapsack problem. The receiver simply applies his secret vector a' to solve the knapsack problem for C' and recover x. For those who do not have the secret information w and m, however, there is no easily implemented method known of transforming C into C' (or translating the difficult vector a into the easy vector a') for efficient deciphering.

In the 10-element example I have been discussing it is easy to verify that the numbers w and m relating a and a' or C and C' are respectively 764 and 2,731, and that w^{-1} is 1,605. Notice that in this public-key system the trapdoor information w, m and a' is virtually synonymous with the secret deciphering key w^{-1}, m and a'. The same is not true of all public-key cryptosystems. (In practical cryptosystems based on the trapdoor knapsack scheme it may be desirable to introduce additional security by iterating the conversion process, so that the public and the private vectors differ by several transformations and several intermediate vectors.)

Since only the numbers w, or w^{-1}, and m and the vector a' must be kept secret, all users of the trapdoor knapsack system can employ the same public computer program for generating both their public key and their secret parameters. Utilizing a random-number generator to provide the program with a', w and m will serve to ensure that each user's pair of keys is distinct. Similarly, a public program could be made available that would encipher messages and, when it was supplied with the secret parameters, decipher messages. Therefore no mathematical ability is required to implement the trapdoor knapsack cryptosystem. Any useful public-key system must have this same characteristic.

The second public-key system I shall describe is based on an NP problem that has an even longer and more distinguished history of resisting solution than the knapsack problem: the problem of factoring a large number, or finding all the primes that divide it evenly. (A prime number is an integer that is divisible only by 1 and itself.) This problem has been studied since the time of the ancient Greeks, and although some progress has been made with it, factoring a 200-digit number would still take the most powerful modern computer about a billion years. To give a smaller example, consider the problem of factoring 29,083. Calculating by hand, it would take the better part of an hour to find the only two factors of this number: 127 and 229. It takes less than a minute, however, to verify that those factors are correct, suggesting that the problem of factoring is a good basis for the construction of a one-way function. Figuring out how to build a trapdoor into such a function presents more difficult obstacles, but they have been overcome by Rivest, Shamir and Adleman, the designers of the RSA system.

To generate a public enciphering key each user of the RSA public-key system (or rather a program run on his computer) chooses two large random prime numbers p and q. The product n of these two numbers and another random number E are placed in the public file as the user's enciphering key. To apply the key a sender first converts his message into a string of numbers, which he then breaks into blocks $P_1, P_2, \ldots$. In this instance it is not necessary to use binary numbers, but each plaintext number P_i must be between 0 and $n - 1$. (The enciphering and deciphering functions operate modulo n and so can distinguish between numbers in this range only.) Locating the user's public key (E, n) in the directory, the sender computes for each plaintext number P_i the ciphertext number $C_i = P_i^E$ modulo n. For example, if p equals 5, q equals 11 and E equals 3, then the user's enciphering key is (3, 55), and to encipher the plaintext information $P = 2$ a sender would compute $C = 2^3 = 8$ modulo 55. (Because the numbers in this example are so small modularity does not yet play a role.)

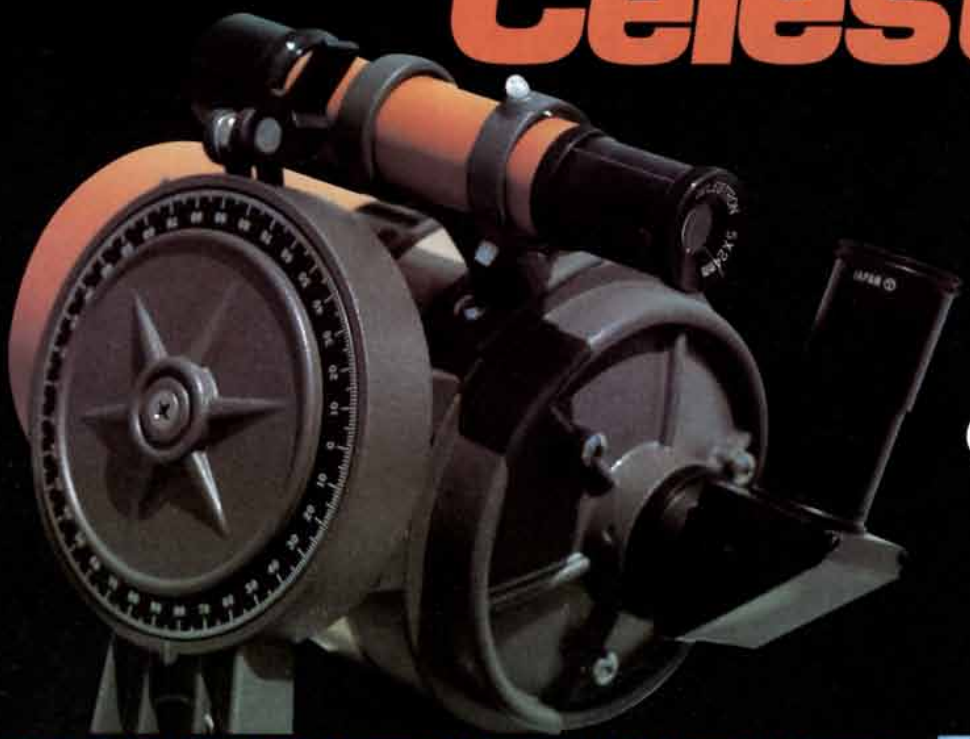
The RSA public-key cryptosystem is based on the fact that although finding large prime numbers is computationally easy, factoring the product of two such numbers is at present computationally infeasible. (It is important to understand that because there are computationally efficient primality tests, determining whether a number is prime is much easier than factoring a number of about the same size.) To decipher a ciphertext C_1, $C_2, \ldots$, the user employs n and a secret deciphering key D derived from the prime factors p and q of n.

To understand how the deciphering key is derived it is necessary to consider the number $(p - 1)(q - 1)$: a well-known object in number theory called Euler's totient function. This function, which is written $\phi(n)$, is defined as the number of integers between 1 and n that have no common factor with n. It is not hard to see that if n equals pq, then $\phi(n)$ equals $(p - 1)(q - 1)$. The number $\phi(n)$ is introduced here because in functions, such as the one used for enciphering in the RSA system, that are calculated modulo n, arithmetic in the exponent is carried out not modulo n but modulo $\phi(n)$. An example may make this idea clearer. Consider the expression 2^{11}

P	0	1	2	3	4	5	6	7	8	9	10	...
$C = P^3$	0	1	8	27	64	125	216	343	512	729	1000	...
$C' = P^3$ modulo 11	0	1	8	5	9	4	7	2	6	3	10	...

MODULAR ARITHMETIC is employed in many cryptosystems to further disguise information already transformed by an enciphering function. As is shown here, the value of an integer a modulo another integer b is defined as the remainder left when a is divided by b. For example, 27 modulo 11 equals 5, because 11 goes into 27 twice with 5 left over. The usefulness of this operation is shown for a simple enciphering function $C = P^3$. As P increases, the continuous way P^3 increases makes it possible to invert the function, or determine what value of P corresponds to a particular value of C, even though there is no simple formula for expressing P as the cube root of C. More precisely, a value of P that gives too small a value of C is itself too small, whereas value of P that gives too large value of C is itself too large. When modularity is added, however, so that C' equals P^3 modulo 11, values of function are thrown into disarray. As P increases, C' changes in a quite discontinuous way, effectively shielding P.

Celestron 90



MAKSUTOV-CASSEGRAIN TELESCOPE

Dealer inquiries invited

FOR THE PHOTOGRAPHER: BEGINNING ASTRONOMER: FOR THE HOME WITH A VIEW: FOR THE NATURALIST:

The Celestron 90 is the most compact and lowest-priced member of the Celestron family of telescopes. And, judging from the overwhelming response to the C90, it may well become the most popular Celestron telescope.

The C90 is unusually small and lightweight for its focal length (1000mm). This — coupled with its unusually-low price and high quality — makes the C90 Astro Telescope extremely desirable as a telephoto lens or spotting scope too. Accordingly, we've designed a complete product line around the C90 optical system.

The heart of the C90 series is the 90mm aperture, f/11, Maksutov-Cassegrain optical system, which is identical on all three C90 versions. As a result, the C90 is a telescope that's also a telephoto or a telephoto that's also a telescope. By adding the appropriate optional accessories, it's possible to use any C90 version in another mode.

The C90 Astro Telescope is supplied with advanced features usually found only on larger, more expensive telescopes — features like star locating setting circles, an automatic tracking system and slow-motion controls on both axes — that make using the telescope easy and convenient. The standard ocular and Barlow lens provide magnifications of 55×, 140× and 200×.

As a daytime telescope, the C90 is remarkably powerful. With it, you'll see bees buzzing in flowers 20 feet away, recognize the faces of friends ½ mile away, read auto license plates 2 miles away, or identify airliners many miles away.

At night, there are countless wonders in the night sky visible through the C90. The Moon, our nearest celestial neighbor, will display a multitude of craters, rugged mountains, rills and relatively smooth plains (maria).

Free 4-page brochure. Or Send $2.00 for 32-page full color catalog on how to select and use a Celestron telescope.

C90 Telephoto Lens

C90 Astro Telescope

C90 Spotting Scope

Celestron

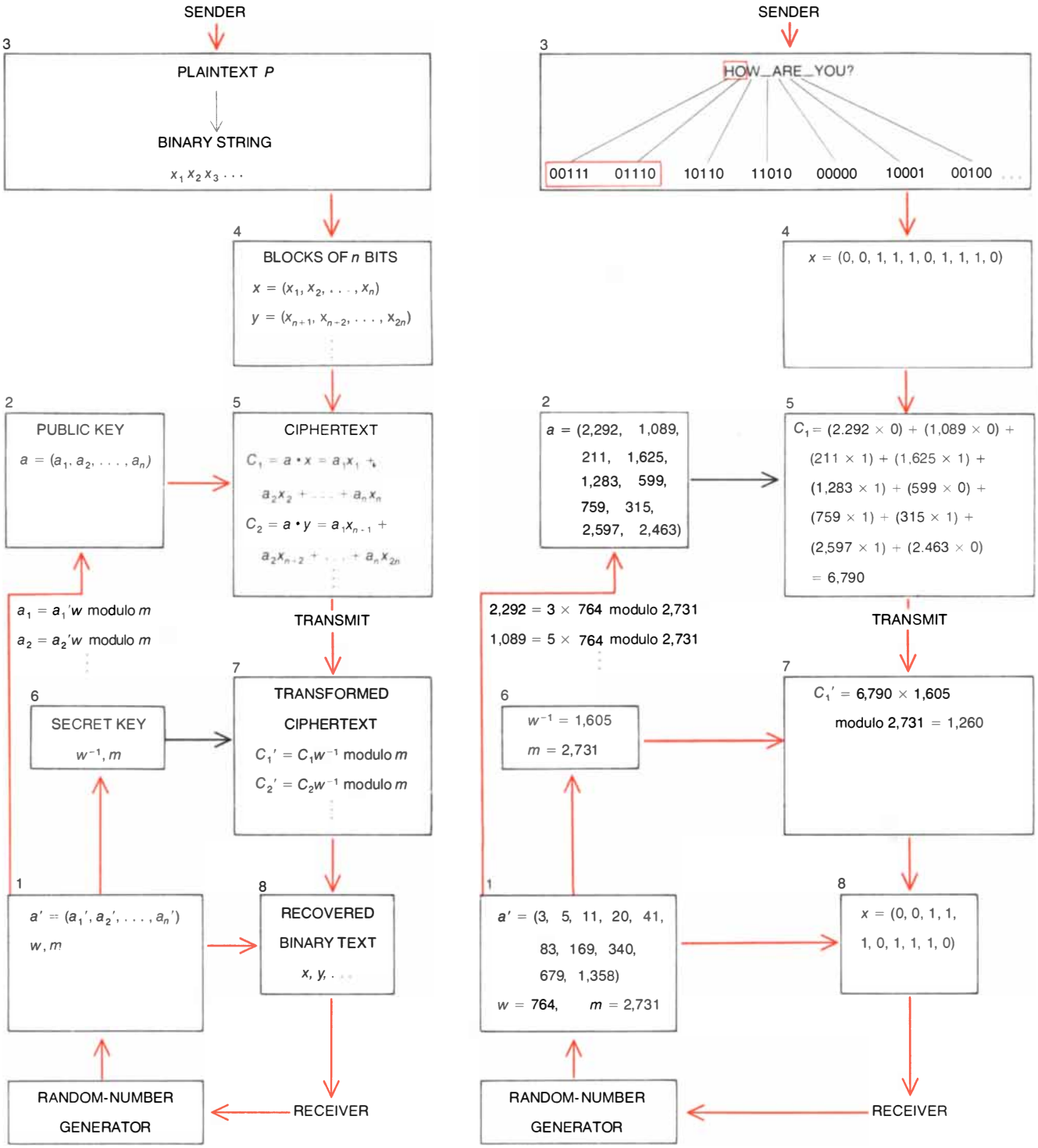
2835 Columbia Street ● Box 3578-2C ● Torrance, CA 90503 ● Phone (213) 328-9560

modulo 10. Since 2^{11} is equal to 2,048 and dividing 10 into 2,048 leaves a remainder of 8, the expression is equal to 8. Note that calculating by first reducing the exponent modulo 10 does not give the correct answer, since 11 modulo 10 equals 1, and 2^1 equals 2. On the other hand, 10 equals 2×5, and so $\phi(10)$ equals $(2-1)(5-1)$, or 4. Since 11 modulo 4 equals 3, calculating 2^{11} modulo 10 by first reducing the exponent modulo 4 gives the correct answer: 2^3, or 8.

Now, the properties of $\phi(n)$ guarantee that there is always a multiplicative inverse D of E modulo $\phi(n)$, that is, ED modulo $(p-1)(q-1)$ is equal to 1. In fact, there is always a fast, computationally easy method for deriving D. (It is not hard to see that in the example discussed above, where p equals 5, q equals 11 and E equals 3, $(p-1)(q-1)$ equals 40 and D equals 27, because 3×27 is one more than 2×40.) This inverse D is the secret deciphering key for the RSA system. To decipher a ciphertext the receiver computes C_i^D modulo n for each ciphertext number C_i. C_i equals P_i^E modulo n, so that C_i^D modulo n equals $(P_i^E)^D$, or P_i^{ED}, modulo n. Because arithmetic in the exponent is performed modulo $\phi(n)$ and ED modulo $\phi(n)$ equals 1, P_i^{ED} modulo n equals P_i^1, or P_i. In other words, raising the ciphertext to the Dth power and reducing modulo n recovers the plaintext.

Hence in the RSA cryptosystem modularity plays a dual role, not only blocking the recovery of the secret deciphering key D from the public enciphering key (E, n) but also, by its presence in the enciphering algorithm, preventing a direct recovery of the plaintext from the ciphertext. The difficulty of computing D from the public information (E, n) depends on the difficulty of factoring n, or of deriving p and q from n. Once again the example I have given is too small to provide real secrecy, but since factoring large numbers is a very difficult problem, the difficulty of breaking the cipher blows up rapidly as n increases. When p and q are chosen so that n is about 200 digits long, it appears to be computationally infeasible for anyone but the intended receiver to decipher the message.

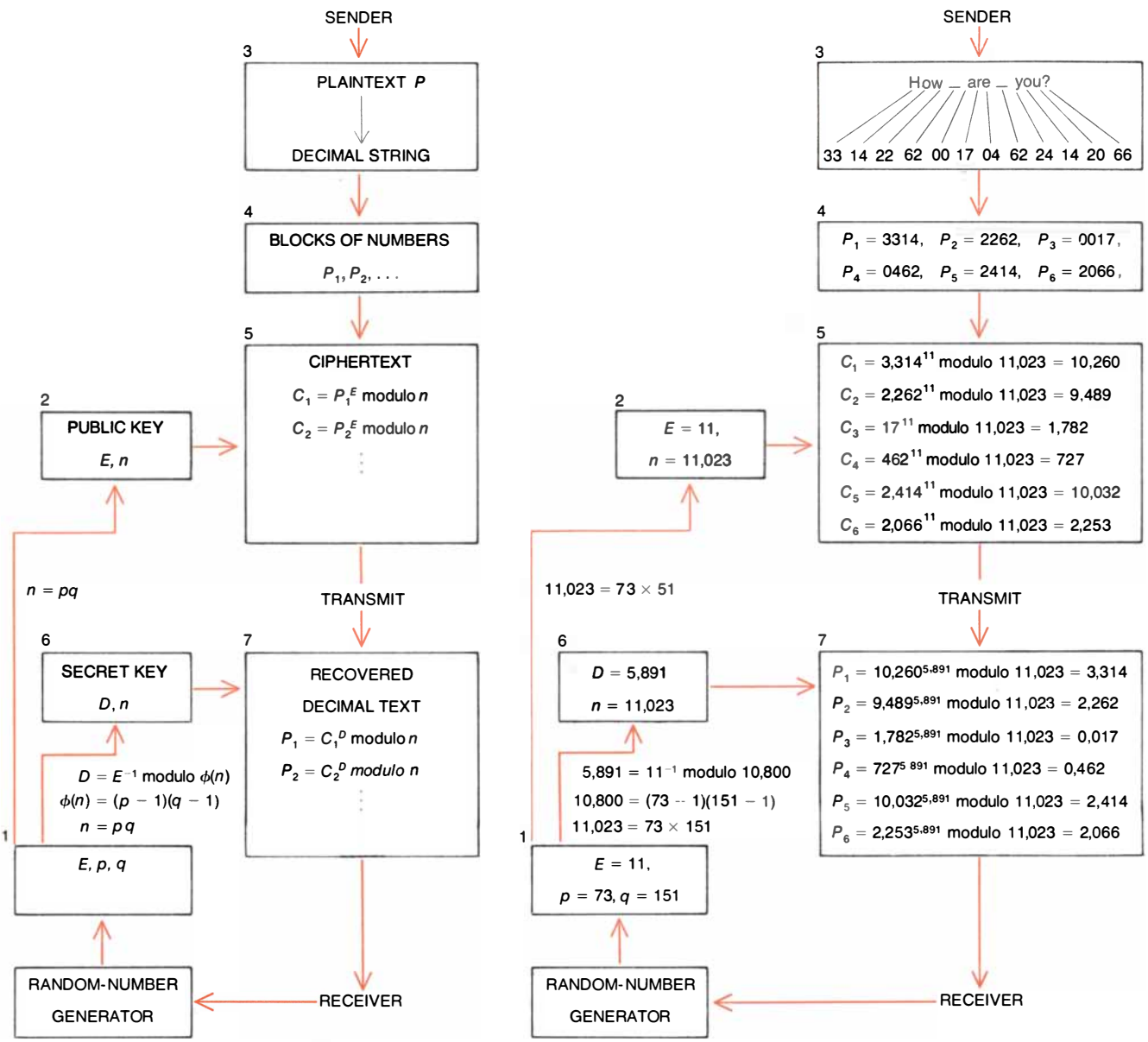
Just as the deciphering procedure (without the trapdoor information) must be computationally infeasible, the public enciphering procedure and secret deciphering procedure must be computationally efficient. At first the implementation of the RSA system appears to present some practical problems in this area. Consider the simple example in which the plaintext number $P = 2$ was transformed into the ciphertext number $C = 8$. To apply the deciphering algorithm $P = C^D$ modulo n it is necessary to calculate 8^{27} modulo 55 (which does indeed equal 2). Multiplying 8 by itself 27 times is, however, a cumbersome process involving large numbers and a great



FLOW OF INFORMATION in the trapdoor knapsack cryptosystem is shown at the left. The corresponding transformations of the first block of plaintext HO are shown at the right. In this public-key system based on the knapsack problem each receiver (by means of a random-number generator) selects a secret vector $a' = (a_1', \ldots, a_n')$ with the property that each element is greater than the sum of the preceding elements and also selects two large random numbers w and m with no common factors (*1*). The numbers w and m are the trapdoor parameters for converting the secret "easy" vector a' into a "difficult" public vector $a = (a_1, \ldots, a_n)$ by means of the equations $a_1 = a_1'w$ modulo m, $a_2 = a_2'w$ modulo m and so on. The difficult vector a is transmitted to the sender over an insecure channel or is listed in a public directory as the receiver's enciphering key (*2*). To encipher a plaintext P a sender begins by converting it into a binary string according to, say, the five-bit binary alphabet given at the top of the illustration on page 146 (*3*). The sender then looks up the receiver's public key and breaks the string into blocks of n binary digits (*4*). For example, in the system shown at the right (in which the numbers are far too small to provide real secrecy) there are 10 elements in the public vector, and so the binary string is divided into blocks of 10 bits each. Every block is enciphered by forming its dot product with a (*5*), that is, if the first block is $x = (x_1, \ldots, x_n)$, the first ciphertext number C_1 equals $a \cdot x$, or $a_1x_1 + \ldots + a_nx_n$, and so on. The ciphertext numbers are transmitted to the receiver over an insecure channel. The receiver recovers, say, x by calculating first w^{-1} (*6*) and then $C_1' = C_1w^{-1}$ modulo m (*7*). (The number w^{-1} is the inverse of w modulo m, the number that when multiplied by w gives 1 modulo m.) Remember that a_1 equals $a_1'w$ modulo m, a_2 equals $a_2'w$ modulo m and so on, and therefore a_1' equals a_1w^{-1} modulo m, a_2' equals a_2w^{-1} modulo m and so on. Hence C_1', or C_1w^{-1} modulo m, which equals $a_1x_1w^{-1} + \ldots + a_nx_nw^{-1}$, or $a_1w^{-1}x_1 + \ldots + a_nw^{-1}x_n$, modulo m, equals $a_1'x_1 + \ldots + a_n'x_n$. In other words, C_1' equals $a' \cdot x$, and the difficult knapsack problem of recovering x from C_1 and a has been converted back into the easy problem of recovering x from C_1' and a'. Only receiver, who possesses trapdoor information w and m and knows secret vector a', can effect transformation and recover x (*8*).

many computational steps. In more realistic RSA systems, where D would be a 200-digit number, this procedure would be impossible to carry out, even on a very powerful computer.

Fortunately there is a much faster method for calculating functions of this kind. First the binary expansion of the exponent (the expression of the exponent as a sum of powers of 2) is utilized to break up the function into a product of smaller factors; for example, 27 equals $1 + 2 + 8 + 16$, and so 8^{27} equals $8 \times 8^2 \times 8^8 \times 8^{16}$. Now, by calculating the smaller factors first and then taking their product, the number of operations required can be limited. For example, 8^2 modulo 55 can be evaluated with only one modular multiplication (an ordinary multiplication followed by an ordinary division), since 8×8, or 64, modulo 55 equals 9. Then 8^4, which is equal to $8^2 \times 8^2$, can be evaluated with an additional modular operation, $9 \times 9 = 81 = 26$ modulo 55, and so on. (Substituting the value of 8^2 modulo 55, namely 9, into the larger factors prevents the size of the numbers involved in the computation from blowing up.) Hence only seven modular multiplications are needed to calculate 8^{27}: four to evaluate 8^2, 8^4, 8^8 and 8^{16} and three more to multiply 8 times 8^2 times 8^8



RSA PUBLIC-KEY CRYPTOSYSTEM is based on the problem of factoring a large number, or finding all the prime numbers that divide it evenly. (A prime number is an integer that is divisible only by 1 and itself.) As is shown at the left, each receiver in the RSA system generates two large random prime numbers p and q, which serve as his secret trapdoor parameters, and a large random number E (*1*). There are computationally efficient tests for identifying primes such as p and q, but when these numbers are sufficiently large, it is computationally infeasible to derive them from their product pq, or n. The receiver lists E and n as his public deciphering key (*2*). To encipher a plaintext P the sender first converts it into a string of numbers, using the decimal alphabet shown at the bottom of the illustration on page 146 (*3*), and then breaks the string into blocks of equal length $P_1, P_2, \ldots$, so that each number P_i in this series is less than n (*4*). Each block is converted into a ciphertext number by raising it to the Eth power and then reducing modulo n, that is, C_1 equals P_1^E modulo n, C_2 equals P_2^E modulo n and so on (*5*). The ciphertext numbers $C_1, C_2, \ldots$ are transmitted over an insecure channel. Arithmetic in the exponent of a function that is calculated modulo n must be carried out modulo $\phi(n)$, where $\phi(n)$ equals $(p-1)(q-1)$, and so the receiver utilizes p and q to determine $\phi(n) = (p-1)(q-1)$ and then $D = E^{-1}$ modulo $\phi(n)$, which serves as his secret deciphering key (*6*). To convert $C_1, C_2, \ldots$ back into plaintext numbers the receiver raises each one to the Dth power and reduces it modulo n (*7*). Because C_i^D modulo n equals $(P_i^E)^D$, or P_i^{ED}, modulo n and ED modulo $\phi(n)$ equals 1, P_i^{ED} equals P_i modulo n. In other words, this operation inverts the enciphering transformation, recovering the plaintext number blocks $P_1, P_2, \ldots$. In the example shown at right the values of E, p and q are too small to provide real security.

times 8^{16}. Even when D is a 200-digit number, this method results in a deciphering procedure that is quite efficient, requiring at most 1,330 modular multiplications rather than the 10^{200} operations necessary in the straightforward approach.

The RSA system is a public-key cryptosystem that allows the direct generation of a digital signature: a number that can be appended to a ciphertext message to solve the problems of authentication mentioned above. To be of real service such a signature must be easy for the sender to generate and for the receiver to check but must be computationally infeasible for a third party or the receiver himself to generate. Of the various methods for generating digital signatures the simplest involves exploiting the inverse relation of the public enciphering and secret deciphering keys by reversing their roles. For example, in the RSA system the sender can utilize his own secret deciphering key D as a signing key, to compute the signature $S_i = P_i^D$ modulo n for each P_i in the series of plaintext numbers P_1, P_2,... that represent a message to be transmitted. (Remember that each P_i is chosen to be between 0 and $n - 1$.) Once the signatures S_1, S_2,... have been generated the sender enciphers each signed message block (P_i, S_i), using the receiver's public enciphering key. This second operation has nothing to do with signature generation; it simply ensures the privacy of the communication.

The receiver uses his own secret key to recover the signed message block (P_i, S_i), and then he looks up the sender's public key (E, n) and computes S_i^E modulo n for each i. D and E effect inverse operations regardless of the order of their application, and so since S_i equals P_i^D modulo n, S_i^E modulo n should be equal to P_i. If that is the case, the receiver can be sure that the message comes from the apparent sender and that it has not been tampered with. Since the digital signature depends on both the sender and the message sent, it offers a level of security different from that of a written signature, which is the same for all messages. With the digital signature neither the receiver nor a third party can alter the message without destroying the validity of the signature. (When the message to be sent is long, rather than signing each submessage separately it may be desirable to compress the message and calculate a single signature S; that compression can be effected in such a way that S still depends on the entire message P.)

The traditional difficulties of solving the knapsack and factoring problems can be taken as an encouraging sign that the public-key cryptosystems based on these problems are in practical terms secure. A past history of intractability cannot, however, be considered a proof that a system is secure. It is always possible, if unlikely, that at some time in the future computationally efficient general methods for solving these problems will be found. An even greater hazard is that a method will be discovered for breaking one of the cryptosystems without solving the corresponding general problem. For example, it is possible that although solving most regular knapsack problems is computationally infeasible, there is an easy way to solve the much smaller set of trapdoor knapsack problems. Similarly, it may be possible to recover the plaintext enciphered by the RSA technique without finding the factors of n. (Michael O. Rabin of the Hebrew University of Jerusalem has recently shown, however, that in the case where the enciphering exponent E is 2, the security of the RSA system is not simply dependent on the difficulty of factoring n but is actually equivalent to it. This finding constitutes an important first step toward the goal of developing provably secure systems.)

Cryptography has not yet advanced to the stage where it can prove the computational security of even a conventional system or a one-way function. Hence it is not surprising that there is no way to establish the security of the public-key systems, which are based on the more complex trapdoor one-way functions. It is hoped, however, that over the next decade or two complexity theory will advance to the point where such proofs can be formulated. Some progress has been made in this direction through the study of a special subset of the NP problems.

Remember that the NP problems are ideal candidates for one-way functions because finding a solution to them is computationally difficult but checking a proposed solution is computationally easy. Some of these problems such as the knapsack problem (but not factoring) belong to the subset of the NP problems that is called NP-complete. The NP-complete problems have the added property that if any one of them had an easily implemented method for finding general solutions, then all the NP problems would. Now, all cryptanalytic problems—problems of breaking cryptosystems—are in the class NP, since it is always easy to check the validity of a proposed key. Therefore if any NP-complete problem can be solved rapidly, it follows that all cryptographic systems can be broken easily. Roughly speaking, then, if the security of a cryptosystem could be shown to be equivalent in difficulty to an NP-complete problem, it would be as secure as any cryptographic system can be.

One flaw in this type of evaluation is that complexity theory deals with the "worst case" computational difficulty of solving a problem, whereas cryptography is concerned with the average or typical difficulty of solving a problem. For example, in current complexity theory a problem whose solution requires $10^{1,000}$ operations 1 percent of the time but only 100 operations 99 percent of the time is considered to be difficult. Obviously a cryptosystem that can be broken 99 percent of the time is worthless. Workers in complexity theory are aware of this shortcoming and are currently developing more suitable measures of computational difficulty.

Although factoring is not an NP-complete problem, it has through the years largely resisted the attack of some of the best mathematical minds. That is why Rabin's proof, which establishes an equivalence between the difficulty of factoring and breaking an RSA scheme, is an extremely important result. Until such time as the security of proposed cryptosystems can be formally evaluated, however, it is a worthwhile (and intellectually challenging) exercise to try to break them.

In electronic communications systems, as in any new technology, there is a potential for misuse. For example, the danger of foreign or domestic intelligence organizations spying on American citizens who rely on these systems is a real one. It has recently been revealed that U.S. microwave telephone traffic is being monitored in at least one foreign embassy in Washington. In the late 1960's the U.S. Government's "Operation Shamrock" intercepted international Telex communications to and from "targeted" individuals, including antiwar activists. If such excesses are to be limited, both legal and technical safeguards are needed.

There is always a trade-off between the rights of citizens to privacy and the desire of government intelligence agencies to limit the availability of secure cryptosystems. A conflict in this area has recently arisen concerning the Federal Data Encryption Standard, a conventional cryptosystem issued by the National Bureau of Standards for nonmilitary encryption purposes. The National Security Agency convinced the International Business Machines Corporation, the company that designed the standard, to reduce the key size to 56 bits. Although there is controversy surrounding the issue, I believe the reduction in key size was meant to weaken the standard so that if it were ever employed by a foreign organization, it could be broken by the National Security Agency. Issues similar to this one will certainly arise as the new public-key systems become commercial realities. It is to be hoped that these issues will be decided by an open discussion of the relative needs of the intelligence community and the citizenry rather than, as appears to have been the case, by the unilateral decision of the intelligence community that its needs take precedence.

Problems in Physics with Many Scales of Length

Physical systems as varied as magnets and fluids are alike in having fluctuations in structure over a vast range of sizes. A novel method called the renormalization group has been invented to explain them

by Kenneth G. Wilson

One of the more conspicuous properties of nature is the great diversity of size or length scales in the structure of the world. An ocean, for example, has currents that persist for thousands of kilometers and has tides of global extent; it also has waves that range in size from less than a centimeter to several meters; at much finer resolution seawater must be regarded as an aggregate of molecules whose characteristic scale of length is roughly 10^{-8} centimeter. From the smallest structure to the largest is a span of some 17 orders of magnitude.

In general, events distinguished by a great disparity in size have little influence on one another; they do not communicate, and so the phenomena associated with each scale can be treated independently. The interaction of two adjacent water molecules is much the same whether the molecules are in the Pacific Ocean or in a teapot. What is equally important, an ocean wave can be described quite accurately as a disturbance of a continuous fluid, ignoring entirely the molecular structure of the liquid. The success of almost all practical theories in physics depends on isolating some limited range of length scales. If it were necessary in the equations of hydrodynamics to specify the motion of every water molecule, a theory of ocean waves would be far beyond the means of 20th-century science.

A class of phenomena does exist, however, where events at many scales of length make contributions of equal importance. An example is the behavior of water when it is heated to boiling under a pressure of 217 atmospheres. At that pressure water does not boil until the temperature reaches 647 degrees Kelvin. This combination of pressure and temperature defines the critical point of water, where the distinction between liquid and gas disappears; at higher pressures there is only a single, undifferentiated fluid phase, and water cannot be made to boil no matter how much the temperature is raised. Near the critical point water develops fluctuations in density at all possible scales. The fluctuations take the form of drops of liquid thoroughly interspersed with bubbles of gas, and there are both drops and bubbles of all sizes from single molecules up to the volume of the specimen. Precisely at the critical point the scale of the largest fluctuations becomes infinite, but the smaller fluctuations are in no way diminished. Any theory that describes water near its critical point must take into account the entire spectrum of length scales.

Multiple scales of length complicate many of the outstanding problems in theoretical physics and in certain other fields of study. Exact solutions have been found for only a few of these problems, and for some others even the best-known approximations are unsatisfactory. In the past decade a new method called the renormalization group has been introduced for dealing with problems that have multiple scales of length. It has by no means made the problems easy, but some that have resisted all other approaches may yield to this one.

The renormalization group is not a descriptive theory of nature but a general method for constructing theories. It can be applied not only to a fluid at the critical point but also to a ferromagnetic material at the temperature where spontaneous magnetization first sets in, or to a mixture of liquids at the temperature where they become fully miscible, or to an alloy at the temperature where two kinds of metal atoms take on an orderly distribution. Other problems that have a suitable form include turbulent flow, the onset of superconductivity and of superfluidity, the conformation of polymers and the binding together of the elementary particles called quarks. A remarkable hypothesis that seems to be confirmed by work with the renormalization group is that some of these phenomena, which superficially seem quite distinct, are identical at a deeper level. For example, the critical behavior of fluids, ferromagnets, liquid mixtures and alloys can all be described by a single theory.

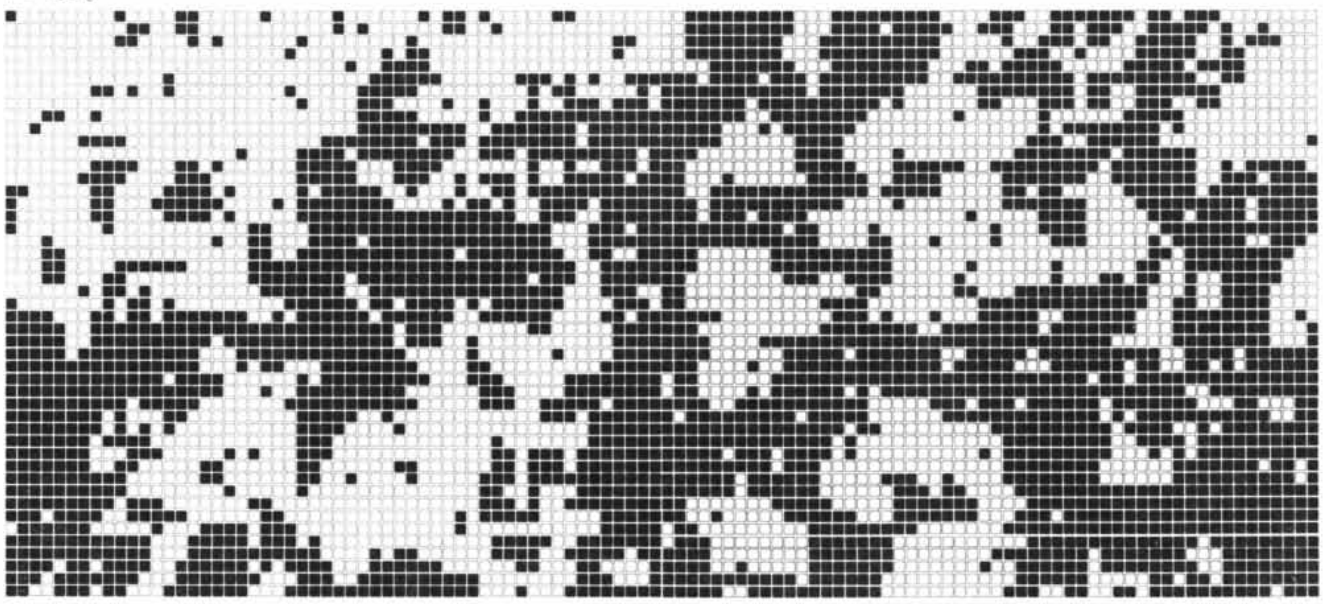
The most convenient context in which to discuss the operation of the renormalization group is a ferromagnet, or permanent magnet. Ferromagnetic materials have a critical point called the Curie point or the Curie temperature, after Pierre Curie, who studied the thermodynamics of ferromagnets at about the turn of the century. For iron the Curie temperature is 1,044 degrees K. At higher temperatures iron has no spontaneous magnetization. As the iron is cooled the magnetization remains zero until the Curie temperature is reached, and then the material abruptly becomes magnetized. If the temperature is reduced further, the strength of the magnetization increases smoothly.

Several properties of ferromagnets besides the magnetization behave oddly

MULTIPLE SCALES OF LENGTH characterize the patterns that emerge when a ferromagnetic solid is cooled to the temperature at which it becomes spontaneously magnetized. Each square represents the magnetic moment associated with a single atom in the solid, and each moment is assumed to have only two possible orientations, labeled "up" (*black squares*) and "down" (*open squares*). At high temperature (*top*) the orientation of the magnetic moments is essentially random, and so there is only short-range order in the pattern. As the temperature is reduced (*middle*) somewhat larger patches in which most of the magnetic moments are lined up in the same direction begin to develop. When the temperature reaches a critical value called the Curie temperature, or T_c (*bottom*), these patches expand to infinite size; significantly, however, fluctuations at smaller scales persist. As a result all scales of length must be included in a theoretical description of the ferromagnet. This simulation of a ferromagnet was carried out with the aid of a computer by Stephen Shenker and Jan Tobochnik of Cornell University.

$T = 2T_c$

$T = 1.05T_c$



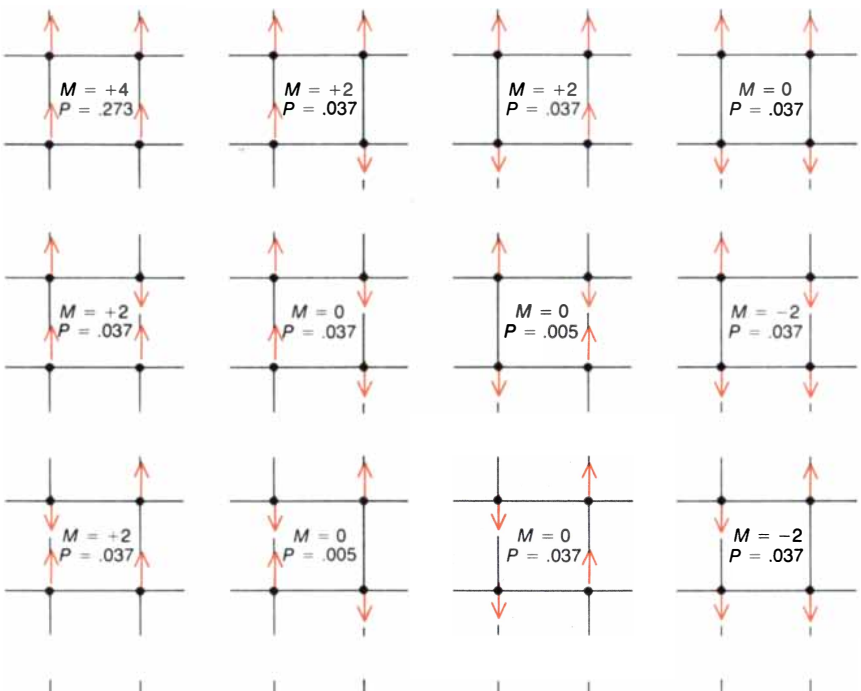
$T = T_c$

near the Curie point. Another property of interest is the magnetic susceptibility, or the change in magnetization induced by a small applied field. Well above the Curie point the susceptibility is small because the iron cannot retain any magnetization; well below the Curie temperature the susceptibility is small again because the material is already magnetized and a weak applied field cannot change the state of the system very much. At temperatures close to 1,044 degrees, however, the susceptibility rises to a sharp peak, and at the Curie point itself the susceptibility becomes infinite.

The ultimate source of ferromagnetism is the quantum-mechanical spinning of electrons. Because each electron rotates it has a small magnetic dipole moment; in other words, it acts as a magnet with one north pole and one south pole. How the spin of the electron gives rise to the magnetic moment will not concern me here. It is sufficient to note that both the spin and the magnetic moment can be represented by a vector, or arrow, which defines the direction of the electron's magnetic field.

A real ferromagnet has a complex atomic structure, but all the essential properties of the system of spins can be illustrated by a quite simple model. Indeed, I shall describe a model that includes no atoms or other material particles but consists only of spin vectors arranged in a lattice. For the sake of simplicity I shall deal with a two-dimensional lattice: a rectilinear grid of uniformly spaced lines in a plane, with a spin vector at each intersection of the grid lines. Furthermore, I shall assume that each spin can point in only two possible directions, designated up and down. The model lattice is said to be magnetized whenever more than half of the spins point in the same direction. The magnetization can be defined as the number of up spins minus the number of down spins.

Every electron has the same spin and the same magnetic dipole moment. What distinguishes a ferromagnet from other materials is a coupling between nearby spins that makes them tend to line up in the same direction. This tendency can be stated more precisely by pointing out that the total energy of any two adjacent spins is smaller when the spins are parallel than it is when they are antiparallel. The interaction responsible for the coupling of the spins has a short range, which is reflected in the model by specifying that only nearest-neighbor spins are coupled to each other. In the two-dimensional rectilinear lattice each spin is influenced by four nearest neighbors; no other spins have any direct effect on it.



MODEL OF A FERROMAGNET consists of vectors, or arrows of fixed length, arranged at the sites of a lattice. Each vector represents the spin angular momentum and the magnetic moment of a single electron, and it can be oriented either up or down. Nearest-neighbor lattice sites are coupled in such a way that adjacent spin vectors are likelier to be parallel than antiparallel. From the strength of the coupling, which declines as the temperature increases, a probability, *P*, can be assigned to every possible configuration of the spin vectors. All the configurations of a lattice made up of just four sites are shown here. The net magnetization, *M*, of each configuration is easily calculated: it is the number of up spins minus the number of down spins. The magnetization of the model at any given temperature is found by multiplying the magnetization of each configuration by the probability of that configuration, then adding up all the results. The probabilities shown were calculated for a coupling strength of .5, which corresponds to a temperature (in arbitrary units) of 2. The model is called the two-dimensional Ising model.

From the nature of the interaction between spins in a ferromagnet one might well predict that all the spins would always be parallel and the material would always have its maximum magnetization. That is the state of lowest energy, and in the absence of any perturbing effects it would be the favored state. In a real ferromagnet, however, there is one perturbation that cannot be neglected: the thermal motion of the atoms and the electrons. At any temperature above absolute zero thermal excitations of the solid randomly flip some of the spins so that the direction of the spin vector is reversed, even when reversing the spin puts the magnet in a state of higher energy. Hence it is no surprise that the magnetization decreases as the temperature increases: that relation simply reflects increasing thermal disruption. What remains curious is that the magnetization is not a smooth function of the temperature but instead disappears abruptly at a certain finite temperature, the Curie point.

Competition between the tendency toward a uniform spin orientation and the thermal introduction of disorder can readily be incorporated into a model of a ferromagnet. The strength of the coupling between adjacent spins is given by a number, K, that must be specified in the design of the model. Thermal effects are included simply by making K inversely proportional to the temperature. With the appropriate units of measurement the coupling strength can be set equal to the reciprocal of the temperature, a relation expressed by the equation $K = 1/T$.

What the coupling strength determines is the probability that two adjacent spins are parallel. When the temperature is zero, there are no thermal effects and adjacent spins are certain to be parallel; the probability is equal to 1 and the coupling strength is infinite. At an infinite temperature the coupling strength falls to zero, so that the spins do not interact at all. Hence each spin is free to choose its direction randomly and is independent of its neighbors. The probability that two spins are parallel is 1/2, and so is the probability that they are antiparallel. The region of interest,

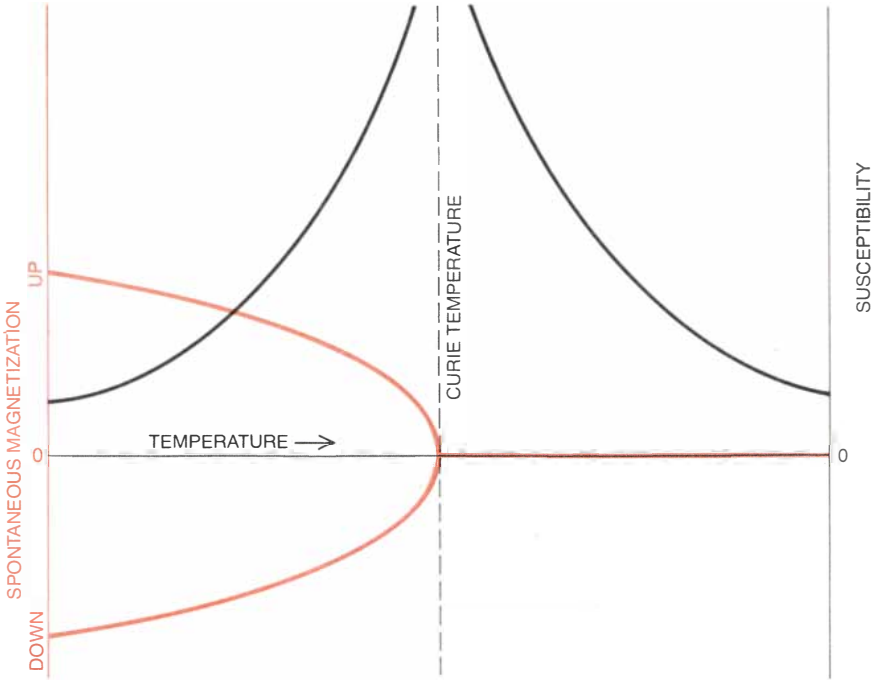
of course, lies between these extremes of temperature, where the probability of the adjacent spins' lining up must always have a value between 1/2 and 1.

Suppose there is a large two-dimensional lattice of spins and that some one spin in it is artificially held fixed in the up orientation. What is the effect on the other spins? The effect on the spins at the four adjacent lattice sites is easy to imagine: since they are directly coupled to the fixed spin, they will have a greater-than-even probability of pointing up. The extent to which the probability is biased depends on the value of K, which is determined in turn by the temperature.

More distant spins have no direct interaction with the fixed spin, but nonetheless the influence of the fixed spin does not end with the immediate neighbors. Because the nearest-neighbor spins tend to point up more often than down they create a similar bias in their own nearest neighbors. In this way the disturbance can propagate over a large area of the lattice. The range of influence of a single fixed spin can be measured by observing the orientation of many spins that are all at the same large distance from the fixed one. If reversing the orientation of the fixed spin from up to down increases the number of down spins in the distant population, then the spins are said to be correlated. The maximum distance over which such a correlation can be detected is called the correlation length. Regions separated by a distance greater than the correlation length are essentially independent.

In a lattice at very high temperature the correlation length is close to zero. The distribution of spins is nearly random, and so the average number of up and down spins must be equal; in other words, the magnetization is zero. As the temperature falls (and the coupling strength increases) correlations over larger distances begin to appear. They take the form of spin fluctuations, or patches of a few spins each that mostly point in the same direction. Over any large area the magnetization is still zero, but the structure of the lattice is much different from what it was near infinite temperature.

As the temperature approaches the Curie point the correlation length grows rapidly. The basic interactions of the model have not changed; they still connect only adjacent lattice sites, but long-range order has emerged from the short-range forces. What is most significant in the growth of the correlation length is that as the maximum size of the spin fluctuations increases, the smaller fluctuations are not suppressed; they merely become a finer structure superimposed on the larger one. The largest fluctuations are not areas of uniform spin alignment; they include many smaller fluctuations and can be distinguished only because they have an overall excess of one spin direction. Thus an ocean of spins that are mostly up may have within it an island of spins that are mostly down, which in turn surrounds a lake of up spins with an islet of down spins. The progression continues to the smallest possible scale: a single spin.



MAGNETIZATION of a ferromagnet has a sudden onset at the Curie temperature. Above this temperature the average numbers of up spins and down spins are equal and so the magnetization is zero. At any temperature below the Curie point two states of magnetization are possible, depending on whether the up spins or the down spins are in the majority; in the absence of an external magnetic field the two states are equally likely. The susceptibility of a ferromagnet measures the change in magnetization induced by an arbitrarily small applied magnetic field. At the Curie point the susceptibility becomes infinite. Near the Curie point a small change in either the temperature or the external field gives rise to a large change in the magnetization.

When the temperature is precisely equal to the Curie temperature, the correlation length becomes infinite. Any two spins are correlated, no matter what the distance between them is. Nevertheless, fluctuations persist at all smaller scales of length. The system remains unmagnetized, but it is exquisitely sensitive to small perturbations. For example, holding a single spin fixed in the up orientation creates a disturbance that spreads throughout the lattice and gives the entire system a net magnetization.

Below the Curie temperature the system becomes magnetized even in the absence of an outside perturbation, but there is no immediate change in the appearance of the lattice. Smaller-scale fluctuations persist; they are remnants of the lakes and islets of opposite spin direction. Merely by looking at the lattice one cannot detect the magnetization. Only when the system is cooled further does the bias become obvious, as the increasing coupling strength coerces more of the spins into conformity with the majority. At zero temperature complete uniformity is attained.

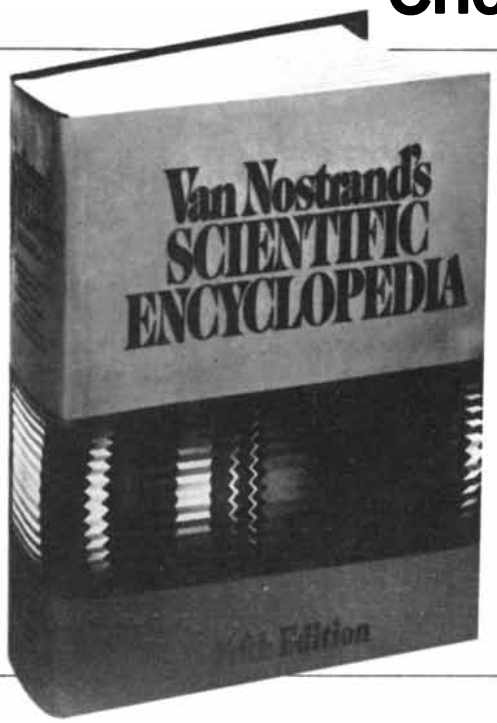
In fluids the fluctuations in density near the critical point are closely analogous to the fluctuations in spin direction observed in ferromagnets. In fluids, however, the presence of fluctuations at all possible scales of length can be observed directly. When the correlation length first reaches a few thousand angstrom units, which is comparable to the wavelength of light, the fluctuations begin to scatter light strongly and the fluid turns milky, a phenomenon called critical opalescence. Significantly, when the temperature comes still closer to the critical point and the maximum scale of the fluctuations becomes much larger (millimeters or centimeters), the critical opalescence is not reduced, indicating that the smaller fluctuations persist. The same phenomenon takes place in spin systems, but because ferromagnetic materials are not transparent to light it cannot be demonstrated as readily. The critical opalescence of ferromagnets has been detected, however, in the scattering of neutrons from a magnetic material near the Curie temperature.

The model I have been describing is not my own invention. It is a version of one introduced in the 1920's by the German physicists Wilhelm Lenz and Ernest Ising, and it is now called the Ising model. The properties of a system of Ising spins on a two-dimensional lattice are known in complete detail because the model was solved exactly in 1944 by Lars Onsager of Yale Universi-

As your introduction to

The Library of Science

Choose either



this $67.50 Classic for only **$3.95** A saving of 94%

Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼" x 12" volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings and charts, and 500 tables. 7200 articles cover mathematics, from information sciences to physics and chemistry. "...an amazing book...for both the general and scientific reader."—The New York Times

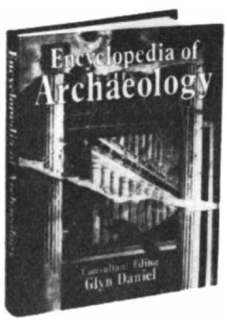
OR

any other 3 books for only $3.95

if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months.

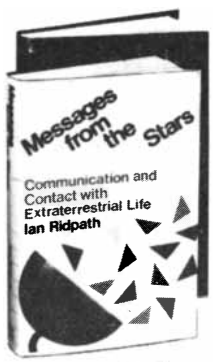
(values to $75.00)

(Publishers' Prices shown)

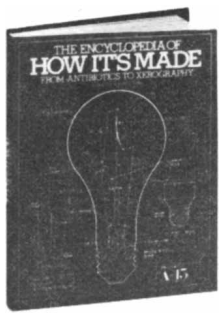


54995. THE ILLUSTRATED ENCYCLOPEDIA OF ARCHAEOLOGY. *Glyn Daniel.* **$17.95**

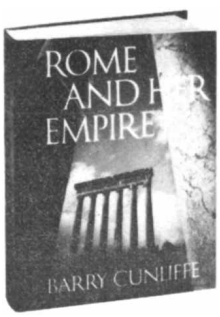
55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. $16.95



74771. MESSAGES FROM THE STARS/ THE RUNAWAY UNIVERSE. A report on the efforts of scientists to locate traces of interstellar communication. *Plus*, the gripping and dramatic story of the birth and death of the cosmos from the Big Bang to titanic future holocaust. *The 2 count as one book.* **$19.95**

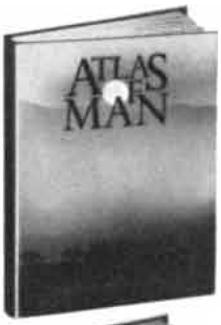


44905. THE ENCYCLOPEDIA OF HOW IT'S MADE. $14.95



74300-2. ROME AND HER EMPIRE. *Barry Cunliffe.* A magnificently illustrated volume that makes the history of one of the world's greatest civilizations come alive. *Counts as 2 of your 3 books.* **$50.00**

63340-2. MYSTERIES OF THE PAST. Captivating investigation of prehistory and great riddles. Outsized. Sumptuously illustrated. *Counts as 2 of your 3 books.* **$34.95**



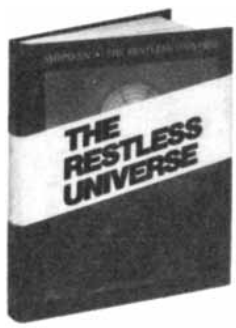
34740. ATLAS OF MAN. Covers 400 groups of peoples, tribes, and nations. 120 color photographs, 350 maps and line illustrations. Covers geography, land use, culture, climate, religion, much more. **$25.00**

81403. SUN, MOON, AND STANDING STONES. *John Edwin Wood.* Includes the latest findings from researchers in archaeology, astronomy, climatology, and physics. Illustrated. **$14.95**

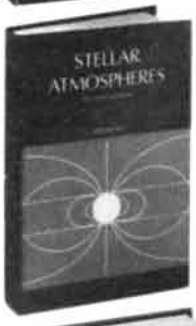
34210. THE ARCHAEOLOGY OF NORTH AMERICA. *Dean Snow.* Prehistoric Indian cultures, from the first crossings into this continent to the 20th century. 195 photos, charts and maps. **$18.95**

46580-2. THE EVOLVING CONTINENTS. *Brian F. Windley.* An integrated global overview of the geosciences. *Counts as 2 of your 3 books.* **$34.95**

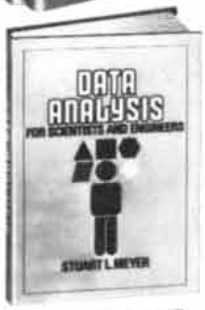
39746-2. COMPANION TO CONCRETE MATHEMATICS. Volumes I and II. *Z. A. Melzak.* Mathematical techniques and various applications. *The set counts as 2 of your 3 books.* **$46.45**



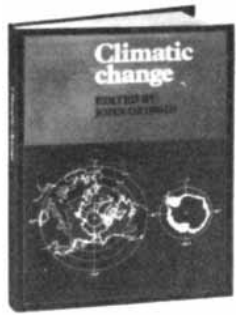
73940. THE RESTLESS UNIVERSE. *Henry L. Shipman.* A highly acclaimed 450-page tour of the ideas, principles, and process of modern astronomy. **$16.95**



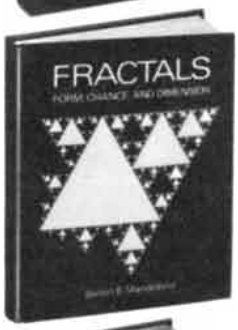
80570. STELLAR ATMOSPHERES. *Dimitri Mihalas.* An authoritative report on the stellar envelope, modes of energy transport in this atmosphere, rates of star mass loss, and much more. **$24.95**



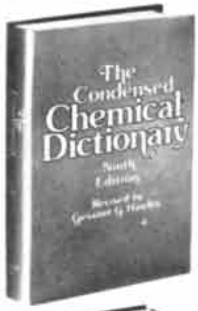
41610. DATA ANALYSIS FOR SCIENTISTS AND ENGINEERS. *Stuart L. Meyer.* 513-page volume covers all the statistical methods, concepts, and analysis techniques that any experimenter will probably ever need. **$19.95**



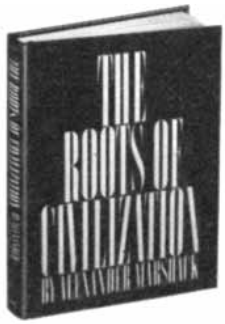
38362-2. CLIMATIC CHANGE. *Dr. John Gribbin.* Provides an A to Z report on the basic pieces and patterns of the earth's rich climatological puzzle. *Counts as 2 of your 3 books.* **$32.50**



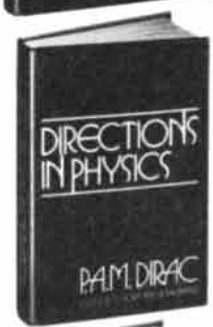
48830: FRACTALS, FORM, CHANCE AND DIMENSION. *Benoit B. Mandelbrot.* A brilliant geometrical investigation into the fragmented and irregular patterns of nature. **$17.50**



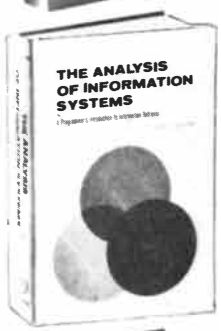
40167-2. THE CONDENSED CHEMICAL DICTIONARY. Updated to meet today's needs. Over 18,000 entries. *Counts as 2 of your 3 books.* **$32.50**



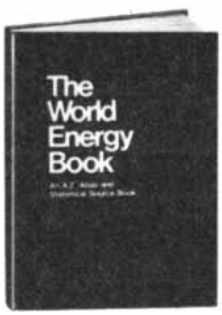
74590. THE ROOTS OF CIVILIZATION. *Alexander Marshack.* A remarkable re-creation of the life of prehistoric man. Illustrated. **$17.50**



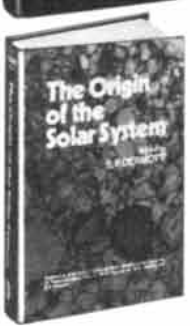
42285. DIRECTIONS IN PHYSICS. *Paul A.M. Dirac.* One of the greatest theoretical physicists of the twentieth century presents his views on cosmological theories, quantum theory, quantum mechanics and quantum electrodynamics. **$12.95**



33430. THE ANALYSIS OF INFORMATION SYSTEMS. Revised Edition. *Charles T. Meadow.* A distinguished computer authority offers an all-inclusive treatment. **$20.95**



87610. THE WORLD ENERGY BOOK. *Crabbe and McBride.* Over 1500 alphabetically arranged entries on everything from fossil fuels to power from sewage. Includes 35 world maps and 40 pages of tables, diagrams, and charts. **$25.00**



66275-2. THE ORIGIN OF THE SOLAR SYSTEM. *Edited by S. F. Dermott.* A compilation of the views of 28 leading astronomers, astrophysicists, cosmologists and mathematicians covering theories of stellar formation. *Counts as 2 of your 3 books.* **$39.00**

37347. CASTE AND ECOLOGY IN THE SOCIAL INSECTS. *Oster and Wilson.* **$20.00**

If the reply card has been removed,
please write to **The Library of Science**
Dept. 2-B6A, Riverside, N.J. 08370
to obtain membership information and application.

ty. Since then solutions have also been found for several other two-dimensional models (whereas no three-dimensional model has yet been solved exactly). Nevertheless, the problems of describing two-dimensional systems are far from trivial. In what follows I shall apply the methods of the renormalization group to the two-dimensional Ising model as if it were a problem still outstanding, and Onsager's solution will serve as a check on the results.

What does it mean to solve or to understand a model of a physical system? In the case of the Ising system the microscopic properties are known completely from the outset, since they were specified in building the model. What is needed is a means of predicting the macroscopic properties of the system from the known microscopic ones. For example, a formula giving the spontaneous magnetization, the susceptibility and the correlation length of the model as a function of temperature would contribute greatly to understanding.

It is not notably difficult to calculate the macroscopic properties of any given configuration of the spins in an Ising model. The magnetization, for example, can be determined simply by counting the number of up spins and the number of down spins and then subtracting. No one configuration of the spins, however, determines the macroscopic properties of the system. Instead all possible configurations contribute to the observed properties, each in proportion to its probability at a given temperature.

In principle the macroscopic properties could be calculated directly as the sum of all the separate contributions. First the magnetization would be found for each configuration and then the corresponding probability. The actual magnetization would be obtained by multiplying each of these pairs of numbers and adding up all the results. The susceptibility and the correlation length could be found by procedures that are not much more elaborate. The common element in all these calculations is the need to determine the probabilities of all possible configurations of the spins. Once the distribution of probabilities is known the macroscopic properties follow directly.

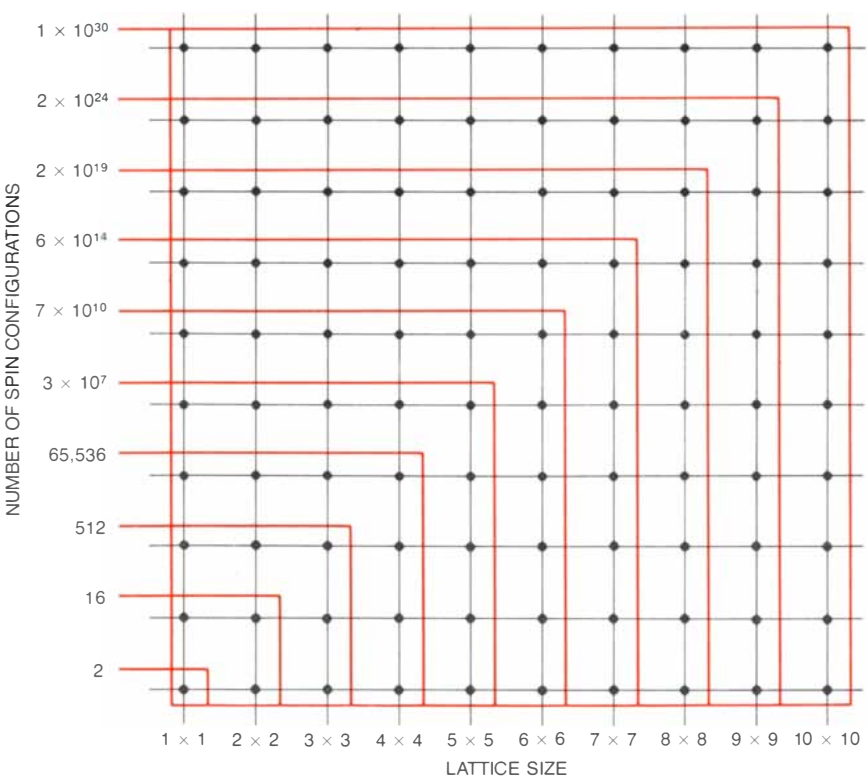
As I pointed out above, the probability of any two adjacent spins' being parallel is determined solely by the coupling strength K, which I have defined as the reciprocal of the temperature. If the probability of two neighboring spins in isolation being parallel is denoted p, then the probability of their being antiparallel must be $1 - p$. From these two values alone the relative probability of any specified configuration of a lattice can be evaluated. All that is required is to multiply together the separate probabilities for every nearest-neighbor pair of spins, in each case taking the value as p when the spins are parallel and as $1 - p$ when they are antiparallel.

Consider a spin system that is made up of just four spins arranged at the corners of a square. Such a lattice has four nearest-neighbor couplings, corresponding to the four sides of the square. Each coupling is considered in turn and is assigned a probability of either p or $1 - p$ according to whether the spins are parallel or antiparallel; then the four separate probabilities are multiplied. In the configuration with all four spins oriented up all four pairs are parallel, and so the relative probability is given by the product $p \times p \times p \times p$. If three spins are up and one is down, the relative probability is $p \times p \times (1 - p) \times (1 - p)$.

The calculation must be carried out for every configuration of the spins; for a system of four spins there are 16 configurations. A final step is to convert the relative probabilities into absolute ones by adjusting each value so that the total of all 16 values is equal to exactly 1. Since the temperature determines the coupling strength and the coupling strength in turn determines the values of p and $1 - p$, the entire sequence of 16 calculations would also have to be repeated for every temperature of interest.

This plan of attack on the Ising model is ambitious but impractical. If the probability of every spin configuration could be calculated, the magnetization and the other macroscopic properties could be evaluated for any specified temperature. The problem lies in the number of spin configurations. For a system made up of n spins, each of which can take on two values, there are 2^n possible configurations. This exponential function grows rapidly as n increases. As I have mentioned, four spins have 2^4, or 16, configurations. A three-by-three block of nine spins has 512 configurations and a four-by-four block has 65,536. The practical limit of computation is not much larger than a six-by-six block of 36 spins, for which there are approximately 7×10^{10} configurations.

What size lattice would be needed in order to determine the critical properties of the two-dimensional Ising model? The array must be at least as large as the largest fluctuations observed at the temperature of interest. At a temperature reasonably close to the Curie point the correlation length, in units of the lattice spacing, might be about 100 and the largest fluctuations would cover about 100^2, or 10,000, lattice sites. A block of spins that large has $2^{10,000}$ possible configurations, a number that is somewhat greater than $10^{3,000}$. The fastest computer conceivable could not carry out such a calculation. Even if the computer had been working continuously since the



NUMBER OF SPIN CONFIGURATIONS rises steeply as the size of a lattice grows. For a system of n spins, each of which has two possible values, the number of configurations is equal to 2^n. When the lattice is large, it becomes impractical to calculate the probability of all the configurations. The limit of practical computation is a lattice somewhat larger than the six-by-six array of 36 spins. In order to observe the critical behavior of the system near the Curie temperature an array of about 100-by-100 spins would be needed, which has $2^{10,000}$ configurations.

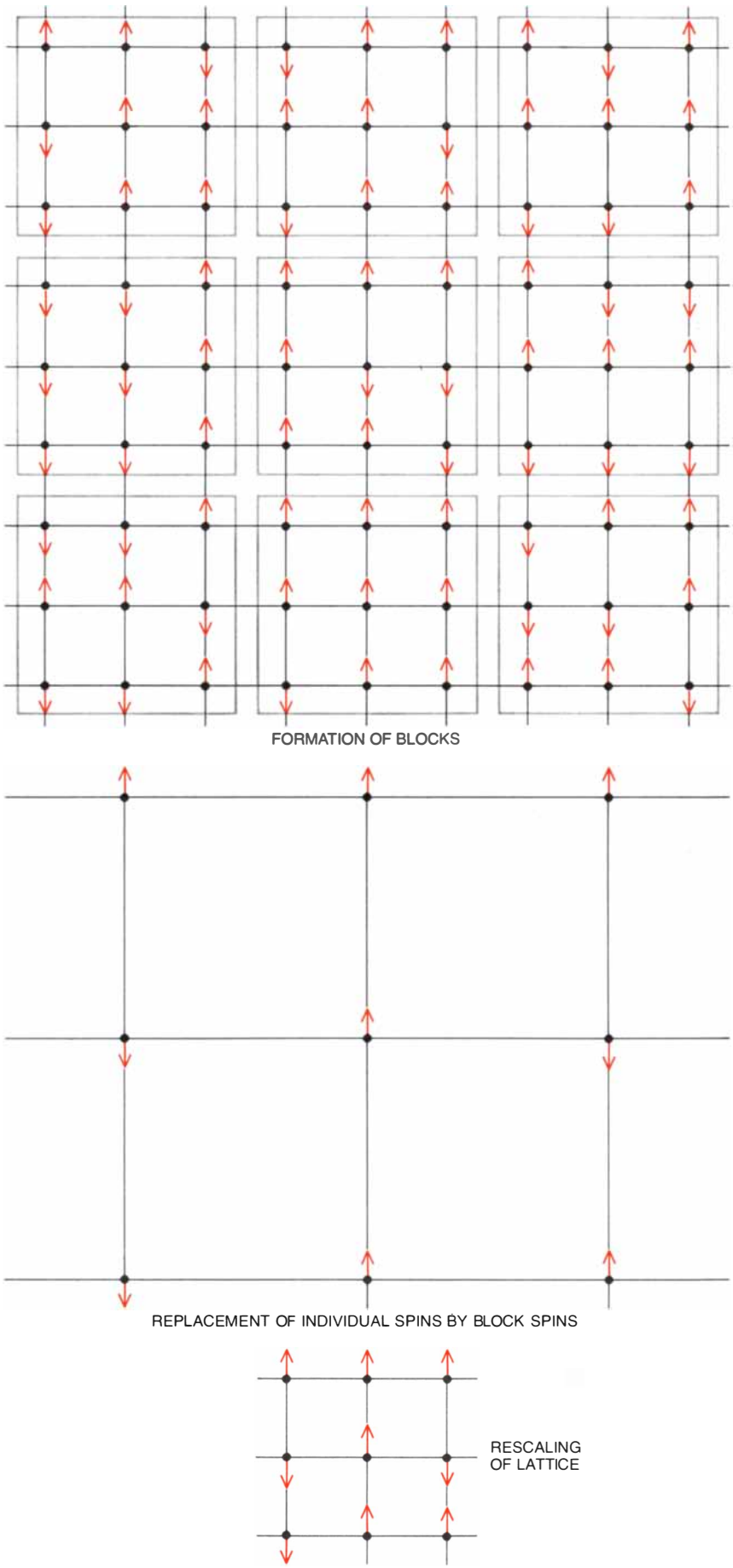
"big bang" with which the universe began, it would not yet have made a significant start on the task.

The need to carry out an almost endless enumeration of spin configurations can be circumvented for two special conditions of the lattice. When the temperature of the system is zero (so that the coupling strength is infinite), all but two of the configurations can be neglected. At zero temperature the probability that a pair of spins will be antiparallel falls to zero, and therefore so does the probability of any configuration that includes even one antiparallel pair. The only configurations that do not have at least one antiparallel pair are those in which all the spins are up or all are down. The lattice is certain to assume one of these configurations, and all other configurations have zero probability.

At infinite temperature, where the coupling strength is zero, the probability distribution is also much simplified. Every spin is then independent of its neighbors and its direction at any instant can be chosen at random. The result is that every configuration of the lattice has equal probability.

Through these two shortcuts to the determination of the probability distribution it is a trivial exercise to calculate exactly the properties of the Ising model at absolute zero and at infinite temperature. Acceptable methods of approximation are also available for any temperature low enough to be considered close to zero or high enough to be considered close to infinity. The troublesome region is between these extremes; it corresponds to the region of the critical point. Until recently there was no practical and direct method of calculating the properties of a system arbitrarily close to the critical point. The renormalization group provides such a method.

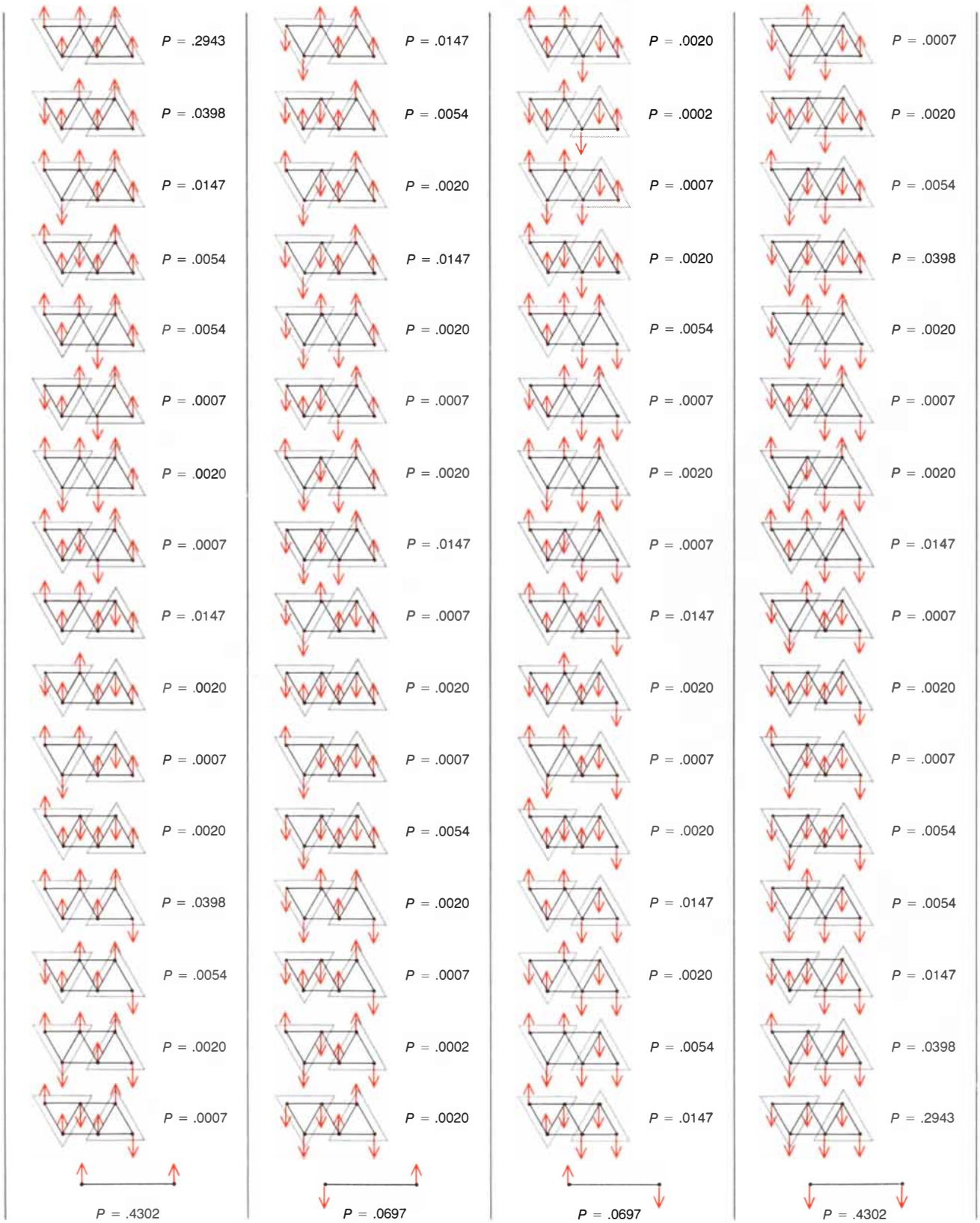
The essence of the renormalization-group method is to break a large problem down into a sequence of smaller and more manageable stages. Instead



RENORMALIZATION-GROUP approach to a model ferromagnet consists in breaking down an intractable problem with multiple scales of length into a sequence of smaller problems, each of which is confined to a single scale of length. One version of the renormalization-group method, called the block-spin transformation, has three steps. First the lattice is divided into blocks of a few spins each, in this case nine. Then each block is replaced by a single spin whose value is the average of all the spins in the block; here the average is determined by majority rule. In this way a new lattice is created, with three times the original lattice spacing and one-third the density of spins. Finally the original scale is restored by reducing all dimensions by a factor of 3. The procedure must be carried out for all configurations of a few spins in the original lattice, so that a probability can be found for every configuration of the block spins.

PROBABILITIES OF NEAREST-NEIGHBOR CONFIGURATIONS IN ORIGINAL LATTICE

PROBABILITIES OF SIX-SPIN CONFIGURATIONS IN ORIGINAL LATTICE



PROBABILITIES OF NEAREST-NEIGHBOR BLOCK-SPIN CONFIGURATIONS

of keeping track of all the spins in a region the size of the correlation length, the long-range properties are deduced from the behavior of a few quantities that incorporate the effects of many spins. There are several ways to do this. I shall describe one, the block-spin technique, in which the principles of the method are revealed with particular clarity. It was introduced by Leo P. Kadanoff of the University of Chicago and was made a practical tool for calculations by Th. Niemeijer and J. M. J. van Leeuwen of the Delft University of Technology in the Netherlands.

The method has three basic steps, each of which must be repeated many times. First the lattice is divided into blocks of a few spins each; I shall employ square blocks with three spins on a side, so that each block includes nine spins. Next all the spins in the block are averaged in some way and the entire block is replaced by a single new spin with the value of the average. Here the averaging can be done by a simple procedure: by following the principle of majority rule. If five or more of the original spins are up, the new spin is also up; otherwise it is down.

The result of these two operations is to create a new lattice whose fundamental spacing is three times as large as that of the old lattice. In the third step the original scale is restored by reducing all dimensions by a factor of 3.

These three steps define a renormalization-group transformation. Its effect is to eliminate from the system all fluctuations in spin direction whose scale is smaller than the block size. In the model given here any fluctuation of the spins over a range of fewer than three lattice units will be smeared out by the averaging of the spins in each block. It is as if one looked at the lattice through an out-of-focus lens, so that the smaller features are blurred but the larger ones are unaffected.

It is not enough to carry out this procedure for any one configuration of the original lattice; once again what is sought is a probability distribution. Suppose one considers only a small region of the initial lattice, consisting of 36 spins that can be arranged in four blocks. The spins in this region have 2^{36}, or about 70 billion, possible configurations. After the block-spin transformation has been applied the 36 original spins are replaced by four block spins with a total of 16 configurations. It is just within the limit of practicality to compute the probability of each of the configurations of the original 36 spins. From those numbers the probabilities of the 16 block-spin configurations can readily be determined. The calculation can be done by sorting all the configurations of the original lattice into 16 classes according to which configuration of the block spins results in each case from applying the principle of majority rule. The total probability for any one configuration of the block spins is then found by adding up the probabilities of all the configurations of the original lattice that fall into that class.

It may well seem that nothing is gained by this procedure. If the complete probability distribution can be calculated for a system of 36 spins, nothing new is learned by condensing that system into a smaller lattice of four block spins. Near the critical point it is still necessary to consider a much larger lattice, with perhaps 10,000 spins instead of 36, and the probability distribution for the block spins generated from this lattice cannot be calculated because there are far too many configurations. As it turns out, however, there is a method for extracting useful information from a small set of block spins. It is a method for observing the behavior of the system over a large region without ever dealing explicitly with the configurations of all the spins in that region.

Each block spin represents nine spins in the original lattice. The complete set of block spins, however, can also be regarded as a spin system in its own right, with properties that can be investigated by the same methods that are applied to the original model. It can be assumed that there are couplings between the block spins, which depend on the temperature and which determine in turn the probability of each possible spin configuration. An initial guess might be that the couplings between block spins are the same ones specified in the original lattice of Ising spins, namely a nearest-neighbor interaction with a strength given by the parameter K, the reciprocal of the temperature.

This guess can easily be checked, because the probability distribution for the configurations of at least a small part of the block-spin system is already known; it was computed from the configurations of the original lattice in the course of defining the block spins. Surprisingly, this hypothesis is generally wrong: the block spins do not have the same couplings as the spins in the original model. Assuming that only adjacent sites interact and that they have a coupling strength equal to K gives the wrong set of probabilities for the configurations of the block spins.

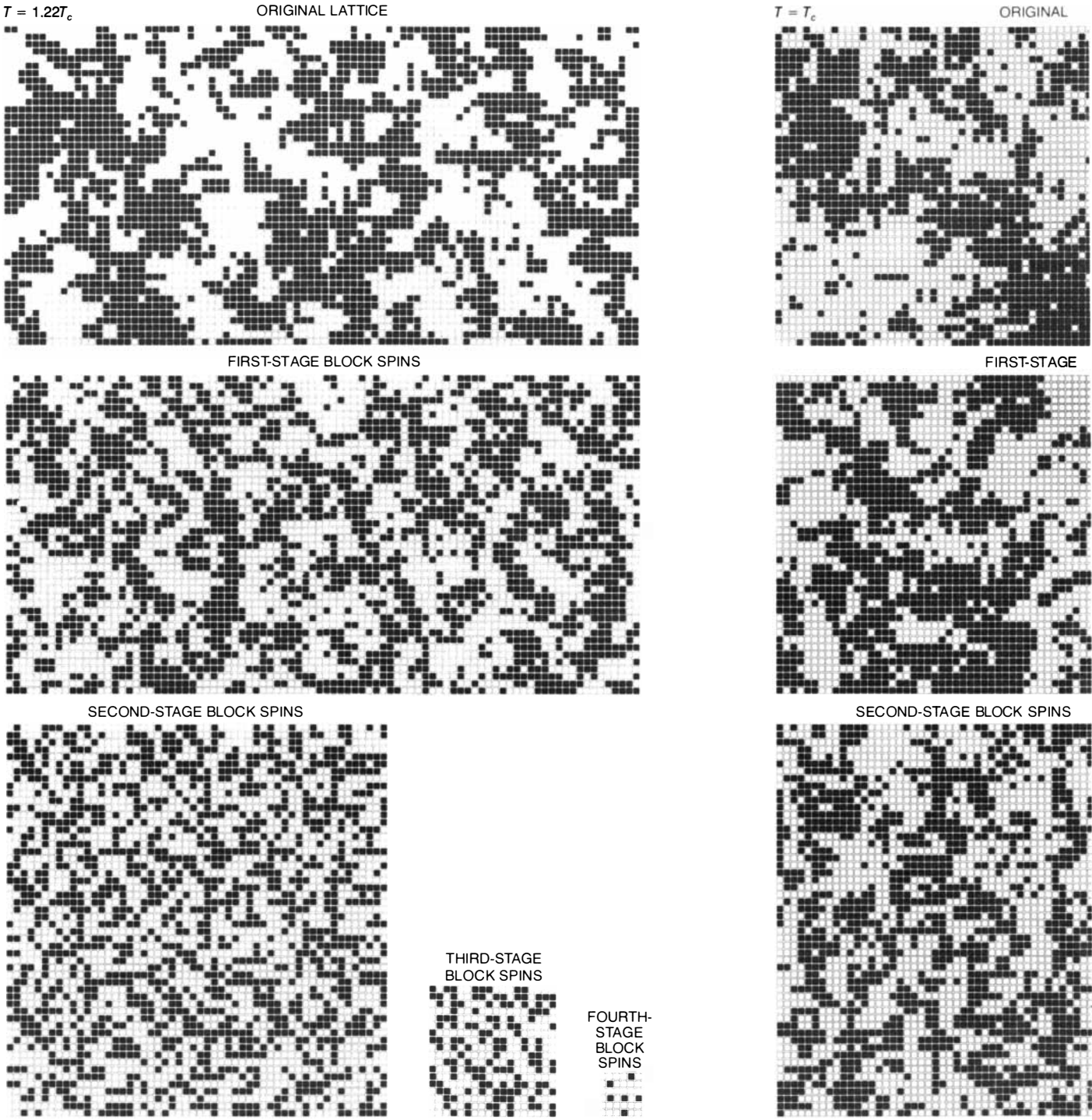
If the specifications of the original model will not describe the system of block spins, then some new set of couplings must be invented. The guiding principle in formulating these new interactions is to reproduce as accurately as possible the observed probability distribution. In general the nearest-neighbor coupling strength must be changed, that is, K must take on a new value. What is more, couplings of longer range, which were excluded by definition from the Ising model, must be introduced. For example, it may be necessary to establish a coupling between spins at the opposite corners of a square. There might also be direct interactions among spins taken three at a time or four at a time. Couplings of still longer range are possible. Hence the block spins can be regarded as a lattice system, but it is a system quite different from the original one. Notably, because the basic couplings have different values, the lattice of block spins is at a temperature different from that of the initial Ising system.

Once a set of couplings has been found that correctly describes the probability distribution for the block spins, a lattice of arbitrary size can be constructed from them. The new lattice is formed the same way the original one was, but now the probability for the spin at each site is determined by the newly derived coupling strengths rather than by the single coupling of the Ising model. The renormalization-group calculation now proceeds by starting all over again, with the new system of block spins as the starting lattice. Once again blocks of nine spins each are formed, and in some small region, such as an array of 36 spins, the probability of every possible configuration is found. This calculation is then employed to define the probability distribution of a second generation of block spins, which are once more formed by majority rule. Examination of the second-generation block spins shows that the couplings have again changed, so that new values must be supplied a second time for each coupling strength. Once the new values have been determined another lattice system (the third generation) can be construct-

PROBABILITY DISTRIBUTION for a system of block spins is found by adding up the probabilities for all the configurations of the original lattice that contribute to each configuration of the block spins. The calculation is shown for a system of six spins on a triangular lattice. Two blocks of three spins each are formed from the lattice, and each block is replaced by a single spin whose orientation is determined by majority rule. The six spins have 64 possible configurations, which are assigned to columns in such a way that all the configurations in each column give rise to the same block-spin configuration. For example, all the configurations in the column at the far left have at least two spins in each block pointing up, so that they are represented by two up block spins. The coupling strength in the original lattice is set equal to .5, which yields the nearest-neighbor probabilities shown at the top of the page. From this set of numbers a probability is calculated for every configuration of the original lattice; then all the probabilities in each column are added up to give the probability of the corresponding block-spin configuration. The block-spin probabilities are not the same as those specified for the original lattice, which implies that the coupling strength is also different, as is the temperature.

ed and the entire procedure can be repeated yet again.

The point of this repetitive operation is that it provides information about the behavior of distinct but related spin systems in which the fundamental scale of length gets larger with each iteration. After the first block-spin transformation the fluctuations at the smallest scale have been eliminated, but those slightly larger, with a scale of roughly three times the original lattice spacing, can be seen more clearly. After the second transformation each block spin represents the 81 spins in a nine-by-nine block of the original lattice, and all fluctuations up to this size range are averaged out, leaving only those larger than nine lattice units. The next iteration removes all fluctuations whose scale is between nine and 27 lattice units, then the following iteration removes those between 27 and 81 units. Eventually fluctuations at all scales up to the correlation length

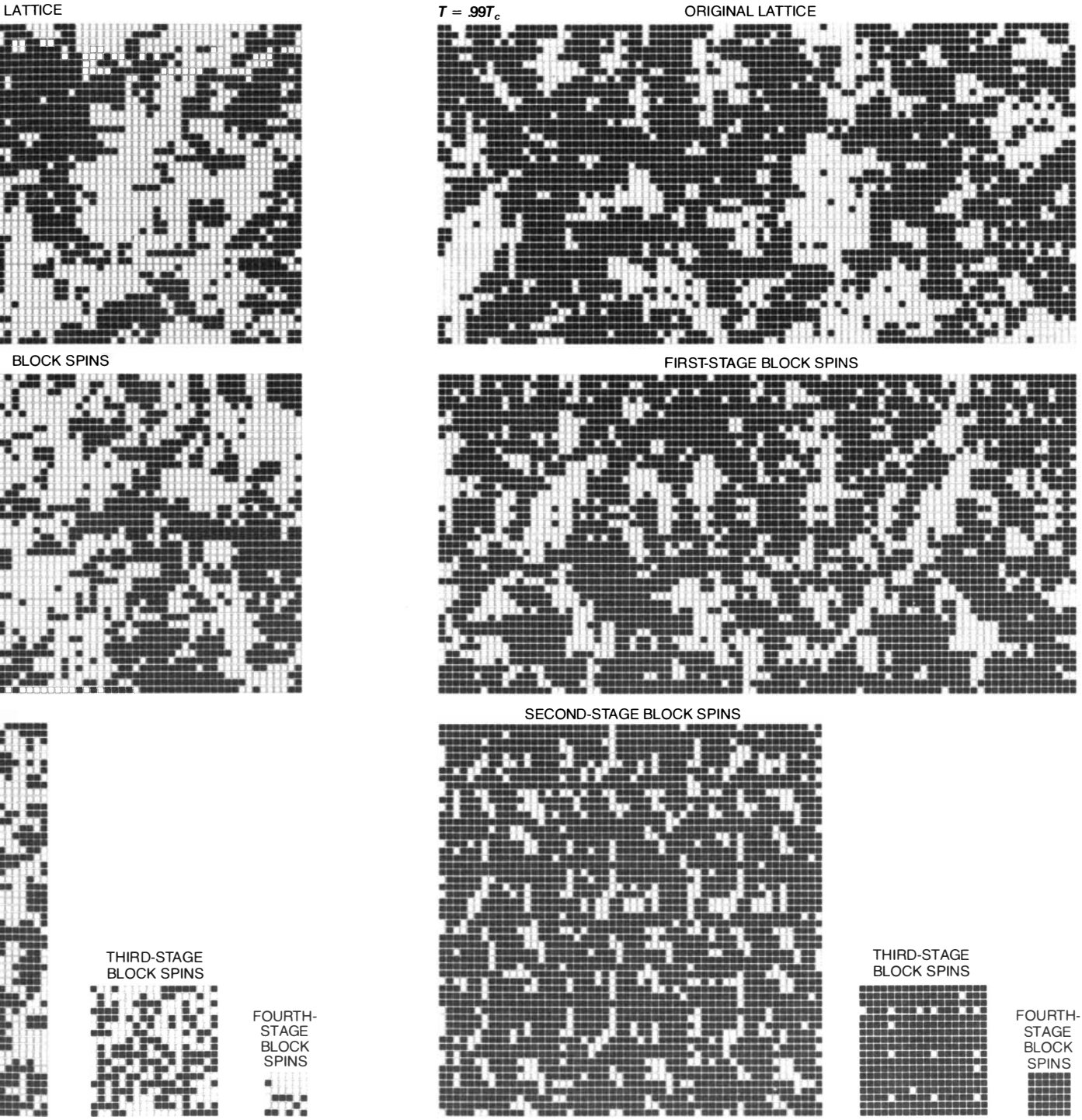


BLOCK-SPIN TRANSFORMATION is applied to a lattice of spins repeatedly, each time elucidating the behavior of the system at a larger scale. The computer simulation, which was carried out by the author, began with an array of some 236,000 spins; a black square represents an up spin and an open square a down spin. The initial temperature was set equal to three values: above the Curie temperature, T_c, at T_c and just below T_c. The transformation begins with the division of the original lattice into three-by-three blocks. Each block is replaced by a single spin whose value is determined by majority rule; these make up the lattice of first-stage block spins. The procedure is then repeated, but with the first-stage block spins serving as the starting lattice. The resulting second-stage spins form the initial configuration for the next transformation, and so on. By the time the third stage is reached the number of spins is small enough for them all

are averaged out. The resulting spin system reflects only the long-range properties of the original Ising system, with all finer-scale fluctuations eliminated.

The value of the block-spin technique can be perceived even through a simple visual inspection of the evolving model. Merely looking at a configuration of Ising spins just below the Curie temperature will seldom reveal that the model is slightly magnetized. At this temperature there is only a small excess of one spin direction over the other, and the many small-scale fluctuations obscure the overall bias. After several applications of the block-spin transformation, however, the smaller fluctuations disappear and the long-range magnetization becomes obvious.

Much of the physical meaning of the block-spin transformation is to be found in the way the couplings between spins change. The rules for deriving the new couplings from the old ones at each



to be shown, and after the fourth stage there are only 36 spins left, each one representing more than 6,000 sites in the original lattice. In the first stage any fluctuations whose scale of length is smaller than three lattice units are eliminated by the averaging procedure. The second stage removes the fluctuations between three and nine lattice units, the third stage those between nine and 27 units, and so on. When the initial temperature is above T_c, the spins become more nearly random in appearance with each iteration and large-scale fluctuations disappear; when the temperature is below T_c, the spins become more nearly uniform and what fluctuations remain are small in scale. When the starting temperature is exactly equal to T_c, large-scale fluctuations remain at all stages. Because the block-spin transformation leaves the large-scale structure of the lattice unchanged at the Curie temperature, a system at that temperature is said to be at a fixed point.

stage are often complicated, but the effect of the change can be illustrated by a quite simple example. Although the assumptions are not realistic, I shall discuss a model in which no couplings with a range longer than the original, nearest-neighbor interaction are introduced. The only change in the coupling is an adjustment in the value of K, which is equivalent to a shift in the temperature. Moreover, this adjustment in K will have a simple form: at each stage in the procedure the coupling strength in the new lattice will be set equal to the square of the coupling in the old lattice. If the new coupling is denoted K', it is given by the equation $K' = K^2$.

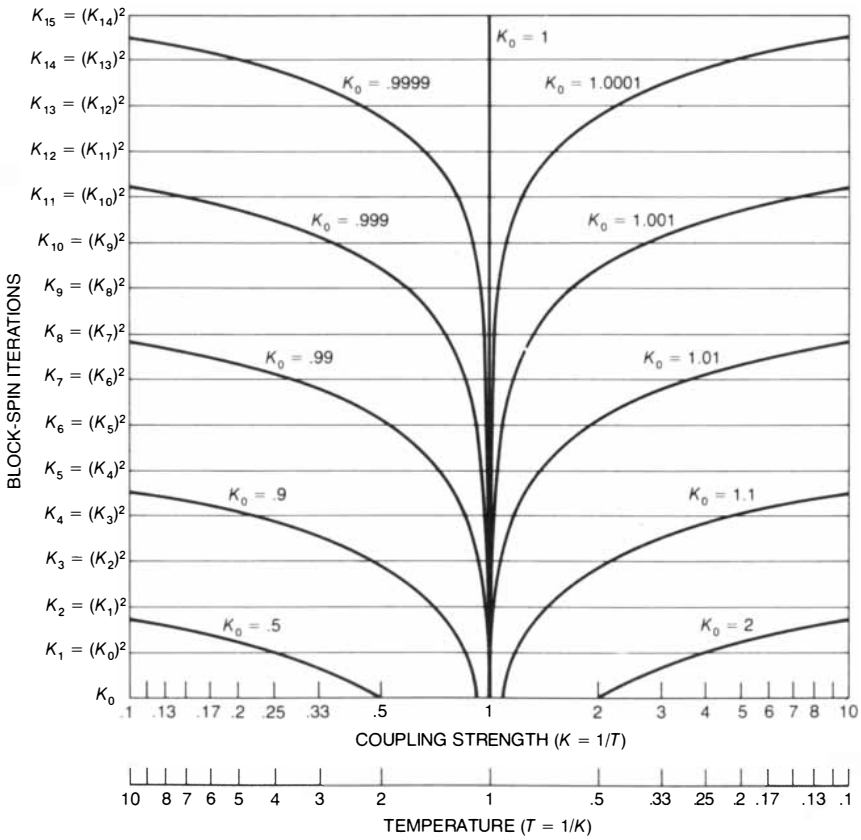
Suppose in some initial state K is equal to 1/2 (which means that the temperature has been given an initial value of 2 in the arbitrary units employed here). In the thinned-out lattice formed as a product of the block-spin transformation K will be replaced by K', with a value of $(1/2)^2$, or 1/4. Repeating the transformation yields successive values of 1/16, 1/256 and so on, in a series that rapidly approaches zero. With each iteration the spin system is converted into a new system that not only has a thinner lattice but also has weaker couplings between the spins. Since K is equal to $1/T$, the temperature increases with each iteration and the lattice approaches the limit of infinite temperature and random spins.

If the initial coupling strength is set equal to 2 (so that the temperature has a value of 1/2), the coupling increases at each stage in the calculation. After the first block-spin transformation the coupling strength is 4, then 16, then 256; ultimately the strength becomes infinite. At the same time, of course, the temperature falls and the system approaches the state of zero temperature, in which all the spins are aligned.

It should be emphasized that what is being observed is not the evolution of any single spin system as the temperature changes. Nothing is being heated or cooled. Instead a new spin system is being created at each stage, a system distinguished by a different set of couplings between spins. The large-scale or long-range behavior of the new lattice is equivalent to the behavior that would be observed in the original lattice at a different temperature.

There is one initial value of K that does not diverge either to infinity or to zero, namely the value $K = 1$. Since 1^2 is equal to 1, K' remains equal to K no matter how many times the transformation is repeated. When K is equal to 1, the system is said to be at a fixed point, where continued application of the renormalization-group transformation leaves all essential properties of the lattice unchanged. Actually the values $K = 0$ and $K =$ infinity also represent fixed points, since zero squared is still zero and infinity squared is still infinity. Zero and infinity, however, are considered trivial fixed points, whereas the value $K = 1$ corresponds to the critical point.

In this discussion of the block-spin technique all the effects of the transformation have been expressed through a single parameter: the nearest-neighbor coupling strength K. Actually many other parameters are introduced by the transformation, each one corresponding to a longer-range coupling. All the possible combinations of these parameters can be represented geometrically by constructing an imaginary multidimensional space in which distance measured along each dimension corresponds to variation in one of the parameters. Every initial state of the spin system and every block-spin transformation of it can be represented by a point on a surface somewhere in this parameter space.



CHANGE IN THE COUPLING BETWEEN SPINS is part of the renormalization-group transformation. The adjustment that must be made to the coupling strength with each iteration can take many forms, but a simple example is presented here: If the strength of the coupling in the original lattice is given by the number K, then in the new lattice the coupling strength is equal to K^2. Any initial value of K greater than 1 must approach infinity when K is squared repeatedly; any value less than 1 must approach zero. The special value $K = 1$ remains unchanged no matter how many times the transformation is repeated. Because the temperature can be defined (in appropriate units) as the reciprocal of the coupling strength, the renormalization-group transformation can be seen as establishing a correspondence between the original lattice and a new, thinned-out lattice that will generally have a different coupling strength and a different temperature. It is only at the fixed point, which corresponds to the Curie temperature, that the coupling and the temperature remain invariant with a value of 1.

In the geometric description of the renormalization-group method the significance of the fixed points becomes apparent. For the two-dimensional Ising system the surface in parameter space has the form of a hilly landscape with two sharp peaks and two deep sinkholes. The ridgeline that connects the peaks and the gully line that connects the sinkholes meet in the center at a saddle point [*see illustration on opposite page*]. One sinkhole is the $K = 0$ fixed point; the other is the $K =$ infinity fixed point. The critical fixed point lies at the point of precarious equilibrium in the saddle.

The transformation of the system from one state to the next can be represented by the motion of a marble rolling on the surface. One can imagine a time-lapse motion picture that would record the marble's position at one-second intervals; then each frame would reveal the effect of one iteration of the block-spin transformation. It is the transformation that allows the marble to move, but the speed and direction of the marble are determined entirely by the slope of the surface at each point it crosses.

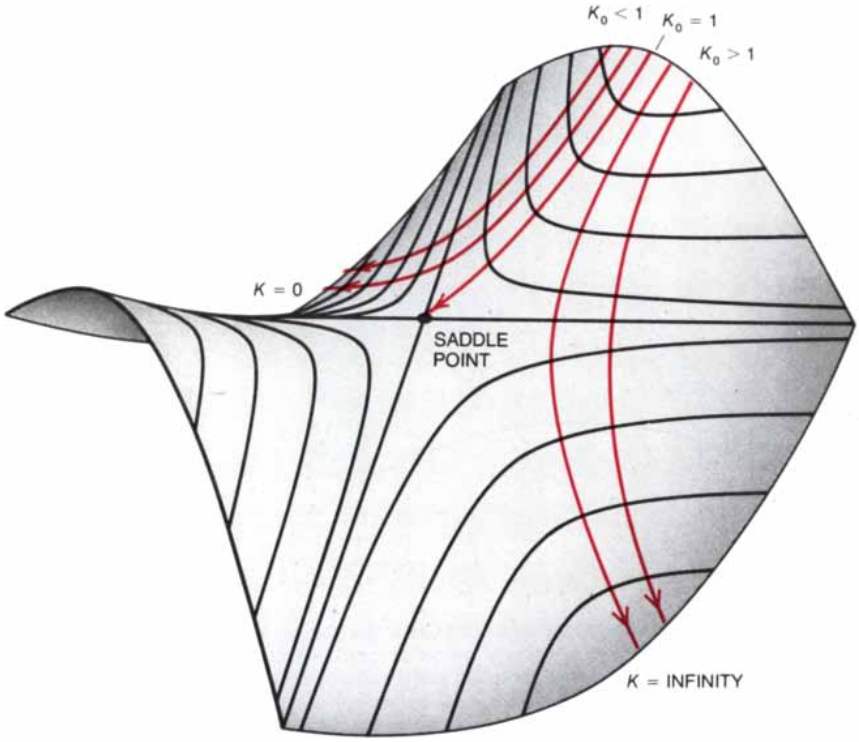
Suppose the marble is initially placed near the top of a hill and just to one side of the ridgeline. At first it moves rapidly,

because the hill is steep near the top, and proceeds in the general direction of the saddle point. As the marble approaches the saddle the slope becomes more gradual and the marble slows down, but it never comes to a complete stop. Moreover, because it started to one side of the ridgeline it does not quite reach the saddle point; instead it is deflected to one side and begins to accelerate again, this time toward a sinkhole.

The trajectory of the marble describes the path followed by the point representing a system of Ising spins as it is transformed repeatedly by the block-spin method. The initial position just off the ridgeline corresponds to an initial value of the various coupling parameters that is equivalent to a temperature just above or just below the critical temperature. In terms of the simplified example described above, with just one parameter, the value of K is either slightly greater than 1 or less than 1. Setting the coupling strength equal to 1 is equivalent to placing the marble exactly on the ridgeline. It then moves directly toward the saddle point, or critical fixed point. Again the motion is rapid at first but becomes slower as the saddle is approached. In this case, however, the marble remains balanced between the two descending slopes. Even after a large number of iterations it remains at the fixed point.

A trajectory on the saddle-shaped surface in parameter space can be made to approach the fixed point as closely as is wished by setting the initial value of K sufficiently close to the critical value. In the example considered here, where the critical value of K is 1, the initial value of K might be .9999, which can be squared several times before it changes appreciably. As a result the trajectory comes quite close to the critical fixed point before it veers off toward the high-temperature sinkhole.

By the examination of many such trajectories the topography of the surface itself can be mapped in the small region surrounding the saddle point. The slope of the surface is what determines how the system approaches the fixed point and how it departs from it. Knowing the slope, then, one can calculate how the properties of the system vary as the initial coupling and the initial temperature are changed. That is precisely the information sought for an understanding of critical phenomena.



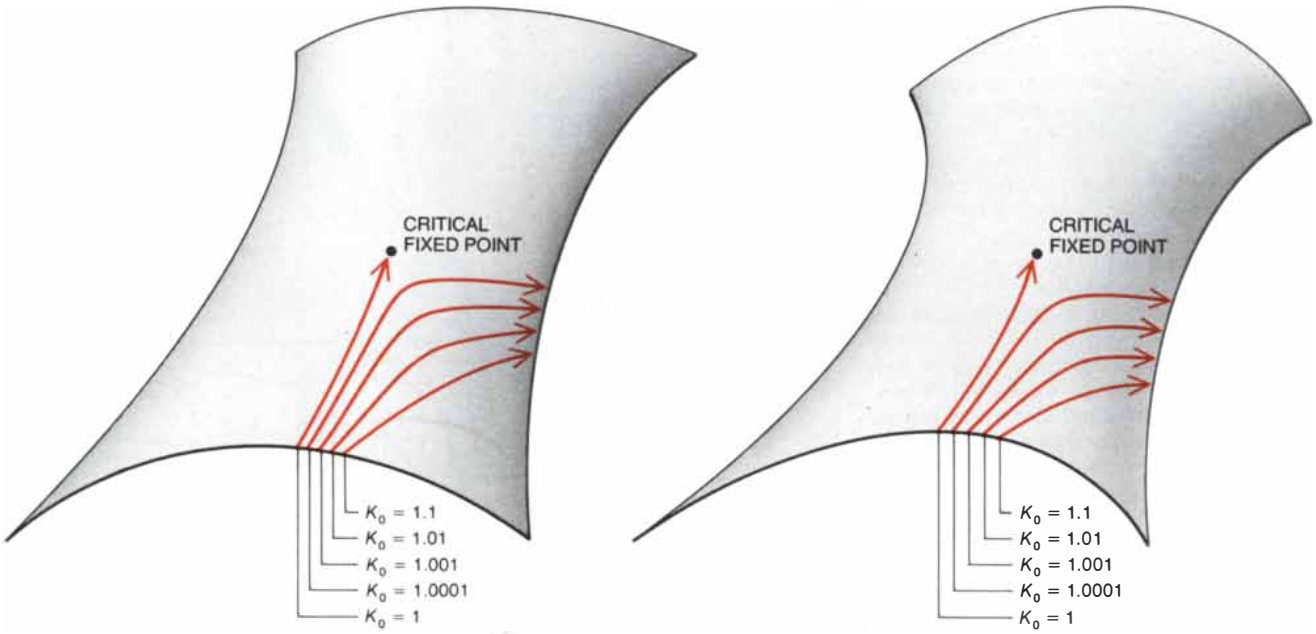
EVOLUTION OF A SPIN SYSTEM in response to repeated renormalization-group transformations can be described as the motion of a point on a surface constructed in an imaginary, multidimensional space: the parameter space. The form of the surface is defined by all the couplings between block spins, but only the nearest-neighbor coupling, K, is considered here. The surface has two peaks and two sinkholes, which are connected to a saddle point. The trajectory followed by the point that represents the state of the system is determined entirely by the slope of the surface. An initial value of K slightly greater than 1 corresponds to an initial position slightly to one side of the ridgeline that connects the peaks. After several block-spin transformations the point rolls down the hill, passes near the saddle point and veers off into one of the sinkholes, where K tends toward infinity. An initial value of K slightly less than 1 leads to a similar trajectory on the other side of the ridgeline and terminates in the other sinkhole, where K approaches zero. When K is equal to exactly 1, the point remains permanently on the ridgeline, approaching equilibrium at the saddle point. Both of the sinkholes are fixed points (since the values of $K = 0$ and $K =$ infinity do not change with further renormalization-group transformations), but they are considered trivial fixed points. The saddle defines the critical fixed point.

The macroscopic properties of a thermodynamic system near the critical point are determined by the temperature. To be more precise, properties such as the spontaneous magnetization, the susceptibility and the correlation length are functions of the amount by which the temperature of the system departs from the critical temperature, T_c. For this reason it is convenient to define the temperature in such a way that all critical points are equivalent. A suitable quantity is the reduced temperature, t, defined as the difference between the actual temperature and the critical temperature, divided by the critical temperature; thus t is equal to $T - T_c/T_c$. On an ordinary temperature scale such as the Kelvin scale the critical temperatures of different systems fall at different values, but all critical points have the same reduced temperature, namely zero.

All critical properties are proportional to the absolute value of the reduced temperature raised to some power. The problem of describing critical phenomena is to determine what that power is, or in other words to determine the values of the critical exponents. For example, the magnetization, M, of a spin system is given by the proportionality $M \quad |t|^{\beta}$, where β (the Greek letter beta) is a critical exponent and where the vertical lines designate the absolute value of t. The magnetic susceptibility is proportional to $1/|t|^{\gamma}$, where γ (the Greek letter gamma) is another exponent. The correlation length is associated with a third exponent, ν (the Greek letter nu), in a relation of the same form: the length is proportional to $1/|t|^{\nu}$.

The earliest attempts to formulate a mathematical description of critical phenomena were theories of a kind that are now called mean-field theories. The first of these was introduced in 1873 by J. D. van der Waals as an explanation of phase changes in fluids. A theory of magnetic phase transitions was proposed in 1907 by Pierre Weiss. In 1937 L. D. Landau of the Academy of Sciences of the U.S.S.R. proposed a more general formulation of mean-field theory, thereby providing a framework in which many physical systems could be discussed. In all these theories the state of any selected particle is determined by the average properties of the material as a whole, properties such as the net magnetization. In effect all particles in the system contribute equally to the force at every site, which is equivalent to assuming that the forces have infinite range.

Mean-field theories are qualitatively successful. They account for important



SLOPE OF THE PARAMETER SURFACE in the vicinity of the critical fixed point determines the macroscopic properties of the Ising model. If trajectories are plotted for many initial values of K near the critical value (which in this case is $K = 1$), it is the slope at the saddle point that determines how quickly the trajectories veer off toward the trivial fixed points at $K = 0$ and $K =$ infinity. If the surface is comparatively flat (*left*), a trajectory with an initial value of K such as $K = 1.01$ passes close to the saddle point. When the surface is more steeply curved (*right*), the corresponding trajectory bends more abruptly toward the sinkhole. Because the temperature is the reciprocal of K the slope near the fixed point reveals how the properties of the system change as the temperature departs from the critical temperature.

features of the phase diagrams of fluids and ferromagnets, the most notable of these features being the existence of a critical point. The quantitative predictions, however, are less satisfactory: the theories give the wrong values for the critical exponents. For β, the exponent that governs the spontaneous magnetization, mean-field theory implies a value of 1/2; in other words, the magnetization varies as the square root of the reduced temperature. The exponent associated with the susceptibility, γ, is assigned a value of 1, so that the susceptibility is proportional to $1/|t|$. The exponent for the correlation length, ν, is 1/2, so that this quantity is proportional to $1/\sqrt{|t|}$.

The exponents calculated from mean-field theory suggest a plausible form for each of these functions. The magnetization has two possible values ($+\sqrt{t}$ and $-\sqrt{t}$) at all temperatures below the critical point, and then it vanishes above the critical temperature. Both the susceptibility and the correlation length approach infinity as t nears zero from either above or below. The actual values of the mean-field exponents, however, are known to be wrong.

For the two-dimensional Ising model the critical exponents are known exactly from Onsager's solution. The correct values are $\beta = 1/8$, $\gamma = 7/4$ and $\nu = 1$, which differ significantly from the predictions of mean-field theory and imply that the system has rather different behavior. For example, the magnetization is proportional not to the square root of the reduced temperature t but to the eighth root of t. Similarly, the susceptibility is given by the reciprocal not of t but of t raised to the 1.75th power, which makes the divergence near the critical point steeper and more abrupt.

The reason for the quantitative failure of mean-field theories is not hard to identify. The infinite range assigned to the forces is not even a good approximation to the truth. Not all spins make equal contributions; the nearest neighbors are more important by far than any other spins. The same objection can be expressed another way: the theories fail to take any notice of fluctuations in spin orientation or in fluid density.

In a renormalization-group calculation the critical exponents are determined from the slope of the parameter surface in the vicinity of the fixed point. A slope is simply a graphic representation of a rate of change; the slope near the fixed point determines the rate at which the properties of the system change as the temperature (or the coupling strength) is varied over some narrow range near the critical temperature. Describing the change in the system as a function of temperature is also the role of the critical exponents, and so it is reasonable that there is a connection between the exponents and the slope.

Renormalization-group calculations for the two-dimensional Ising system have been carried out by several workers. In 1973 Niemeijer and van Leeuwen employed a block-spin method to study the properties of a system of Ising spins constructed on a triangular lattice. I have applied a somewhat different renormalization-group technique, called spin decimation, to a square lattice. In spin decimation, instead of assembling blocks of a few spins each, every other spin in the lattice is held fixed while a probability distribution is computed for the remaining spins. These calculations were much more elaborate than the model calculation described here; in my own work, for example, 217 couplings between spins were included. The critical exponents derived from the calculation agree with Onsager's values to within about .2 percent.

Because an exact solution is known for the two-dimensional Ising model, the application of the renormalization group to it is something of an academic exercise. For a system of Ising spins in a three-dimensional lattice, however, no exact solution is known. A method has been devised, by Cyril Domb of University College London and many others, for finding approximate values of the exponents in the three-dimensional case. First the properties of the system at high temperature are determined with great precision, then these properties are extrapolated to the critical temperature. The best results obtained so far by this method give values for the exponents of $\beta = .33$, $\gamma = 1.25$ and $\nu = .63$.

Although extrapolation from a high-temperature solution leads to good approximations for the critical exponents, it provides little intuitive understanding of how the system behaves near the critical point. A renormalization-group calculation gives essentially the same values for the exponents, but it also explains important universal features of critical behavior.

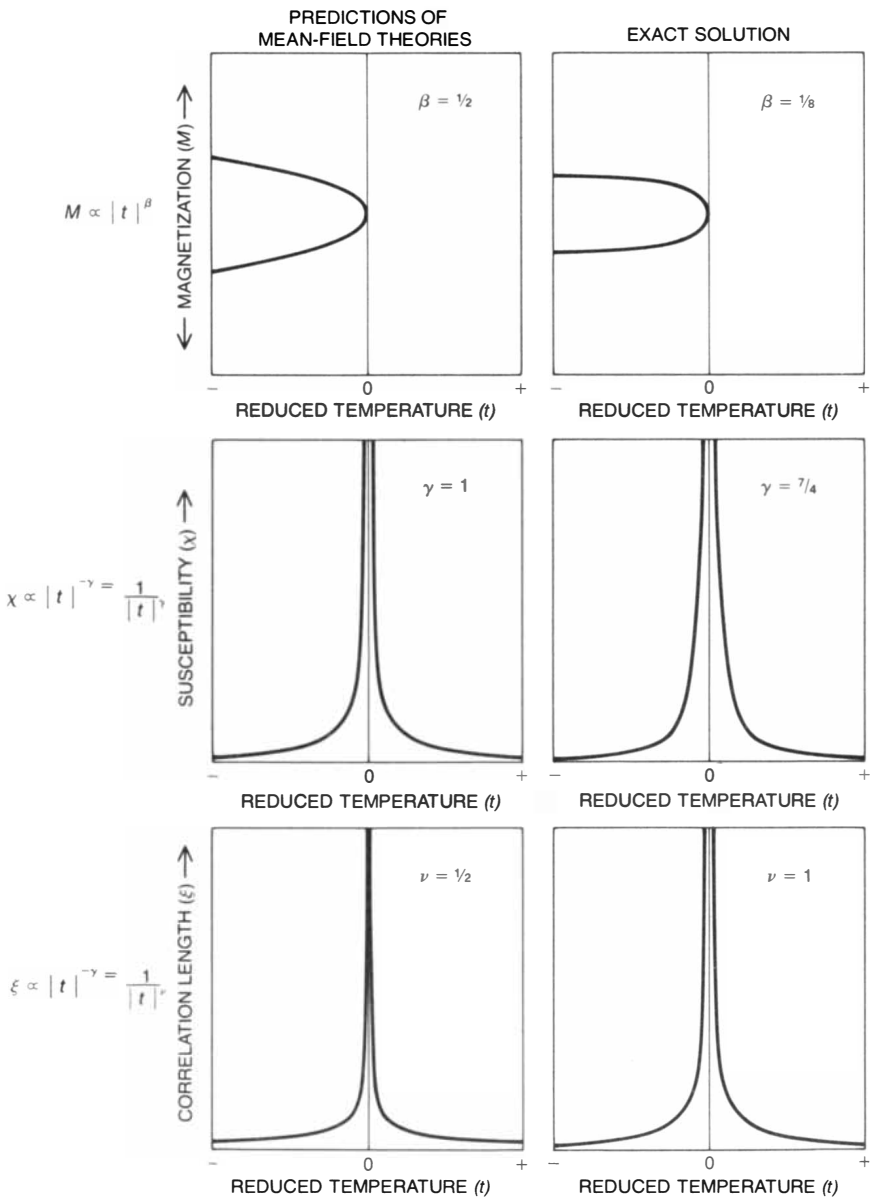
Two remarkable facts about the exponents in the three-dimensional Ising model should not be overlooked. The first is simply that the values are different from those for the two-dimensional model. In mean-field theories the dimensionality of space does not enter the calculations and so the critical exponents have the same values in any space. The second surprise is that the exponents are not integers or ratios of small integers, as they are in mean-field theories. They may even be irrational numbers.

If it is surprising that the spatial dimensionality influences the critical exponents, it is equally remarkable that certain other properties of the model have no effect at all. An example of such an irrelevant parameter is the structure of the lattice. In the two-dimensional Ising model it makes no difference whether the lattice is rectilinear, as in my own work, or triangular, as in the model employed by Niemeijer and van Leeuwen; the critical exponents are the same. By extension, in a real ferromagnet the great variety of crystal structures all yield identical critical behavior.

There is an intuitive justification for the irrelevance of the lattice structure and of other microscopic properties. A change in the form of the lattice has a large effect on events at the scale of the lattice spacing, but the effect diminishes as the scale of interest increases. In a renormalization-group calculation the fluctuations at the scale of the lattice spacing are averaged out after the first few iterations, and so models with many different lattices have the same critical behavior. Through the renormalization group the appearance of the same critical exponents in many systems is seen to result from the topography of the surface in parameter space. Each lattice structure corresponds to a different position in parameter space, but at the critical temperature every lattice is represented by a point somewhere along the ridgeline. After repeated renormalization-group transformations all these systems converge on the same fixed point, namely the saddle point.

The idea that certain variables are irrelevant to critical phenomena can be extended to systems other than ferromagnets. A fluid near its critical point, for example, has the same properties as the three-dimensional Ising model of a ferromagnet. In order for this identity to be understood some correspondence must be established between the macroscopic properties of the fluid and those of the magnet. The magnetization, which is the number of up spins minus the number of down spins, can be identified with the density difference in the fluid: the density of the liquid phase minus the density of the vapor phase. Just as the magnetization vanishes at the Curie temperature, so the density difference falls to zero at the critical point of the fluid. These quantities—the magnetization and the density difference—are called the order parameters of their respective systems. The susceptibility of the magnet, which is the change in magnetization for a given small change in the applied magnetic field, is analogous to the compressibility of the fluid: the change in density that results from a given small change in pressure. Like the susceptibility, the compressibility becomes infinite at the critical point. The critical behavior of the fluid and that of the three-dimensional Ising model are identical in that they have the same surface in parameter space. The two systems have different initial positions on the surface, but they converge on the same saddle point and hence have the same critical exponents.

The similarity observed in the critical behavior of fluids and of ferromagnets is an instance of a more general hypothesis called critical-point univer-



CRITICAL EXPONENTS express the dependence of macroscopic properties on the extent to which the temperature of the system departs from the critical temperature. The temperature is most conveniently given in the form of the reduced temperature, t, defined by the equation $t = T - T_c / T_c$. All macroscopic properties are then proportional to the absolute value of t raised to some power; the power is the critical exponent for that property. The exponents and power laws in the graphs at the left are those predicted by mean-field theories, which ignore all fluctuations. The exponents in the graphs at the right are derived from an exact solution of the two-dimensional Ising model reported in 1944 by Lars Onsager of Yale University. The exponents show how the properties of the system change as the temperature or the coupling strength is changed; that is the same information conveyed by the slope of the surface in parameter space near the critical fixed point. The exponents can be determined from the slope, and calculations by the author and others for the two-dimensional Ising model give values close to Onsager's.

sality. According to the hypothesis, only two quantities determine the critical behavior of most systems: the dimensionality of space and the dimensionality of the order parameter. These quantities are labeled respectively d and n. All systems that have the same values of d and n are thought to have the same surface in parameter space and the same critical exponents. They are said to be members of the same universality class.

The dimensionality of space is seldom difficult to determine, but the dimensionality of the order parameter requires more careful consideration. In magnetic systems, where the order parameter is the magnetization, n is the number of components needed to define the spin vector. The vector of an Ising spin can be oriented only along a single axis, and so it has only one component; for the Ising model n is equal to 1. A spin vector that is allowed to point anywhere in a plane has two components, which are customarily drawn along the two axes that define a plane. Similarly, a vector that can point anywhere in three-dimensional space has three components, so that n equals 3.

For the three-dimensional Ising model d equals 3 and n equals 1. Ordinary fluids belong to the same universality class. The space in which the fluid exists clearly has three dimensions. The order parameter—the difference in density between the liquid and the vapor phases—is a quantity that has only a magnitude and hence only one component; it can be expressed as a single number, just as the value of an Ising spin can be.

Several other physical systems are members of this class. A mixture of two liquids such as oil and water exhibits critical behavior near the temperature where the component fluids become completely miscible in each other, a temperature called the consolute point. At temperatures below the consolute point the mixture separates into two phases, and the order parameter is defined as the concentration difference between the two phases, another quantity that can be expressed as a single number. Alloys such as brass have a transition between an ordered phase, where the two metals occupy alternate sites in a regular lattice, and a disordered phase, where their distribution is less uniform. The order parameter in this system is again a concentration difference, so that n equals 1. All these systems are expected to have the same critical exponents as the three-dimensional Ising model. So are some real ferromagnets, those that are easily magnetized only along a single axis. The available experimental evidence confirms these predictions.

The universality hypothesis would be trivial if the critical exponents had the values of integers or simple fractions such as 1/2. Many physical laws share such exponents, and there is no compelling reason for postulating a connection between them. Gravitation and electromagnetism both have an inverse-square law (an exponent of -2), but that coincidence does not demonstrate that the two forces are identical. The correspondence of exponents does seem remarkable, however, when the values are not round numbers but fractions such as .63. The convergence of many systems on these values cannot be coincidental. It is evidence that all the details of physical structure distinguishing a fluid from a magnet are less important than the geometric properties expressed by the values of d and n.

The two-dimensional Ising model ($d = 2$, $n = 1$) typifies a class of systems that are confined to two-dimensional space. One example is a thin film of liquid; another is a gas adsorbed on a solid surface. An ordinary ferromagnet falls into the class with $d = 3$ and $n = 3$, that is, the lattice is three-dimensional and each spin has three components, so that it can point in any direction. When the spins are constrained to lie in a plane, the class is reduced to $d = 3$ and $n = 2$. In this same class are the superfluid transition of liquid helium 4 and the superconducting transitions of various metals.

Other universality classes have values of d and n whose interpretation is somewhat less obvious. The case of $d = 4$ is of interest in the physics of elementary particles, where one of the four spatial dimensions corresponds to the axis of time. In a theoretical lattice of spins called the spherical model, where an individual spin can have any magnitude and only the total of all the spins is constrained, n is effectively infinite. A self-avoiding random walk through a lattice of points, or in other words a random walk that never occupies the same lattice site more than once, describes the folding up in space of a long-chain polymer; Pierre Gilles de Gennes of the Collège de France has shown that this problem belongs to a universality class with n equal to zero. In theoretical models n

UNIVERSALITY CLASS		THEORETICAL MODEL	PHYSICAL SYSTEM	ORDER PARAMETER
$d = 2$	$n = 1$	Ising model in two dimensions	Adsorbed films	Surface density
	$n = 2$	XY model in two dimensions	Helium-4 films	Amplitude of superfluid phase
	$n = 3$	Heisenberg model in two dimensions		Magnetization
$d > 2$	$n = \infty$	"Spherical" model	None	
$d = 3$	$n = 0$	Self-avoiding random walk	Conformation of long-chain polymers	Density of chain ends
	$n = 1$	Ising model in three dimensions	Uniaxial ferromagnet	Magnetization
			Fluid near a critical point	Density difference between phases
			Mixture of liquids near consolute point	Concentration difference
			Alloy near order-disorder transition	Concentration difference
	$n = 2$	XY model in three dimensions	Planar ferromagnet	Magnetization
			Helium 4 near superfluid transition	Amplitude of superfluid phase
	$n = 3$	Heisenberg model in three dimensions	Isotropic ferromagnet	Magnetization
$d \leq 4$	$n = -2$		None	
	$n = 32$	Quantum chromodynamics	Quarks bound in protons, neutrons, etc.	

UNIVERSALITY HYPOTHESIS states that diverse physical systems behave identically near their critical points. In most cases the only factors that determine the critical properties are the dimensionality of space, d, and the dimensionality of the order parameter, n. For magnetic systems the order parameter is the magnetization, and its dimensionality is the number of components needed to describe the spin vector. Most systems with the same values of d and n are members of the same universality class and share the same critical exponents. For example, ferromagnets that resemble the three-dimensional Ising model, fluids, mixtures of liquids and certain alloys are all members of the class with $d = 3$ and $n = 1$; graphs of their properties near a critical point should all have the same form. The interpretation of some values of d and n is less obvious, and values such as $n = -2$ can be defined mathematically but correspond to no known physical system. The XY model and the Heisenberg model are similar to the Ising model but describe ferromagnets whose spin vectors have two and three components respectively.

The gift of a **Garden**, free!

If it grows it's in Garden!

Garden magazine explores the obvious and the obscure, the how-to and the how-come, the superstition and the science of botany, horticulture, floriculture, agriculture and the environment. A gift of year-round entertainment for the enterprising and inquisitive gardener.

A **Garden** sampler:
The Spiderwort Assay: A Japanese plant geneticist has developed clones of a wildflower that can detect ionizing radiation with more sensitivity than a Geiger counter.

The Bamboozling Thamnocalamus: In synchronization throughout the world, a popular garden bamboo is blooming for the first time in a century —signaling the death of this entire generation of the species.

The Non-Lawn Grass: There is a beautiful natural alternative to turf that requires no mowing, spraying or fertilizing.

Power Lines in the Wake of 2,4,5-T: There is a non-chemical alternative to the country's most toxic herbicide—currently used by utility companies to maintain fully 32,000 square miles of transmission line rights-of-way.

The Directors of the Chicago Horticultural Society, The Horticultural Society of New York, Los Angeles Arboretum and The New York Botanical Garden invite you to subscribe now at a cost of $10 for six bi-monthly, full-color issues.

Yes, start my subscription with a free current issue.
I enclose a $10 check ($12 foreign) for six bi-monthly full-color issues.
Please send **Garden** to:
Name _____

Address _____

_____

Edition wanted:
☐Chicago Horticultural Society
☐Horticultural Society of New York
☐Los Angeles Arboretum
☐New York Botanical Garden/ Queens Botanical Garden
☐No regional news, please

Clip and mail to:
Subscription Department, **Garden** magazine, Botanical Garden, Bronx, N.Y. 10458

If you have ever taken a luxury sports car through a tight turn, you know the feeling. It's the sense of supreme precision with which this trim, compact camera proclaims its Nikon heritage. A feeling that is borne out by the professional quality pictures the Nikon FE delivers with automatic ease. And one that, unlike other fine things in life, is readily affordable.

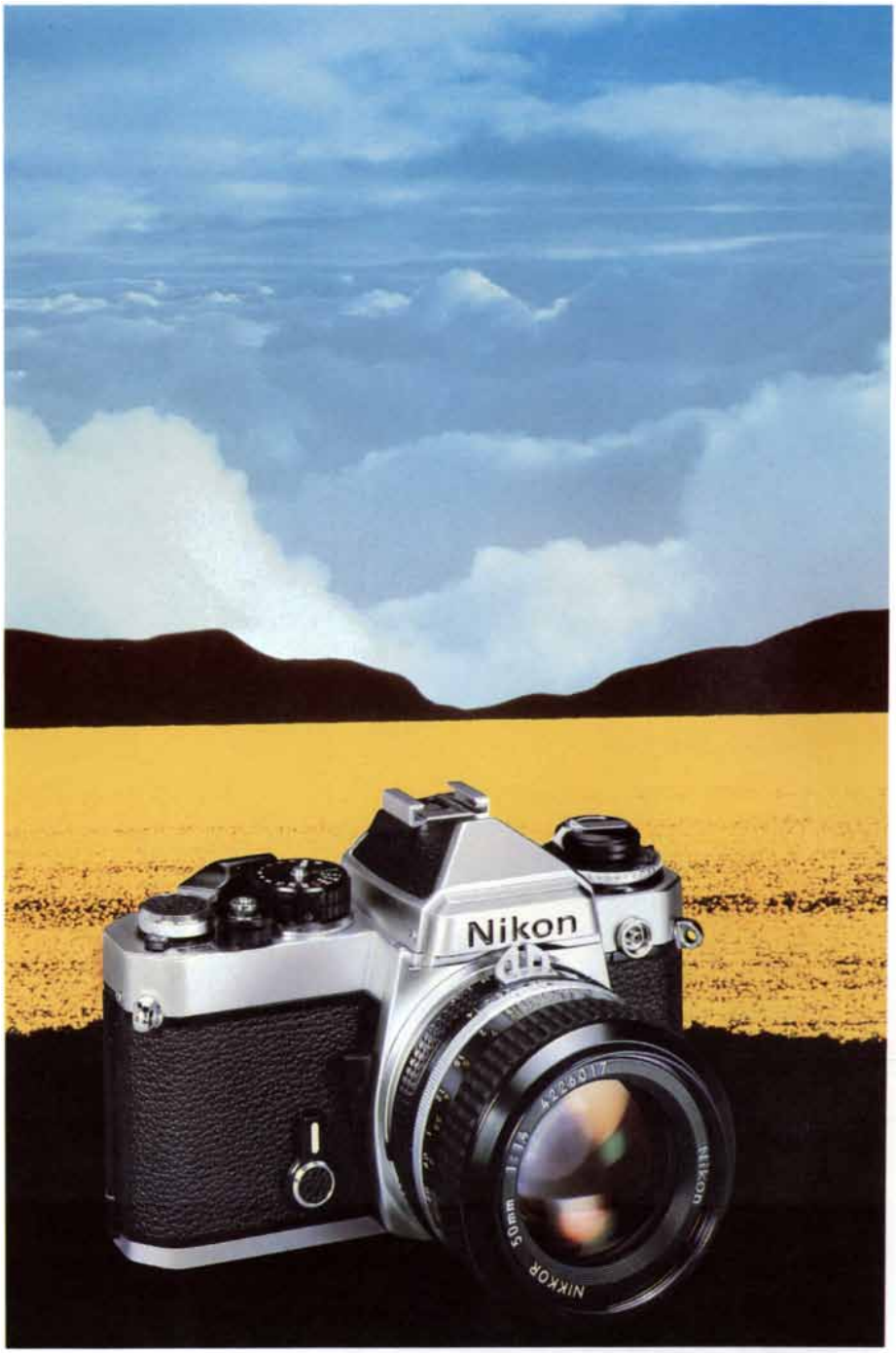
With the Nikon FE, you can simply focus and shoot... and rely on its Nikon electronics to give you sharp, magnificently exposed photographs, automatically. Or, switch to manual operation and enjoy complete creative control over every exposure, more easily than you ever thought possible.

Above all, this is a camera that makes no compromise in its supreme Nikon quality. Stroke the advance lever, and feel the smoothness of precision gearing machined to microscopic tolerances. Press the exposure button, and hear the shutter respond with hushed precision. Look through the bright, silver-coated viewfinder, and see your picture snap into sharp focus with a fingertip touch.

Know, too, that the world's greatest photographic system stands behind your Nikon FE. Add the dynamic firepower of motor drive, at up to 3.5 shots a second. Banish darkness with the ingenious automatic thyristor flash. Explore new perspectives through more than 60 Nikkor lenses, the same superb optics chosen by most professionals for their sharpness and color fidelity.

For the purist:
The Nikon FM
For those who prefer only manual exposure control, the Nikon FM offers the reliable guidance of one-step electronic metering. It's as compact and precisely responsive as the FE and costs even less. At your Nikon dealer.

Nikon Inc., Garden City, New York 11530.
In Canada: Nikon Canada, Inc.



Experience a sense of perfection. The Nikon FE.

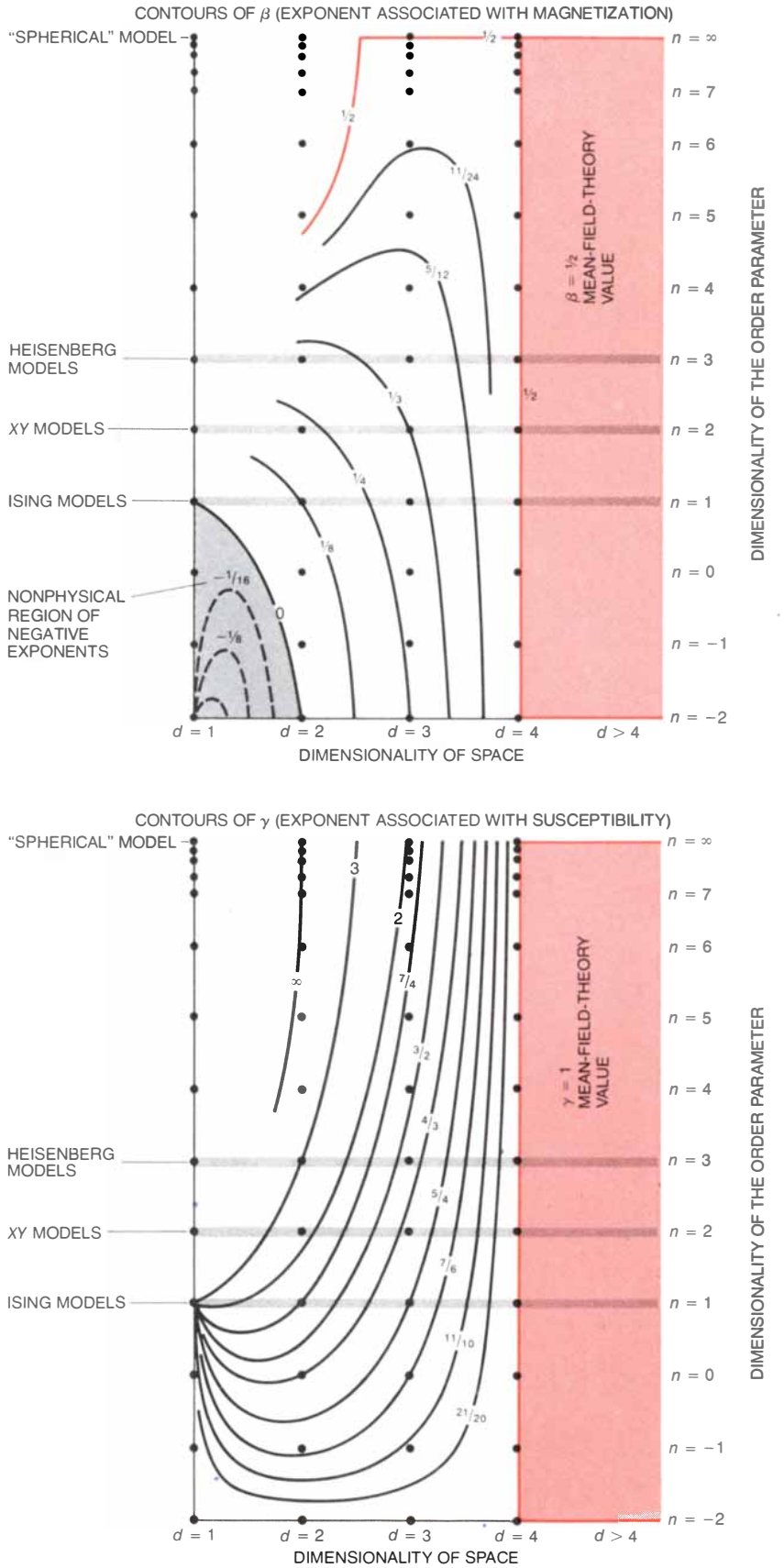
can even take on the value -2, although the physical meaning of a negative number of vector components is not clear.

The only values of d and n that have a straightforward physical meaning are integer values. This is particularly clear in the case of d, since a space with a noninteger number of dimensions is hard even to imagine. In renormalization-group calculations, however, d and n appear in equations where they can be allowed to vary continuously over some range. It is even possible to draw a graph in which the values of critical exponents are plotted as continuous functions of d and n. The exponents have well-defined values not only for integer dimensions but also for all fractional dimensions between the integers. Such a graph shows that the exponents approach the values given by mean-field theory as the number of spatial dimensions approaches 4. When d is equal to exactly 4, and at all higher values of d, the mean-field values are exact. This observation has given rise to an important method of performing renormalization-group calculations. The dimensionality of space is expressed as being equal to $4 - \epsilon$, where ϵ (the Greek letter epsilon) is a number that is assumed to be small. The critical exponents can then be calculated as the sum of an infinite series of terms including progressively higher powers of ϵ. If ϵ is less than 1, a high power of ϵ will have a small value, and reasonable accuracy can be obtained by neglecting all but the first few terms in the infinite series.

This calculation method, which is called the epsilon expansion, was developed by Michael E. Fisher of Cornell University and me. It is a general method for solving all the problems to which mean-field theory can be applied, and it represents the natural successor to Landau's theory. Indeed, it supplies answers in the form of corrections to the values given by mean-field theory. The block-spin method is the more transparent technique, but the epsilon expansion is the more powerful one.

It is not entirely surprising that the critical exponents should converge on the mean-field values as the number of spatial dimensions increases. The fundamental assumption of mean-field theories is that the force at each lattice site is influenced by conditions at many other sites. The number of nearest-neighbor sites increases along with the number of spatial dimensions. In a one-dimensional lattice each site has just two nearest neighbors, in a two-dimensional lattice four, in a three-dimensional lattice six and in a four-dimensional lattice eight. Hence as the dimensionality increases, the physical situation begins to resemble more closely the underlying hypothesis of mean-field theory. It remains a mystery, however, why $d = 4$ should mark a sharp boundary above which the mean-field exponents are exact.

In this article I have discussed main-



VARIATION OF CRITICAL EXPONENTS with the dimensionality of space (d) and of the order parameter (n) suggests that physical systems in different universality classes should have different critical properties. The exponents can be calculated as continuous functions of d and n, but only systems with an integral number of dimensions are physically possible. In a space with four or more dimensions all the critical exponents take on the values predicted by mean-field theories. The graphs were prepared by Michael E. Fisher of Cornell University.

ly the applications of the renormalization group to critical phenomena. The technique is not confined to those problems, however, and indeed it did not begin with them.

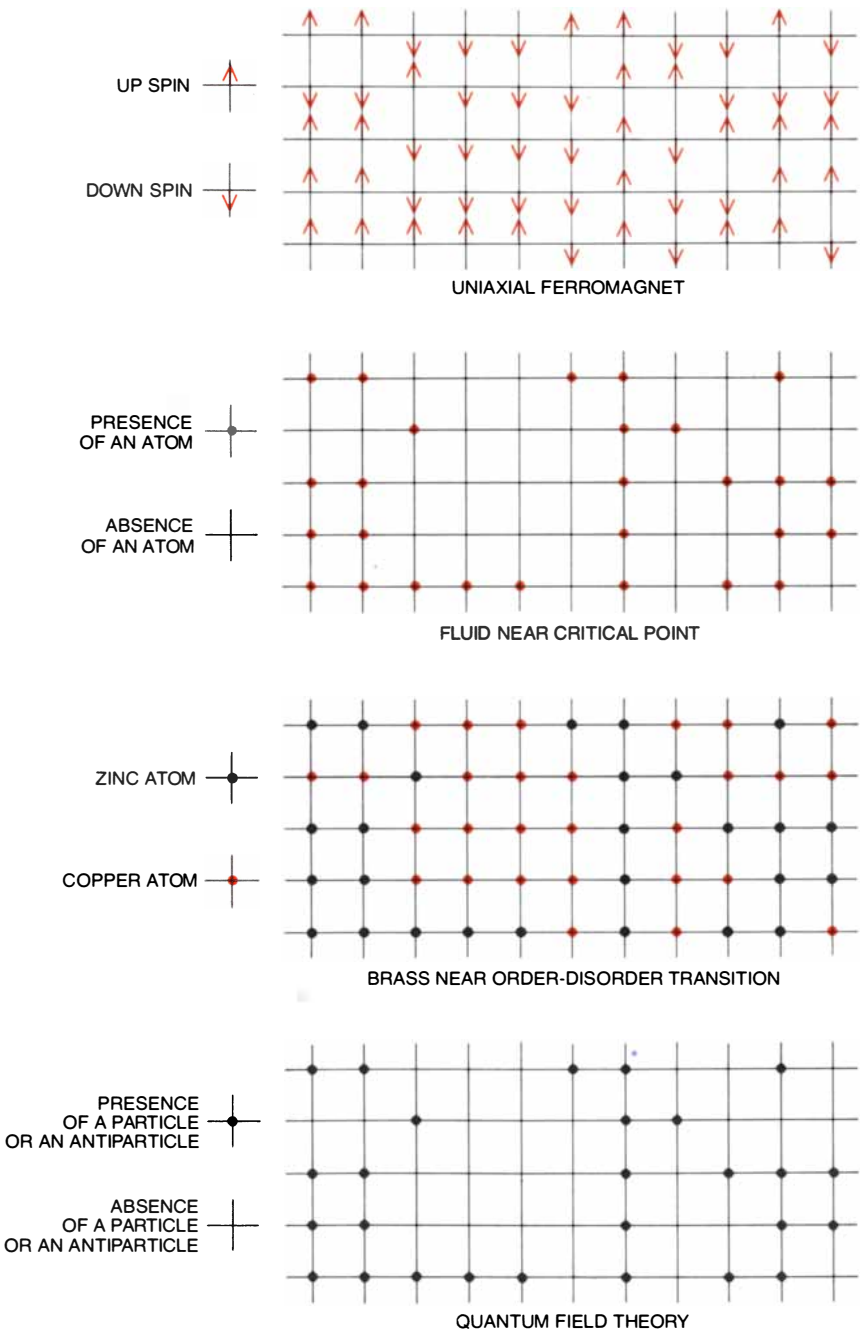
The procedure called renormalization was invented in the 1940's as part of the development of quantum electrodynamics, the modern theory of interactions between electrically charged particles and the electromagnetic field. The difficulty encountered in the formulation of the theory can be understood as one of multiple scales of length. For some time it had been apparent that the charge of the electron predicted by quantum-mechanical theories was infinite, a prediction that was in serious conflict with the measured charge. The renormalized theory of electrodynamics does not abolish the infinity; on the contrary, the electron is defined as a point particle whose "bare" charge is infinite. In quantum electrodynamics the bare charge has the effect, however, of inducing a charge of opposite polarity in the surrounding vacuum, which cancels most of the infinity, leaving only the small net charge that is observed in ordinary experiments.

One can imagine a probe particle that could measure the electron's charge at arbitrarily close range. At long range it would find the familiar finite value, which is the difference between the bare charge and the induced charge. As the layers of shielding were penetrated the measured charge would increase, and as the range was reduced to zero the charge would become infinite. The renormalization procedure provides a means for subtracting the infinite shielding charge from the infinite bare charge so that a finite difference results.

In the 1950's it was pointed out by several workers, among them Murray Gell-Mann and Francis E. Low, that the renormalization procedure adopted for quantum electrodynamics is not unique. They proposed a more general formulation, which is the original version of the renormalization group. In their application of the method to quantum electrodynamics a mathematical expression is constructed that gives the magnitude of the charge at some definite distance from the electron. Then the form of the expression is examined as the distance at which the measurement is made is allowed to approach its limiting value of zero. The arbitrariness of the procedure is in the choice of the initial distance. Any value can be selected without changing the ultimate results, so that there is an infinite set of equivalent renormalization procedures.

A "group" in mathematics is a set of transformations that meets a special requirement: the product of any two transformations must also be a member of the set. For example, rotations are transformations that make up a group, since the product of any two rotations is also a rotation. What this means in the case of the renormalization group is that the procedure can be iterated indefinitely, since applying the procedure twice is equivalent to applying the product of two transformations. Actually the renormalization group is properly called a semigroup because the inverse of the transformation is not defined. The reason for this can be seen plainly in the block-spin technique applied to the two-dimensional Ising model. A block of nine spins can be condensed into a single average spin, but the original spin configuration cannot be recovered from the average because essential information has been lost.

The version of the renormalization



LATTICE SYSTEM can be interpreted as a model not only of a ferromagnet but also of other physical systems that have fluctuations on many scales. The Ising model describes a uniaxial ferromagnet, one with a preferred axis of magnetization. It can also be applied to a fluid near its critical point, where each lattice site either is occupied by an atom or is vacant, so that the fluctuations become variations in density. An alloy such as brass has a similar structure, where each site is occupied by one kind of metal or the other. In all these systems the fluctuations are thermal; in the quantum field theories that describe the interactions of elementary particles there are quantum fluctuations of the vacuum, which allow particles and antiparticles to appear spontaneously. A simple quantum field theory can be formulated on a lattice by specifying that the particles and antiparticles can be created and annihilated only at the lattice sites.

group outlined in this article differs in several respects from the one introduced by Gell-Mann and Low. The earlier version of the technique is useful only for understanding problems that can be solved by one of the traditional methods of physics: by finding some approximate expression for the behavior of a system and then calculating better approximations as a series of perturbations departing from the original expression. Moreover, in the original formulation only one quantity is allowed to vary; in the example given above it is the charge of the electron. As a consequence the surface in parameter space is not a multidimensional landscape but a mere line. The modern version of the renormalization group, which was introduced by me in 1971, gives access to a much broader spectrum of physical problems. What is equally important, it gives a physical meaning to the renormalization procedure, which otherwise seems purely formal.

In the past few years I have been attempting to apply the newer version of the renormalization group to a problem in the physics of elementary particles. The problem is how to describe the interactions of quarks, the hypothetical elementary particles thought to compose protons, neutrons and a multitude of related particles. In one sense the problem is much like the original renormalization problem of quantum electrodynamics; in another sense it is just the opposite.

In quantum electrodynamics the charge of the electron is found to increase as the electron is approached more closely. For interactions of quarks the property analogous to electric charge is called color, and for that reason the theory of quark interactions has been named quantum chromodynamics. When the color charge of a quark is measured at close range, it seems to diminish as the distance becomes smaller. As a result two quarks that are very close together interact hardly at all: the coupling between them is weak. On the other hand, when the quarks are pulled apart, the effective color charge increases and they become tightly bound. Whereas an electron induces a compensating charge in the surrounding space, a quark seems to induce a color charge with the same polarity, which augments its own charge at long range. Indeed, it is a widely accepted hypothesis that the effective coupling between quarks increases without limit when the distance between them exceeds the diameter of a proton, which is about 10^{-13} centimeter. If that is true, a quark could be torn loose from a proton only by expending an infinite quantity of energy. The quarks would be permanently confined.

One way of visualizing the binding of quarks is to construct imaginary lines of force between them. The strength of the coupling is then proportional to the number of lines per unit area that cross any surface between the particles. In the case of electrons when the particles are separated, the lines of force spread out in space, so that there are fewer lines per unit area. The density of lines declines as the square of the separation, which yields the familiar inverse-square law for the electromagnetic force. With quarks, on the other hand, the prevailing hypothesis holds that the lines of force do not spread out in space; they remain confined to a thin tube, or string, that directly links the quarks. As a result the number of lines per unit area remains constant no matter what the distance is, and the quarks cannot be separated. Although this account of quark confinement has an intuitive appeal, it is a qualitative explanation only. No one has yet been able to derive the confinement of quarks from the underlying theory of quantum chromodynamics.

The confinement problem is one with many scales of length and energy and hence is a candidate for renormalization-group methods. I have formulated a version of the problem in which the quarks occupy the sites of a lattice in four-dimensional space-time and in which they are connected by "strings" that follow the lines connecting sites. The lattice is a strictly artificial structure with no analogue in real space-time, and it must ultimately disappear from the theory. That can be accomplished by allowing the lattice spacing to approach zero.

As in the study of ferromagnetic systems, a renormalization-group transformation is applied repeatedly to the lattice of quarks and strings. In this way the interaction of the quarks can be examined at progressively larger separations. The question to be answered is whether the lines of force remain confined to tubelike bundles or spread out in the lattice as the scale of length is increased. The calculations are near the limit of practicality for the present generation of digital computers. As yet I do not have the answers.

There are many other problems seemingly suitable for renormalization-group methods but that have not yet been expressed in such a way that they can be solved. The percolation of a fluid through a solid matrix, such as water migrating through the soil or coffee through ground coffee beans, involves aggregations of fluid on many scales. Turbulence in fluids represents a problem of notorious difficulty that has resisted more than a century of effort to describe it mathematically. It is characterized by patterns with many characteristic sizes. In the atmosphere, for example, turbulent flows range in scale from small "dust devils" to hurricanes.

One problem that has yielded to the renormalization group is a phenomenon in solid-state physics called the Kondo effect, after the Japanese physicist Jun Kondo. The effect is observed in nonmagnetic metals, such as copper, when they are contaminated with a small concentration of magnetic atoms. The simplest theories predict that the electrical resistance of such a metal will fall continuously as the temperature is reduced. Actually the resistance reaches a minimum value at a finite temperature and then rises again as the temperature is reduced further. The anomaly was never one of pressing importance because an explanation of it does not illuminate more general properties of solids, but it tantalized physicists for more than 40 years, always seeming just beyond the reach of the available methods. The root of the difficulty is that the conduction electrons in the metal can have any energy over a range of a few volts, but perturbations in that energy are significant down to a level of about 10^{-4} volt. The problem was ultimately solved in 1974, when I completed a renormalization-group calculation of the electron energies at all temperatures down to absolute zero.

A more recent series of renormalization-group calculations is notable in that it makes predictions that have been directly confirmed by experiment. The calculations concern the lattice-spin model in which d equals 2 and n equals 2, or in other words concern a two-dimensional lattice of two-component spins. It has been proved that no phase with long-range order is possible in this spin system, but renormalization-group studies done by J. M. Kosterlitz of the University of Birmingham and David J. Thouless of Yale University have shown that the behavior of the system does change abruptly at a critical temperature. These findings have been applied to studies of thin films of superfluid helium 4, which also fall into the universality class of $d = 2$ and $n = 2$. In particular Kosterlitz and David R. Nelson of Harvard University have predicted a discontinuous jump in the density of the superfluid fraction of the film. Such a jump has since been observed experimentally by John D. Reppy of Cornell and others and has been found to have the predicted magnitude.

For all the work that has been invested in the renormalization group it may seem the results obtained so far are rather scanty. It should be kept in mind that the problems to which the method is being applied are among the hardest problems known in the physical sciences. If they were not, they would have been solved by easier methods long ago. Indeed, a substantial number of the unsolved problems in physics trace their difficulty to a multiplicity of scales. The most promising path to their solution, even if it is an arduous path, is the further refinement of renormalization-group methods.

The Hormonal Control of Behavior in a Lizard

The green anole, often mistakenly called a chameleon, is a good subject for examining how the sex glands and the brain interact in orchestrating the sexual behavior of both males and females

by David Crews

Generations of American children have become familiar with the green anole lizard (*Anolis carolinensis*) as the chameleon, since it is the animal usually sold by that name in pet shops. Recently biologists have become familiar with *A. carolinensis* as an excellent animal for laboratory studies of the interaction of behavior and hormones: the "chemical messengers" in the body that act at a distance from the site of their manufacture. It was a boyhood interest in the green anole (which I then thought was a chameleon) that led me as a behavioral biologist to work with the animal as a means of investigating the bases of reproductive behavior. The findings among other things provide a new outlook on the adaptive function of the relation between behavior and hormones that is seen in animals of quite different kinds.

The true chameleon and the green anole have much in common. They are both lizards. Most species live in trees or bushes, subsisting mainly on insects. Both can change color, although the anole's ability to do so is considerably more limited than the chameleon's. This is the trait that has made chameleons and anoles popular as pets, but the anole's color change is not (as many people think) related to the color of the background. Instead it is determined by such factors as light and temperature or by such emotions as fright, triumph and defeat. The chameleon is an animal of the Old World, whereas the anoles are found in the warmer regions of North and South America. The chameleon lays from two to 40 eggs at a time, the anole only a single egg.

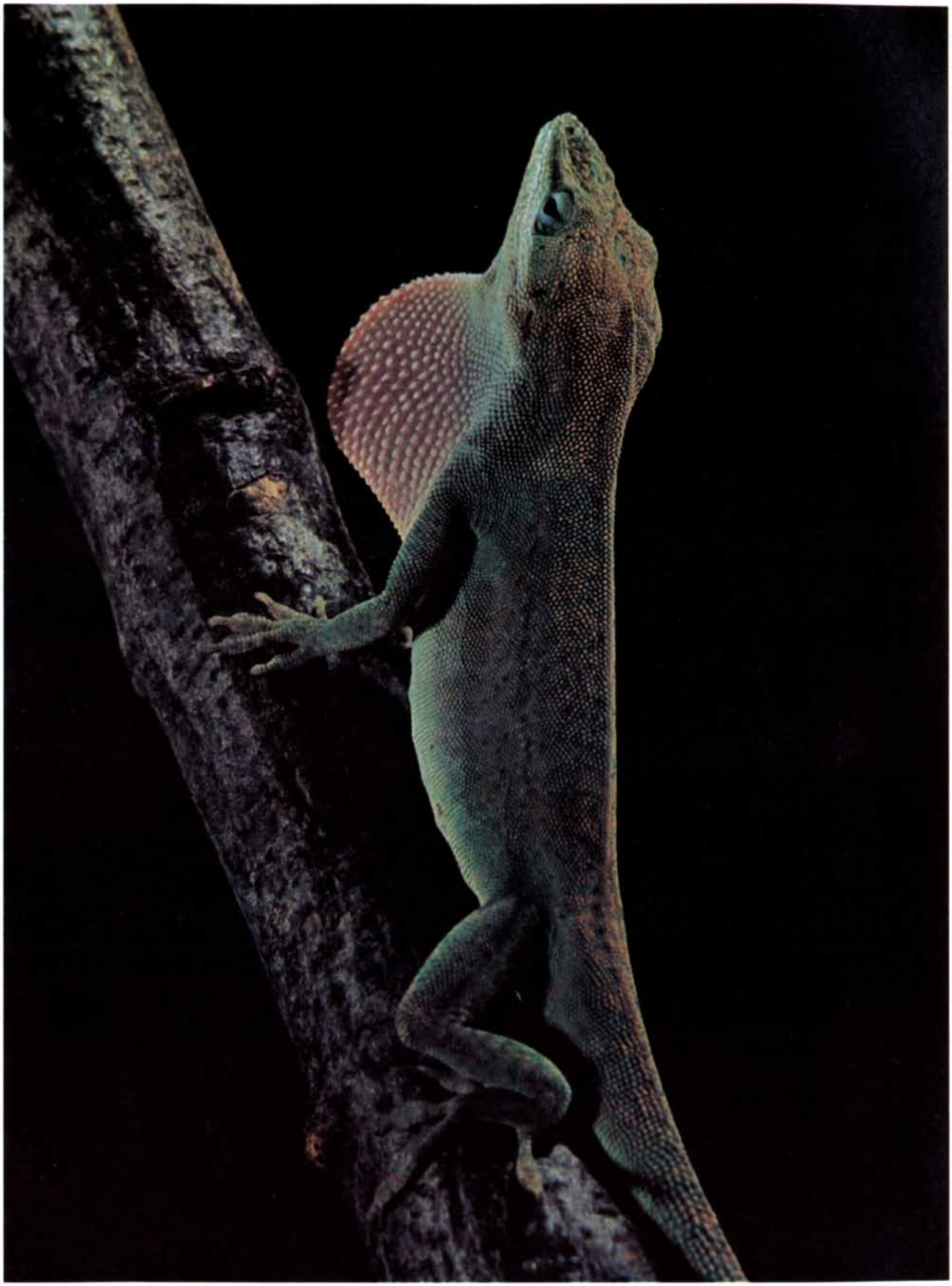
The particular value of the green anole as an experimental animal is that it is abundant and that under the appropriate conditions it will establish in the laboratory the same social system and behavior it displays in its natural environment. A typical population might be found in southern Louisiana. There the anoles are reproductively inactive and essentially dormant from late September through late January. During that time they cluster in groups behind the loose bark of dead trees or under rocks and fallen logs. Beginning late in January or early in February the males emerge and establish breeding territories. The females begin to be active about a month later; by May each female is laying her single egg in the ground at intervals of from 10 to 14 days. The breeding season, which extends through August, is followed by about a month in which both males and females are refractory, that is, insensitive to the environmental and social stimuli that in the spring induced them to begin breeding.

Research on the reproductive behavior of a number of kinds of animal, most notably the work of the late Daniel S. Lehrman with the ring dove, has shown that the behavior is the consequence of two interlocking systems, one system represented by the animal and its internal state and the other by the animal and its environment. For example, among many birds and mammals the increasing amount of daylight in the spring acts on the brain of males to stimulate the pituitary gland to secrete increasing quantities of gonadotropic hormones. In lizards the primary stimulatory cue is the increase in temperature. Paul Licht of the University of California at Berkeley has shown that lizards also differ from birds and mammals in secreting one gonadotropic hormone rather than two of them.

The gonadotropin is transported by the bloodstream to the testes, where it stimulates the production of sperm and the secretion of androgens (the collective term for male hormones), specifically testosterone. The rising concentration of testosterone in the blood feeds back to specific areas of the brain to modulate further secretion of gonadotropin by the pituitary and to activate male sexual behavior. The male's behavior then becomes an important feature of the environment that stimulates the pituitary of the female to secrete gonadotropin, which rapidly induces ovarian development, steroid-hormone secretion and female sexual behavior. As the female responds to the male, her behavior influences his behavior and indirectly his physiology, and so the cycle begins anew.

The green anole has a rich repertory of behavior. A sexually active male will patrol his territory, stopping at prominent perches to execute an "assertion display." It is characterized by a bobbing movement that is typical of the species and is coordinated with the extension of a red dewlap at the throat. If a strange male green anole enters the territory, the resident male responds with a "challenge display." It is identified by an extreme lateral compression of the body and a highly stereotyped bobbing movement. If the intruding animal does not immediately respond with a "submission display" (a rapid nodding of the head) or responds aggressively to the challenge, the resident approaches the intruder and a fight ensues. During such a fight the dewlap is not extended but the throat region is engorged by a lowering of the tongue-supporting apparatus. As the fight progresses a crest is erected along the back and neck and a black spot forms directly behind the eye. It is not uncommon for males to lock jaws while they circle each other, each one trying to throw the other off the perch. After a fight the winning male (almost always the resident) is likely to climb to a prominent perch and execute a series of assertion displays.

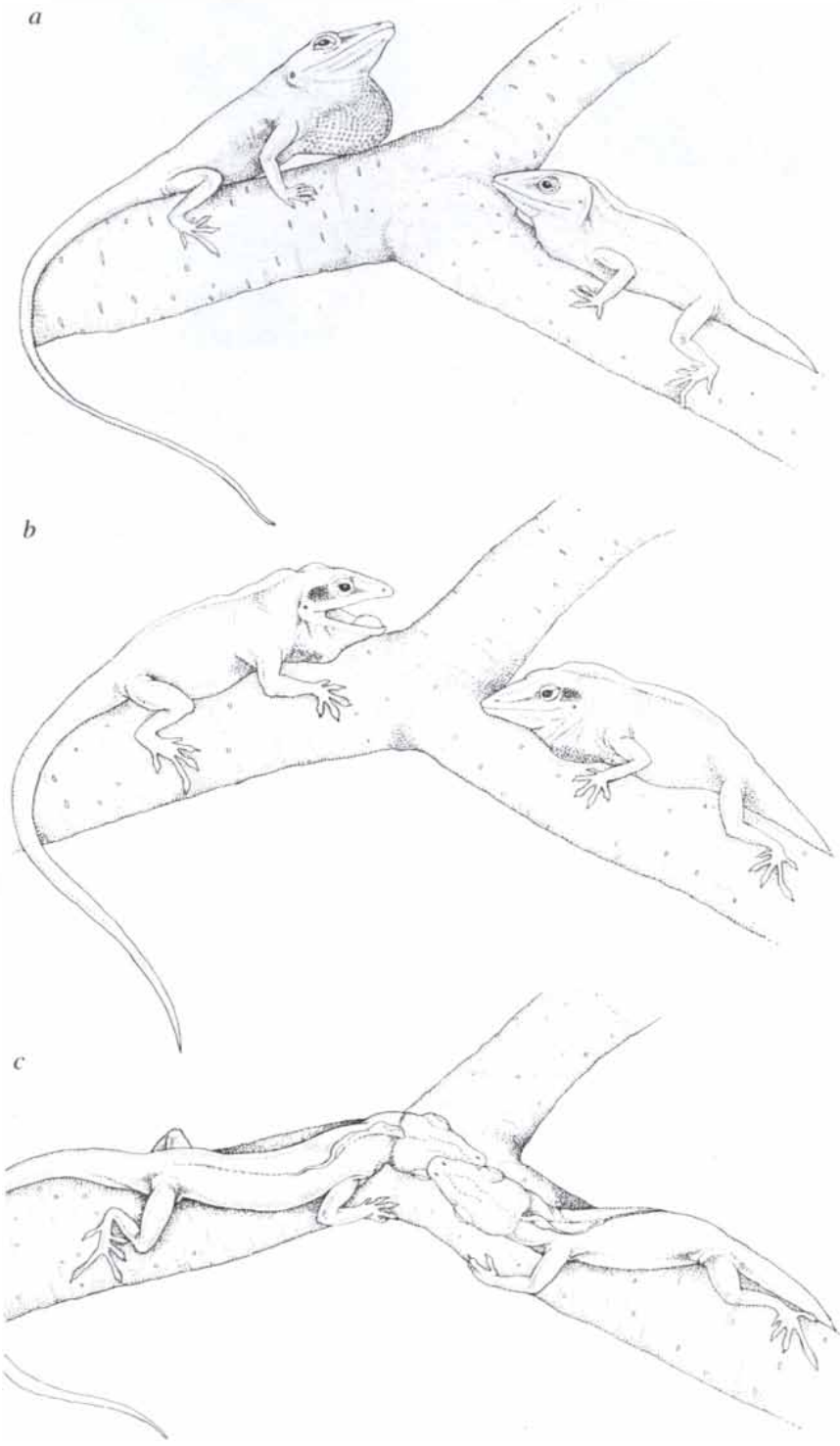
In a "courtship display" a male advances toward a female, pausing for a series of bobbing movements and exhibitions of the dewlap. The number of bobbing movements varies greatly among individual males; the variation may help the female to distinguish one male from another. If the female is sexu-



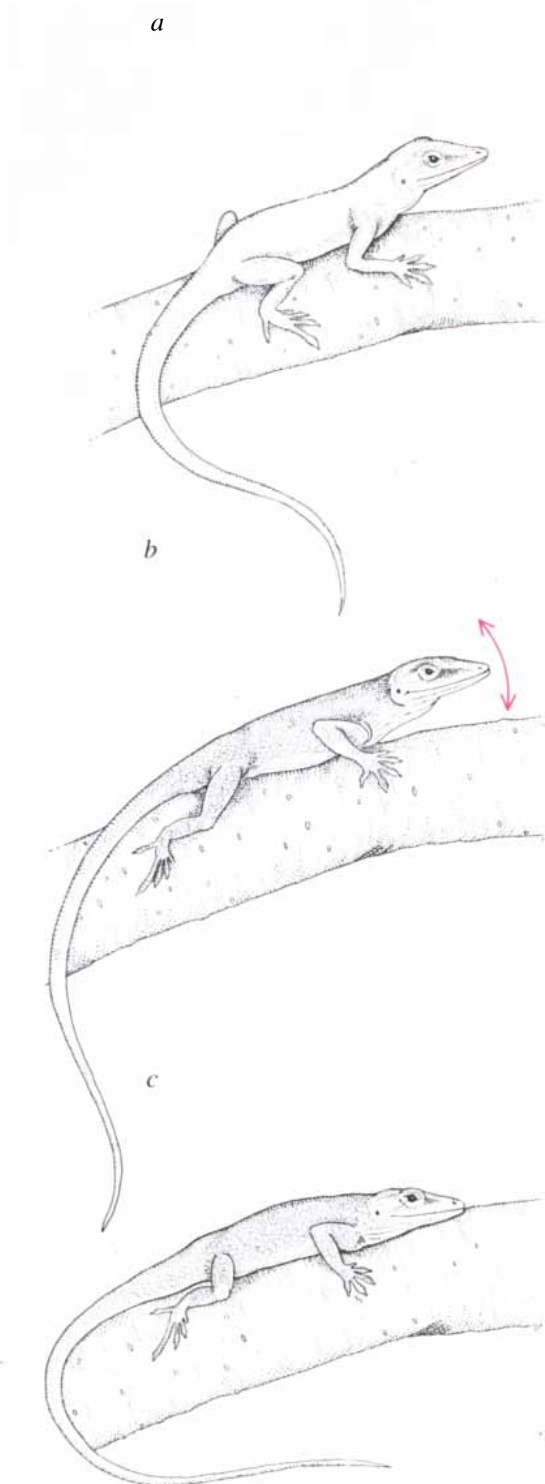
COURTSHIP DISPLAY is given by a male green anole lizard (*Anolis carolinensis*) that has sighted a female, which is not visible in the photograph. The male advances toward the female, pausing from time to time to extend the red dewlap at his throat and to execute a series of bobbing movements. The amount of bobbing varies from male to male and so perhaps enables a female to distinguish males.

ally receptive, she will allow the male to approach her and grip her neck. The male then mounts the female and curves his tail under hers. This action brings the cloacal regions into apposition so that the male can evert one of his two penises. (Lizards and snakes are unique among the vertebrates in having two of these organs, each of which is called a hemipenis.)

Much has been learned about the physiological control of the *A. carolinensis* male's reproductive behavior. As one might expect, male sexual behavior depends on the gonads. Castration leads to a rapid decline in courtship activity, but the administration of androgen reinstates it. Aggressive behavior appears to be less affected by castration. It is strongly affected, however, by environmental factors. For example, if a male



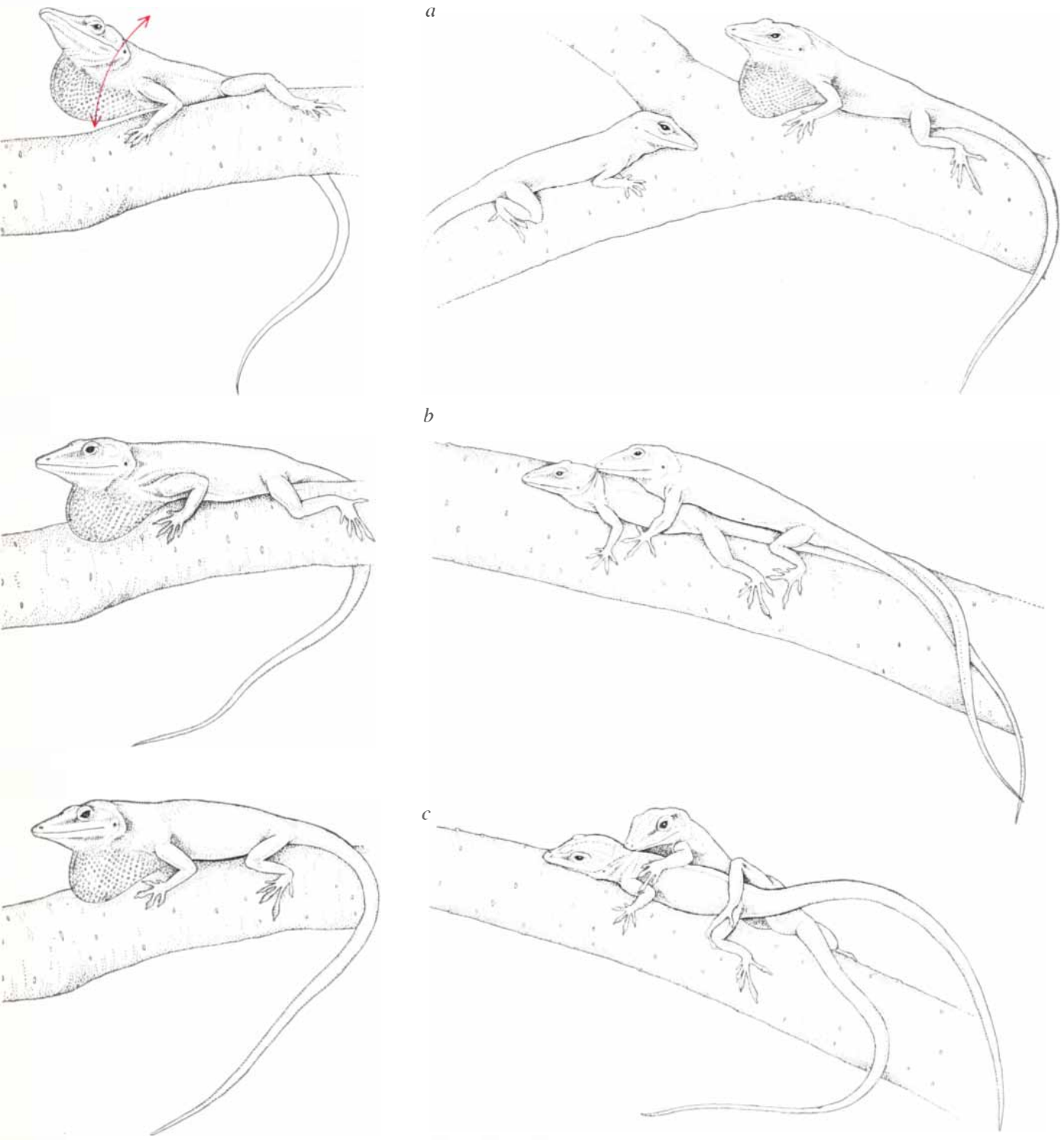
CHALLENGE DISPLAY given by a resident male (*a*) when a strange male intrudes on his territory involves an extension of the dewlap, a pronounced lateral compression of the body and a stereotyped bobbing of the body. If the intruder does not respond with a submission display, the resident moves toward the intruder (*b*), often with mouth agape, and they fight. During the fight a crest is erected along the back and neck and a black spot appears behind the eye. The males may lock jaws (*c*). The fight ends when one lizard throws the other off the perch.



CHALLENGE AND SUBMISSION are portrayed. A sequence begins (*a*) when a strange male, here the one on the left, intrudes on the territory of a resident male. The resident responds with a challenge display. This time, however, the intruder does not react aggres-

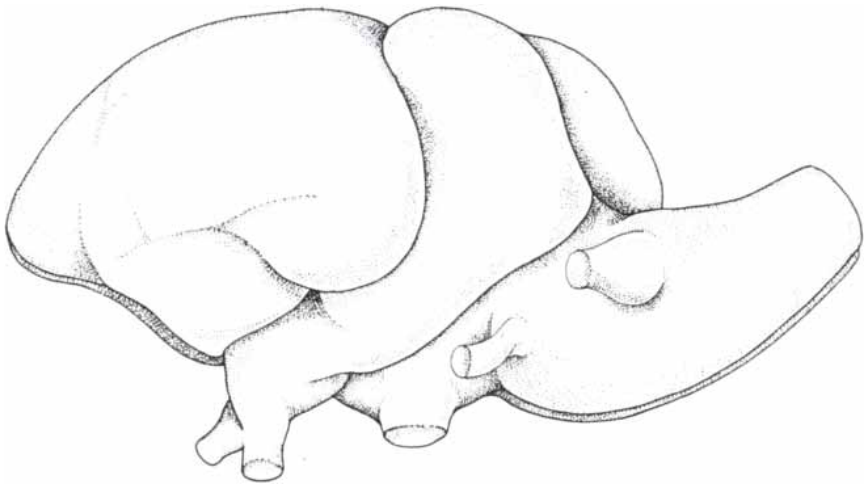
is returned to his home cage after castration, his aggressive behavior declines slowly or not at all, whereas if he is put in an unfamiliar cage, aggressiveness decreases quickly.

It has long been assumed that in vertebrate animals a male's reproductive behavior is triggered by the action of androgens on the brain. An alternative hypothesis was recently proposed by Frederick Naftolin of the Yale University School of Medicine. In essence it is that the "male" hormone testosterone is converted by enzymes in the brain into a "female" hormone, estradiol, which then activates the male's sexual behavior. The process that characterizes the structural change in the hormone is known as aromatization. In stimulating secondary sex structures, however, testosterone is believed to act directly or



sively but instead gives a submission display (*b*), which entails a rapid nodding of the head, and turns from green (*light*) to brown (*dark*). With this response the resident male does not erect a crest and a black spot does not form behind his eye (*c*). In time the intruder leaves.

COURTSHIP AND MATING of the green anole begin (*a*) when a sexually active male sees a female and responds with a courtship display, which includes repeated extension of the dewlap (in the manner shown in the photograph on page 181) and the individually characteristic series of head-bobbing movements. If the female is sexually receptive, she arches her neck and allows the male to approach and grip it (*b*). In mounting (*c*) the male everts one of his two penises, each termed a hemipenis; if he is on the right side of the female, it is the left one, and vice versa.



BRAIN OF THE GREEN ANOLE is only .003 percent the size of the human brain; here it is depicted at an enlargement of about 20 diameters. Because the brain is so small a special stereotaxic apparatus was needed to accurately place tiny needles and tubes in it for experiments.

by its conversion into nonaromatizable androgens such as dihydrotestosterone.

Several studies support the concept that the aromatization of androgens to estrogens (the collective term for a group of female sex hormones) occurs normally in the brain of many species of mammals and that estradiol, not testosterone, activates male sexual behavior in many species. Gloria Callard, Zoltan Petro and Kenneth Ryan of the Harvard Medical School have identified aromatizing enzymes in the brain of turtles and snakes. They have reported that when specific parts of the brain of the freshwater turtle *Chrysemys picta* are incubated with androgens that have been labeled with radioactive androgens, some radioactively labeled estrogens are found in the tissue later. This discovery and others led them to propose that the aromatization of androgens is an evolutionarily ancient property of the vertebrate nervous system.

Before the androgen-aromatization hypothesis can be generalized to all reptiles the effects of aromatizable and nonaromatizable androgens must be tested in living animals. Lizards are well suited to such an investigation. In both lizards and snakes a section of the kidney has been modified to form a sex segment. This structure manufactures the seminal fluid and is similar to the prostate and seminal vesicles of mammals. The secretory activity of the kidney sex segment, which is highly sensitive to stimulation by androgen, can be determined by measuring increases in the height of the epithelial cells of the kidney tubules after hormonal stimulation.

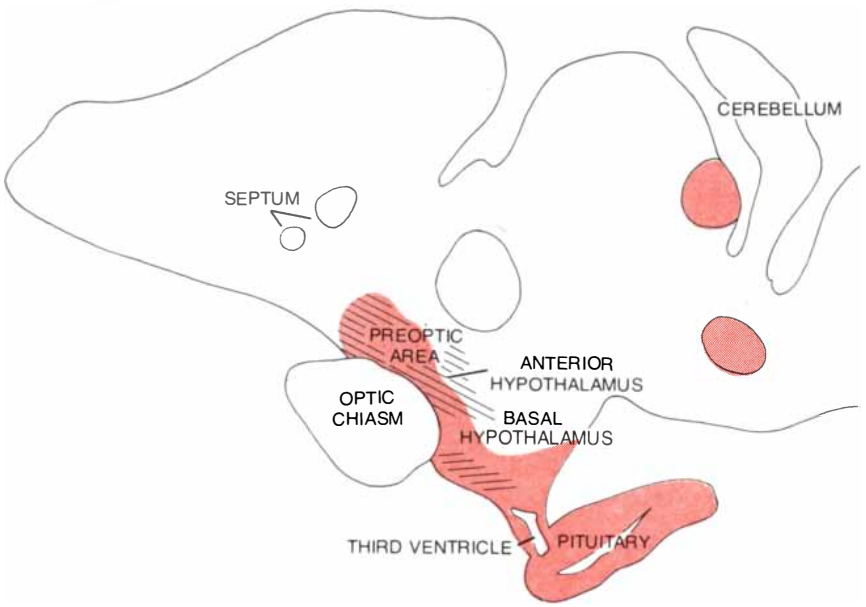
Working with this measure as an indication that hormones were affecting the sex organs directly and with the stimulation of sexual behavior such as courtship displays as an indication that hormones were acting on the brain, we did an experiment in which castrated males were treated with testosterone, dihydrotestosterone or estradiol. Examination of the kidney sex segment indicated that only the first two hormones were acting directly on the sex organs. Only testosterone reinstated sexual behavior.

This finding does not exclude the possibility that estradiol acts in the male brain. Several workers have shown that administering dihydrotestosterone and estradiol simultaneously stimulates copulatory behavior in castrated rats. Results of this kind have led to the suggestion that male sexual performance requires both the central or brain action of estradiol and the peripheral action of androgens to stimulate the sex organs. Support for the hypothesis has come from the work of two groups of investigators, one group headed by Julian M. Davidson of Stanford University and the other by Paula Davis and Ronald J. Barfield of Rutgers University. They found that implants of estradiol in the brain of rats will induce sexual behavior only if dihydrotestosterone is administered systemically at the same time.

My colleagues and I at Harvard University also examined the possibility that dihydrotestosterone and estradiol act in concert in the green anole. When both hormones were administered simultaneously to castrated lizards, half of the animals (in two separate experiments) suddenly exhibited the complete reproductive pattern within a single testing period. The results were quite different from the outcome when only testosterone is given; it typically induces a gradual restoration of behavior over several days. Evidently estrogen can play as vital a role in the control of male sexual behavior in lizards as it does in some mammals.

Half of the castrated lizards in our experiments, however, showed no alteration in sexual behavior after the combination treatment. The animals for our experiments were collected from a variety of sites, and so it is possible that the behavioral variability reflects underlying differences between populations in the amount of aromatizing enzyme in the reptilian brain. The behavioral variability may also reflect genetic differences in an individual's sensitivity to hormones.

The individual variability in response



SECTION OF THE BRAIN shows the areas (*color*) where the cells that concentrate sex steroid hormones are situated, as revealed by autoradiography. In this technique male and female lizards from which the gonads have been removed are injected with radioactively labeled hormones; after the animal has been killed sections of its brain are put on slides coated with photographic emulsion. When the emulsion is developed, the areas exposed by radioactivity show where the hormones have been concentrated. Slanted lines indicate where brain lesions abolish male sexual and aggressive behavior, horizontal lines where lesions decrease pituitary function.

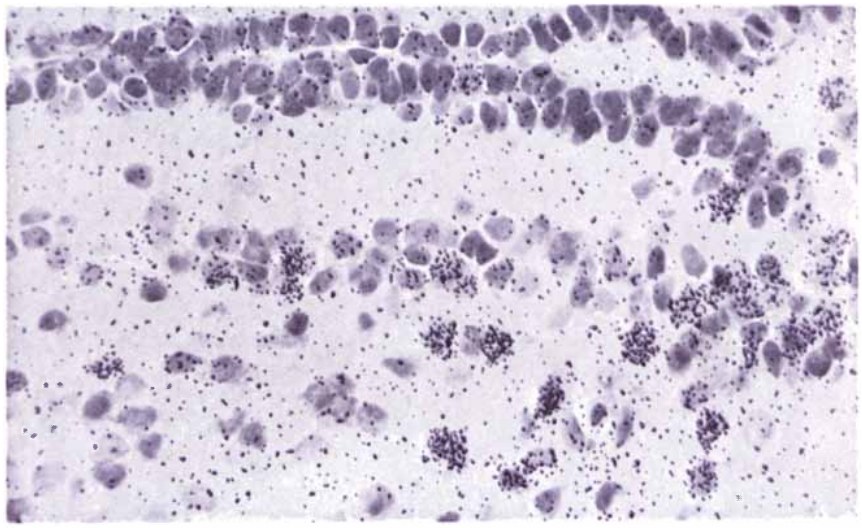
makes one wonder about the wisdom of relying on highly inbred stocks of animals for tests of drugs. Many of the inbred stocks of rodents result from repeated brother-sister matings, so that each strain is more like a single individual and does not reflect the normal spectrum of variability found in most natural populations. This important point is often overlooked in the interpretation of data on drug effects and in the determination of appropriate clinical treatments.

Hormones modulate both behavior and the function of the pituitary gland through their influence on the central nervous system. Our group has recently begun to study the interaction of hormones, the brain and behavior in the green anole.

Although much information is available about the areas of the mammalian brain that participate in the control of reproductive behavior and physiology, I was not sure that the information could be extrapolated to reptiles, and so I joined forces with Joan I. Morrell and Donald W. Pfaff of Rockefeller University. We first identified the sites in the brain of *A. carolinensis* that concentrate sex steroid hormones. Our technique was autoradiography. Male and female lizards from which the gonads had been removed were injected with a radioactively labeled hormone: estradiol, testosterone or dihydrotestosterone. After two hours the animals were killed; the brain and the pituitary gland were quickly removed and frozen. Then the brains were cut in thin sections in a darkroom, and the sections were placed on slides that had been coated with photographic emulsion. Half of these preparations were stored in lightproof boxes for six months, half for nine months. When the slides were removed and the emulsion was developed, the areas having a dense grouping of black grains over cell nuclei indicated where the radioactively labeled hormone had concentrated.

The results of this study show that in the lizard, as in all other vertebrate species that have been investigated, sex steroid hormones are concentrated in specific areas of the brain: the preoptic area, the anterior and basal hypothalamus and the limbic system. These structures have been found to play a crucial role in reproduction in mammals and birds. An examination of their function in reptiles, which have a more primitive level of organization, would shed light on the evolution of the mechanisms that control reproduction.

Since the brain of an adult male *A. carolinensis* is quite small, being only .003 percent the size of the human brain, we had to develop special techniques in order to investigate the function of the steroid-concentrating areas. Our efforts were greatly aided by the work of Neil B. Greenberg of the University of Tennessee. He devised a stereotaxic apparatus with which one can precisely direct the tip of an ultrafine needle or tube three-dimensionally to a particular place in a small brain, and he also prepared an atlas of the forebrain of *A. carolinensis*. Working with this atlas as a "road map" of the brain, we have begun to study the role the different hormone-sensitive areas play in the control of the animal's reproductive behavior and physiology.



CONCENTRATION OF ESTRADIOL in the brain of a female *A. carolinensis* is indicated by the black grains in this autoradiograph of a section through the periventricular nucleus of the hypothalamus. The enlargement is 550 diameters. The injected estradiol was radioactively labeled; hence it exposed grains of silver in the emulsion on which the brain section was placed.

A large body of research with mammals and birds indicates that the anterior hypothalamus-preoptic area is a major center for the integration of behavior. Since it is also a major site of steroid uptake in the lizard, we first examined its role in the regulation of male reproductive behavior. Destruction of the area in intact, sexually active males results in the immediate cessation of sexual behavior, whereas in sexually inactive castrated males the implantation of minute amounts of testosterone directly in the preoptic area reinstates sexual behavior.

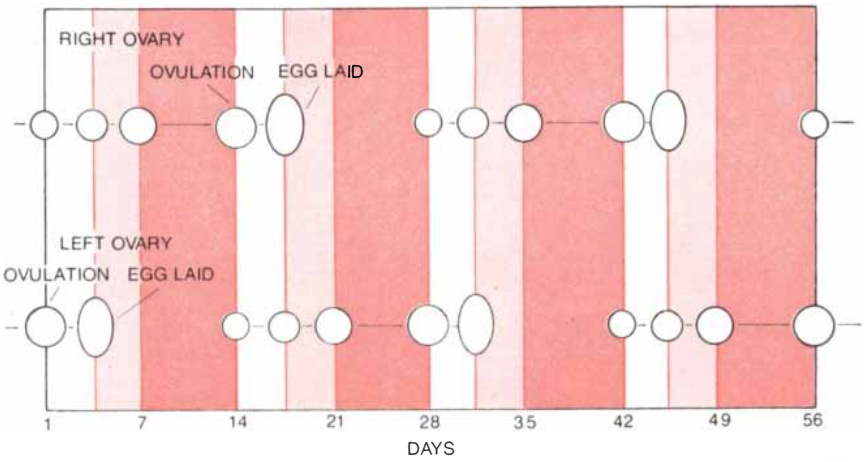
When either the area immediately forward of the preoptic area or the anterior basal hypothalamus is destroyed, the release of gonadotropin from the pituitary is shut off and the testes soon collapse. In the first instance the effect is probably on the cells that produce the releasing hormones; cells of this kind have been found in the forebrain of a large variety of animals. The anterior basal hypothalamus is where these hormones are transferred to the pituitary circulation.

It should be kept in mind that an animal's behavior is not solely a consequence of the action of hormones on discrete areas of the brain. A number of workers have demonstrated that in mammals and birds the feedback of peripheral sensations from the secondary sex structures is important in the coordination and completion of behavior patterns. The relation is evident in the mounting behavior of male lizards.

It is possible to determine which hemipenis a male lizard or snake is using by noting the direction in which the tail is turned. If the male is on the left side of the female, he will intromit the right hemipenis by curving his tail to the right, and vice versa. Experiments have shown that sensations from the hemipenises play an important role not only in the control of the initial orientation of the male during copulation but also in the termination of copulation. The finding is in agreement with a large number of studies of mammals showing that sensory feedback from the penis is necessary for normal mating behavior.

The behavior of the male that results from hormone-brain interactions has a major influence on the reproductive physiology of the female. If reproductively inactive females are exposed to a springlike environment when they are housed either alone or with other females, the growth of the ovaries is stimulated. If such a female is also exposed to a sexually active, courting male, the rate of ovarian development is increased significantly, indicating that the behavior of the male is facilitating the stimulative effects of the environment. Indeed, one finds not only that male courtship behavior is necessary for the normal secretion of gonadotropin by the pituitary in the female but also that the amount of gonadotropin secreted is correlated with the amount or frequency of male courtship behavior to which the female is exposed.

On the other hand, aggressive behav-



SEXUAL CYCLE of the female green anole lizard is represented schematically. The white regions indicate when the female is sexually unreceptive, the light color shows the period of onset of receptivity and the dark color represents the receptive period. The ovaries alternate in producing the follicle; its state of maturity is indicated by the size at which it is represented.

ior by males not only fails to facilitate the effects of the changing environment but also completely inhibits them. Females housed with males that are constantly fighting among themselves show little or no ovarian growth. To my knowledge this finding is the first demonstration that it is possible to turn female fertility off as well as on by varying the sociobehavioral environment.

In the female during the breeding season a single ovarian follicle develops and is ovulated about every two weeks. As in many human females, the ovaries alternate in the production of the follicle. Hence a sexually active female anole will typically have in one ovary a follicle that is in a more advanced state than the largest follicle in the other ovary. Moreover, like many mammals, a female *A. carolinensis* goes through cycles of sexual receptivity that are correlated with the pattern of follicle growth. The cycles presumably reflect corresponding fluctuations of sex steroids in the blood.

As one would expect, removal of the ovaries abolishes sexual receptivity in the female green anole. The administration of estrogen to such a female restores the behavior in a dose-related manner. The threshold dose appears to be .6 microgram of estradiol benzoate. A single injection of .8 microgram will induce receptivity in 85 percent of the females in 24 hours and in all of them in 48 hours.

In many mammals progesterone is important in the regulation of female receptivity. It does not function alone, but when it is administered at an appropriate time after the animal has been "primed" with estrogen, the two hormones act synergistically to facilitate receptivity. Studies of the pattern of hormone secretion by the ovaries in these animals reveal that this facilitation by progesterone coordinates the receptivity of the female with ovulation.

Progesterone appears to serve a similar function in lizards. For example, a female from which the ovaries have been removed will not be receptive to male courtship after a subthreshold dose of estradiol (.4 microgram) unless that dose is followed 24 hours later by 60 micrograms of progesterone.

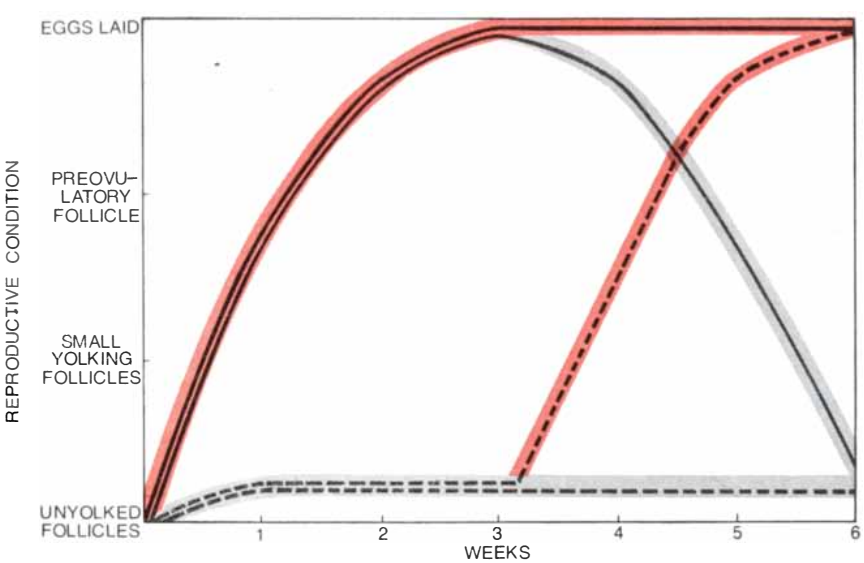
As in mammals, progesterone can have a quite different effect if it is administered simultaneously with estrogen or sufficiently long afterward. Then it inhibits the stimulatory effect of estrogen. In the lizard a single injection of 160 micrograms of progesterone 48 hours after an above-threshold dose (.8 microgram) of estradiol completely inhibits the effect of the estrogen.

With the female as with the male both central and peripheral stimuli are involved in the regulation of receptivity. When a female that is preovulatory or has been primed with estrogen is put in a cage containing a male, the male will immediately begin to court her. Less than a minute after mating, however, the female stops being receptive to male courtship. Experiments have demonstrated that intromission of the hemipenis is the trigger of the transformation. A preovulatory female that is courted and mounted but not mated continues to be receptive. The duration of intromission also seems to be important, since females interrupted during mating continue to be receptive until they are allowed to mate to completion.

We have found that the presence of the ovaries is critical in the turning off of receptivity. If females lacking ovaries but primed with estrogen are mated, they will be receptive again within 24 hours. Indeed, preliminary experiments indicate that such females are unreceptive for only about six hours after mating. Normal females, on the other hand, continue to be unreceptive for several days, suggesting that some change in the production of ovarian hormones is the key to the long-term inhibition of receptivity.

Just as the behavior of the male influences the reproductive physiology of the female, so the presence of females has a strong effect on the male. Males housed with females have a rapider testicular growth and sperm development than males housed with other males.

In sum, the sequence of events in the



INTERACTION of male behavior and female reproductive status in *A. carolinensis* is charted. Four groups of females were involved. One of them (*bottom curve*) was exposed only to aggressive male behavior. Another (*top curve*) was exposed only to male courtship. The curve that ascends after three weeks represents a group that was exposed to three weeks of aggression and three weeks of courtship; curve that descends after three weeks represents a group that received the opposite treatment. Periods of courtship are in color, periods of aggression in gray. Follicle condition (and hence the reproductive status of the females) appears at the left.

reproductive cycle of the lizard begins with environmental stimulation of testicular activity. Hormones produced by the testes as a result of this stimulation act on discrete areas of the brain to influence the secretion of gonadotropin by the pituitary gland and also to activate sexual behavior in the male. Depending on the male's behavior, the secretion of gonadotropin in the female, and hence ovarian growth, is either stimulated or inhibited. The sexual behavior of the female, which is dependent in part on ovarian hormones and in part on external stimuli, then influences male behavior and testicular activity.

The experiments I have described show *A. carolinensis* to be a convenient animal for investigating biomedically important phenomena that in mammals are difficult to isolate and dissect into their component parts, particularly for studies of the neurological pathologies underlying sexual dysfunction. The work has also revealed a certain degree of conservatism in the neural and endocrine control of reproductive behavior in vertebrates. For example, we have found that the areas of the brain concentrating steroid hormones are similar in lizards and mammals and have found that the function of these areas in regulating reproductive behavior is similar too. We have also found that progesterone has a distinctive role in the hormonal control of female sexual receptivity (either stimulating it or inhibiting it) in lizards as in many mammals, suggesting that this biphasic effect of progesterone is ancient in evolutionary terms.

It is noteworthy also that the many unique characteristics of lizards and the integration of laboratory findings with observations in the field lead to a clearer understanding of how behavior interacts with physiology in the adaptation of an animal to its environment. For example, the work suggests that the typical time sequence observed in many species between aggression and courtship, and the physiological mechanisms underlying these actions, may reflect common evolutionary origins. The opposite effects on the female of male aggression and courtship may in turn account for the pattern often seen in vertebrates of the Temperate Zone in which males establish territories in the breeding grounds before the females arrive. Females arriving early would be prevented from reproducing by the aggressive behavior among the males, whereas females arriving later would be rapidly stimulated by the courtship behavior of the territorial males. Finally, the termination of sexual receptivity by mating has several advantages for the female. She locks in her choice of a mate and decreases her vulnerability to predators by limiting the time spent in mating, thereby increasing the chance that she will survive to leave a large number of offspring.

THE AMATEUR SCIENTIST

Delights of forming water into sheets and bells with knives, spoons and other objects

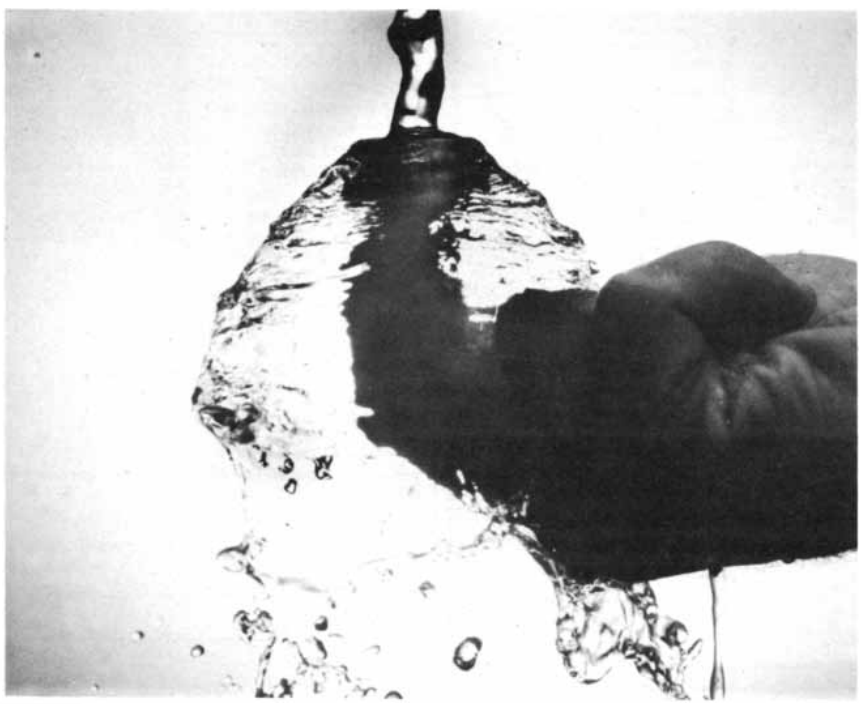
by Jearl Walker

One of the few pleasures I can find in the sheer drudgery of washing dishes is forming the thin films of water known as water bells or water sheets. For example, if I hold the flat part of a knife perpendicular to a stream of water, I can send thin sheets of water off to each side of the blade by adjusting the rate of flow and the distance between the knife and the faucet. Under the right conditions the sheets fold back on themselves and form lovely curved surfaces.

Spoons and other small objects that are either flat or slightly curved can also create sheets, but they may generate other shapes as well. A small circular obstacle can be made to send out sheets that fold back on themselves to form a symmetrical bell shape—a water bell. You may have seen similar sheets and bells in fountain displays. Sometimes the sheets in a fountain are so large that they are almost water sculptures. Although other factors figure in shaping the sheets, it is primarily surface tension that holds them together against disintegration and folds them back on themselves to form the bells.

The easiest way for me to form a bell is with a screw cap from a soft-drink bottle. The flat top surface is about the right size for a good bell and I can wedge two fingers into the bottom of the cap for support. To avoid spoiling the bell with my hand I reach up into the stream of water from below, positioning the flat part of the cap in the stream at the right distance from the faucet. The sheets fold back on my forearm to form a large bell.

Water sheets can also be made from a stream of water projected upward, which fountains usually are. A small obstacle in the stream then sends sheets of water to the sides, either upward or downward depending on the shape of the obstacle. You can make a water bell in an upward stream, as in a downward one, by adjusting the speed at which the water strikes the obstacle, but the shape of the bell may not be the same. If the obstacle has a concave surface, the sheets are directed downward to form a bell of about the same shape as one formed under a faucet. If the obstacle has a convex surface, the sheets are initially directed upward. Then they may curve over and come down. The bell may have a top that is more saucerlike than the rounded tops you will usually get in the sink.



A water bell made with a bottle cap

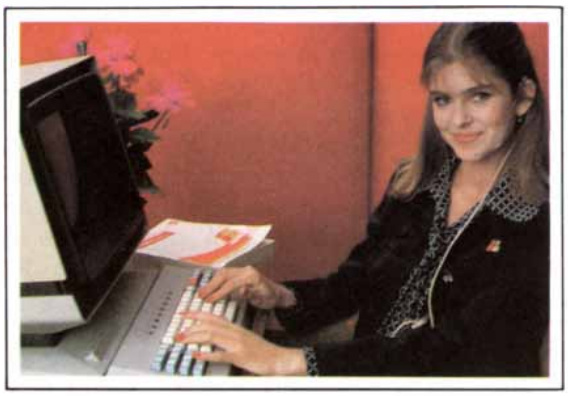
Water sheets have been studied since at least 1833, when Félix Savart wrote about them. By 1935 most of the principles determining the shapes of the sheets and bells were understood, although perhaps not in detail because of the difficulty of their mathematics. In the past 20 years or so interest in water sheets has revived because of the need to control fluid layers in such areas as spray painting. Particularly important is the way thin liquid films break up into drops. The maximum radius of a sheet in the sink is primarily determined by the surface tension of the water. The sheet disintegrates at the radius where the surface tension is no longer strong enough to hold the film together against the outward push of the moving water, that is, against the inertial force of the fluid as it flows radially outward from the impact point.

Except for the bottle cap the best obstacles for my work with water bells were the flat bottom of a 20-milliliter plastic beaker and a curved watch glass about six centimeters in diameter. The beaker was just large enough for me to squeeze in two fingers, so that I could make water sheets uninterrupted by my hands. With the watch glass I had a curved surface to investigate, but my fingers invariably disturbed the water sheets.

I first played with a normal falling stream of water, creating big sheets and bells that ended up to a large extent on my clothes and the kitchen floor. When the flow rate was fairly high, the sheets from the beaker were deflected to the sides with such momentum that they reached the disintegration radius before they had a chance to curve over. With lower impact speeds the sheets did curve before disintegrating. I got horizontal sheets at high speeds and nearly perfect water bells at lower speeds. The watch glass was less successful. The concave side yielded short sheets that were almost immediately spoiled by turbulence. The convex side sent water downward initially and generated better sheets.

To obtain an upward stream of water I attached a rubber hose to the faucet, directed the other end upward and held

Iberia Reservations. Electronics is just part of the game, not the game.



A sophisticated electronic reservations system is absolutely essential to an international airline. And being the second largest airline in Europe, in number of passengers flown, you can be sure ours is one of the most advanced in the world.

But at Iberia, we know that coordinating the flights of over 13.500.000 passengers a year takes more than electronics. It takes our colleagues at Reservations.

When they receive a request for a seat they don't think of you as a number. They think of you as a guest. And, in Spain, guests are friends.

So they put all their effort in trying to book you a seat, your connecting flight, or any other service. Because a computer can offer you an efficient flight, but only a friend can help you enjoy it.

That's why at our Reservations Department, electronics is just part of the game. Not the game.

This is Iberia today. But we want to be better.

IBERIA
INTERNATIONAL AIRLINES OF SPAIN

MORE THAN 50 YEARS MAKING FRIENDS.

How to figure the cost of living ...a longer life.

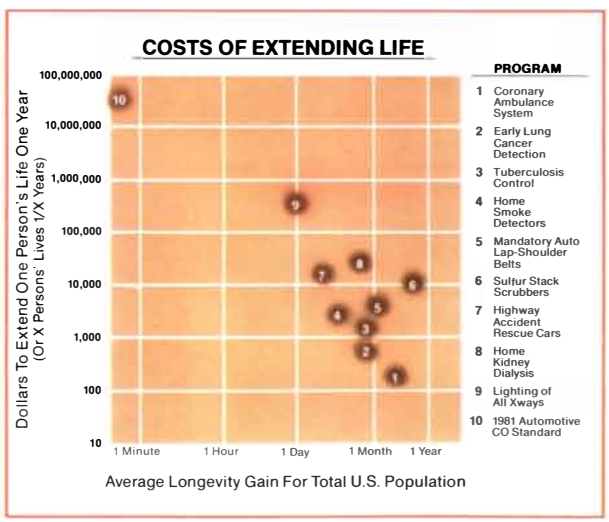
What's the fairest way of allocating the nation's limited funds to reduce various risks to human life and health? A loaded question, to be sure.

One way of evaluating a given risk-reduction program is to compare estimated costs with expected benefits, both measured in dollars. But this kind of analysis is controversial. For one thing, it requires placing a price on life itself.

Here at the General Motors Research Laboratories, societal analysts have developed a method which avoids that problem completely. It focuses on longevity and rests on the simple logic that since all life inevitably ends, no amount of risk reduction can *save* lives . . . only *lengthen* them.

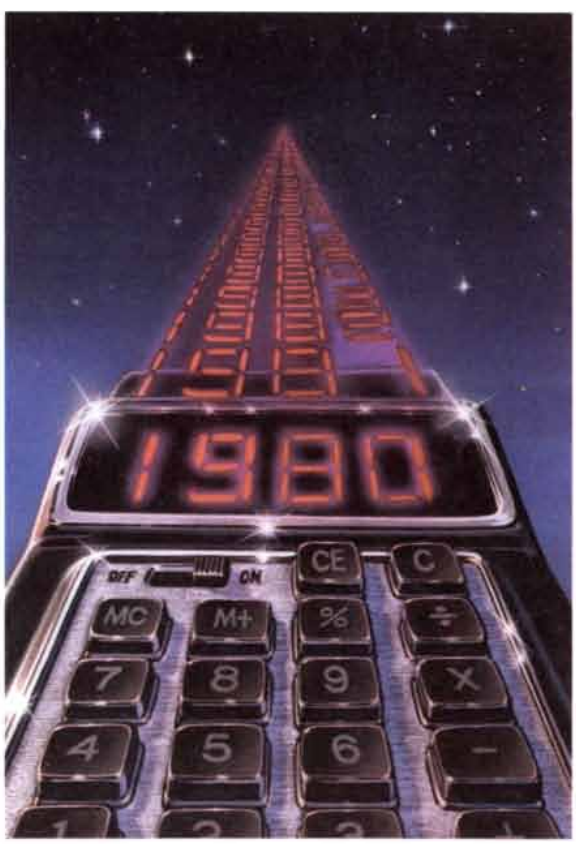
The method involves using the extensive data for all categories of mortality risks and determining the effect on longevity of each category independently. The results can be summarized for each risk by the equation: Average Years Of Longer Life = 0.2 x Annual Deaths Per Million Population.

This equation serves two purposes. First, it provides a perspective of days or years gained from risk-reduction programs. Second, combined with cost estimates, it helps rate the effectiveness of those programs.



To illustrate its utility, we performed a study to compare the cost-effectiveness of several medical, environmental, and safety programs presently under serious consideration. The chart above shows the extreme variation in the costs of extending life by implementing those options.

Through such unbiased comparisons, policymakers can obtain a clearer picture of which programs offer the greatest potential gain for a fixed budget and, thereby, have a better basis for decision.



GM

General Motors Research Laboratories
Warren, Michigan 48090

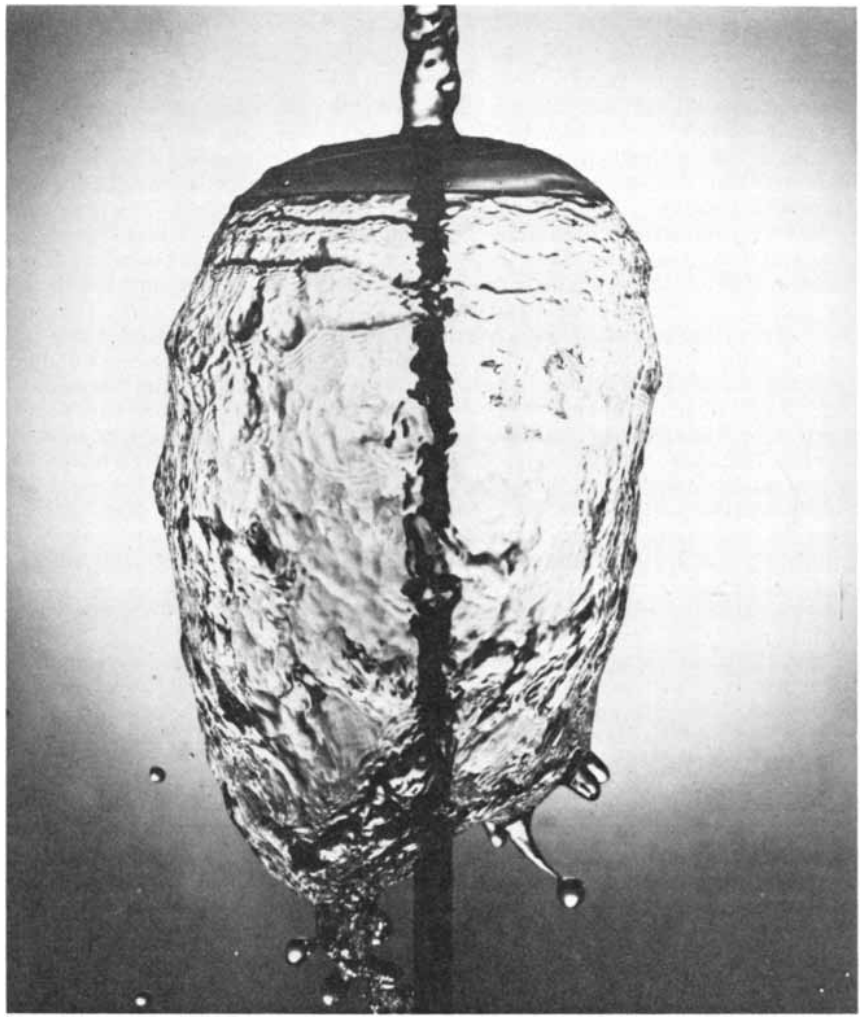
the rig in place with a clamp attached to a ring stand. The arrangement had the advantage of producing the water sheets near the basin of the sink so that they were less likely to wet me and the kitchen. The beaker made large sheets and bells, depending on how I adjusted the flow rate of the water and how much distance I kept between the open end of the hose and the bottom of the beaker. The concave side of the watch glass gave similar results. The convex side sent the water sheets upward; by adjusting the rate of flow I could vary the shape of the sheet as it curved over at its maximum height and began to fall downward. The turnover was either smooth and gradual or sharp and abrupt. When I put a fork in the upward stream, several water sheets were created, one sheet through each space between the prongs, with all the sheets angled off in different directions. In addition the portion of the stream hitting the concave section of the fork produced part of a water bell. The combined effect was quite lovely.

A good description of water bells was written in 1953 by Frank L. Hopwood, who studied their shapes at different rates of flow. He made bells by directing water through a brass tube that was about 10 centimeters long and 2.1 centimeters in external diameter. A coaxial rod at the top of the tube had fixed to it either a short cylinder or a nut. The water had to pass between the rod and the walls of the tube and then go out through the narrow slit between the top of the tube and the obstacle on top. By screwing either the cylinder or the nut up or down on the rod Hopwood could adjust the opening of the slit.

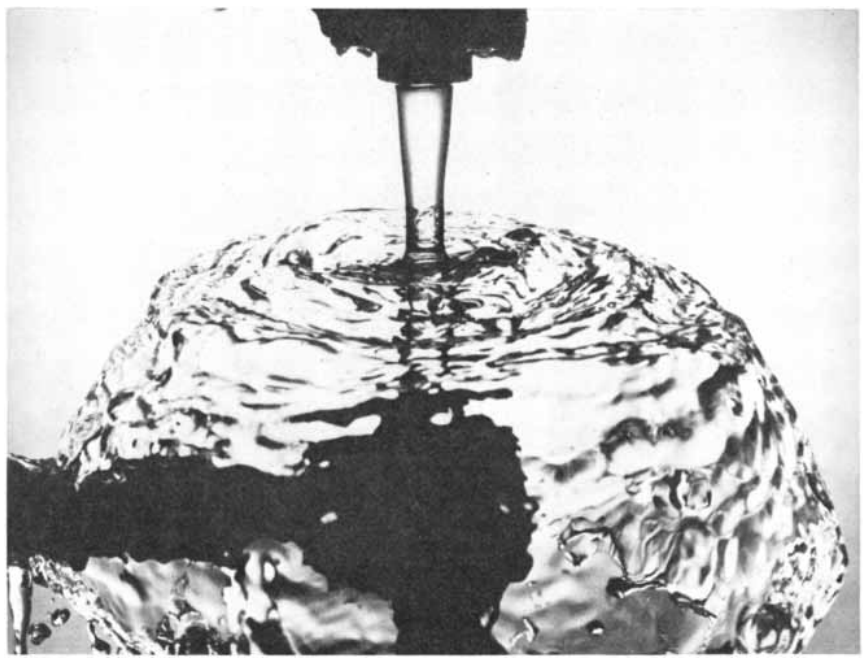
The tube was mounted vertically in a large basin that was filled with the falling water. An overflow pipe drained the additional water into a nearby sink. The flow of water from the faucet was controlled by a stopcock. With this setup Hopwood's water bells joined with the surface of the water in the basin, giving the bells a closed interior.

Hopwood created either round-top or saucer-top bells by adjusting the height of the slit above the water level in the basin, the width of the slit and the rate of flow. Some of the bells with saucerlike tops had sides that either bent inward or were straight but slanted. In some instances the edge at the point of turnover was quite sharp. At times the edge formed cusps, and Hopwood saw small drops being thrown over the edge. When he made a large bell and then slowly decreased the flow rate of the water, the bell generated a series of these stable shapes (sides either straight or concave) in addition to unstable ones as it shrank. When it had become quite small, it could assume the shape of a cylinder with straight vertical sides.

As Hopwood increased the size of a water bell he punctured the side with his



A watch-glass bell



Sheets formed by the coaxial collision of two jets

finger and found that the bell immediately doubled in size. Apparently the interior air pressure was lower than the atmospheric pressure outside the bell. If he punctured the side while he was decreasing the flow rate and the bell size, the result was a slight decrease in the size, implying that the pressure conditions were different.

To further investigate the air pressure inside a bell, Hopwood slowly blew bubbles into it by allowing them to rise through the water in the basin and into its interior. If the bell was initially round on top and gently curved, the extra air pressure inside increased the diameter near the base; the diameter at the top contracted. The sides began to curve inward, a rim rose at the top (giving the top a saucer shape) and eventually a fold or ditch formed along the perimeter near the base. These changes continued until the extra pressure inside the bell was enough to blow air out from under the base. Then the bell shrank to its original shape.

You can achieve similar results by running a flexible straw through the water in the basin and up into the bell. Gently blow air into the bell. You could substitute for the straw a small rubber hose that has a squeeze bulb at one end. With either setup you could also investigate the changes brought about by removing air from the interior of the bell. Does a bell with a round top simply shrink, showing no other change in shape?

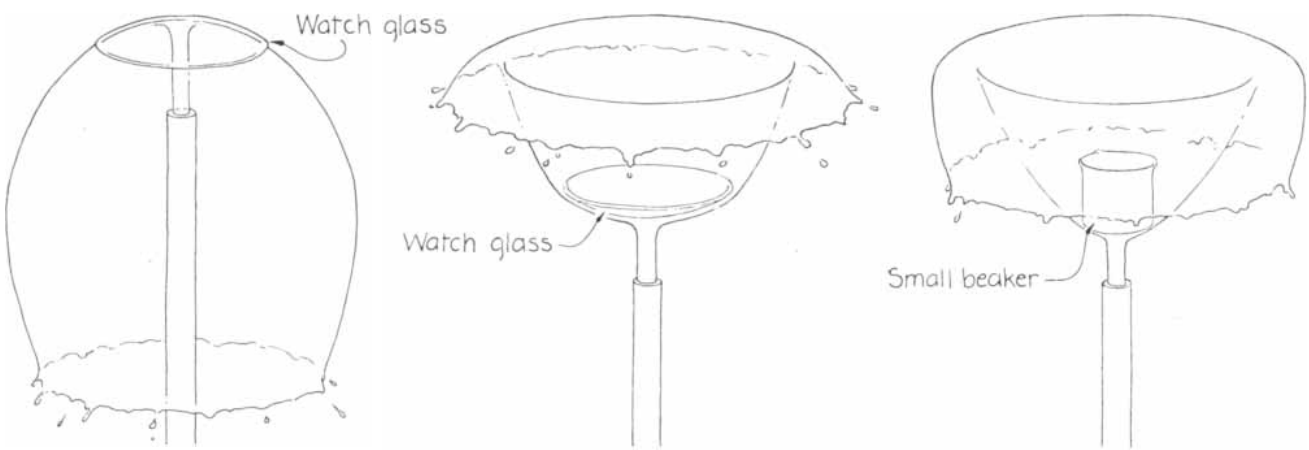
Hopwood suggested that two water bells, one inside the other, can be made with two slotted tubes that are on the same axial rod but are attached to independent water supplies. I made two incomplete water bells more easily with my curved watch glass. The hose from my faucet ran to a *Y* hose connector. One arm of the *Y* was connected to the hose I had already mounted vertically with a clamp and a ring stand. The other arm was connected to a hose that was similarly mounted on the ring stand but was higher up and pointed downward. I put clamps on each hose so that I could control the amount of water directed to each. When I turned on the water supply, I placed the watch glass in the stream. I could adjust the flow rate from each hose, the distance between orifices and the distance between the obstacle and each orifice.

The top and bottom surfaces of the watch glass each produced a water sheet. Usually the sheets merged at the rim of the watch glass. With some fiddling I could get the concave top to send upward a water sheet that was not affected by the water sheet sent down by the convex bottom. Except for the area where I held the watch glass I had two water bells well separated from each other.

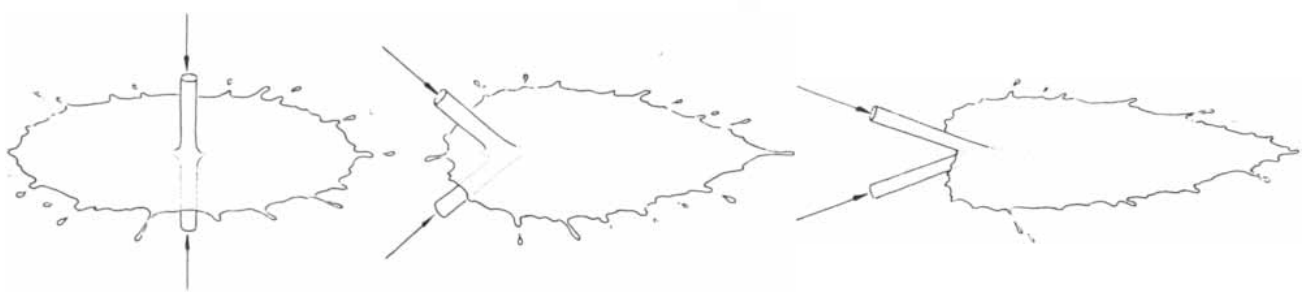
Twisting the watch glass made the two sheets touch and then merge. Separating them again took some work: the surface tension tended to keep them joined. Sometimes I could separate the sheets only in a small area on one side of the watch glass. Then a layer of air would be trapped between the two separated sheets. Nearby there might be large and beautiful distortions of the sheets with gentle curves up and down. I replaced the watch glass with a fork and found similar distortions of the multiple sheets wherever they merged.

The shape of a bell is determined not only by the difference in pressure between the inside and the outside but also by gravity, the viscous flow of the air inside the bell and the pull of surface tension. The effect of surface tension is twofold: a pull up and down along the meridian of the bell and a horizontal pull along the circumference. If the bell were spherical, the two pulls would be the same, but the different radii of curvature in the circumferential and meridional directions result in different pulls.

Whether you investigate water bells and sheets casually or seriously you are sure to have a lot of fun with them. You may want to substitute new shapes for generating the figures or to create water sculptures with multiple and merging water bells. Still another way to make curved water sheets is by letting water pour over an edge, as in a waterfall. A shopping mall in Chicago has a stepped



Examples of water sheets that can be made with an inverted stream



Shapes of sheets generated by colliding streams

waterfall at least one story high, and each step displays curved water sheets. A shape resembling a bell can also be made by allowing water to pour down a vertical cylinder so that the sheets leaving the bottom of the cylinder join at a point below it.

If you photograph water bells made in your sink, you may find that the camera picks up much turbulence you did not see. The only way to eliminate this drawback is to run the water first through a number of thin, narrow tubes and several sheets of wire gauze.

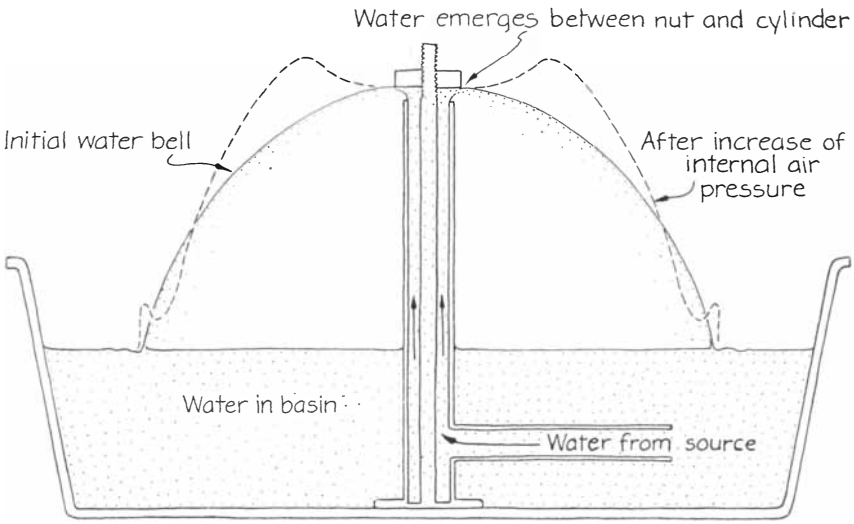
Water sheets and bells can also be created by having two laminar water jets collide either head on or at a glancing angle. Under the right conditions the water spreads into a sheet that can be either flat (out to the disintegration radius) or curved. By adjusting the flow rate and the diameter of each stream you can force curvature in either direction: the sheet curves away from the stream that has the greater momentum. The essentially flat sheets appear when the momentum of the streams is the same. At a suitable imbalance in the momentum the curved sheet folds back on a source to form a water bell.

This method of making water sheets has been employed ever since Savart first wrote about water bells. He calculated that the diameter of the sheet or bell is proportional to the pressure of the water creating the streams and is also proportional to the square of the diameters of the two openings through which the streams emerge. Experimental measurements put the disintegration radius just short of Savart's formula.

Extensive studies of colliding streams were done by J. H. Lienhard and J. C. P. Huang in the 1960's. They arranged for the streams to emerge from a large container of water some one foot to two feet deep. The depth could be adjusted by an overflow device that maintained a constant level of water. Two outlets on the side of the container consisted of tubes perpendicular to its wall. Precisely drilled holes were made in the projecting tubes so that when the container was filled with water, streams emerged from the holes toward each other. The arrangement was stable and had the advantage of making the speed of the streams easy to control by adjusting the height of the water in the container. The holes were also easily replaced with other holes of different diameters.

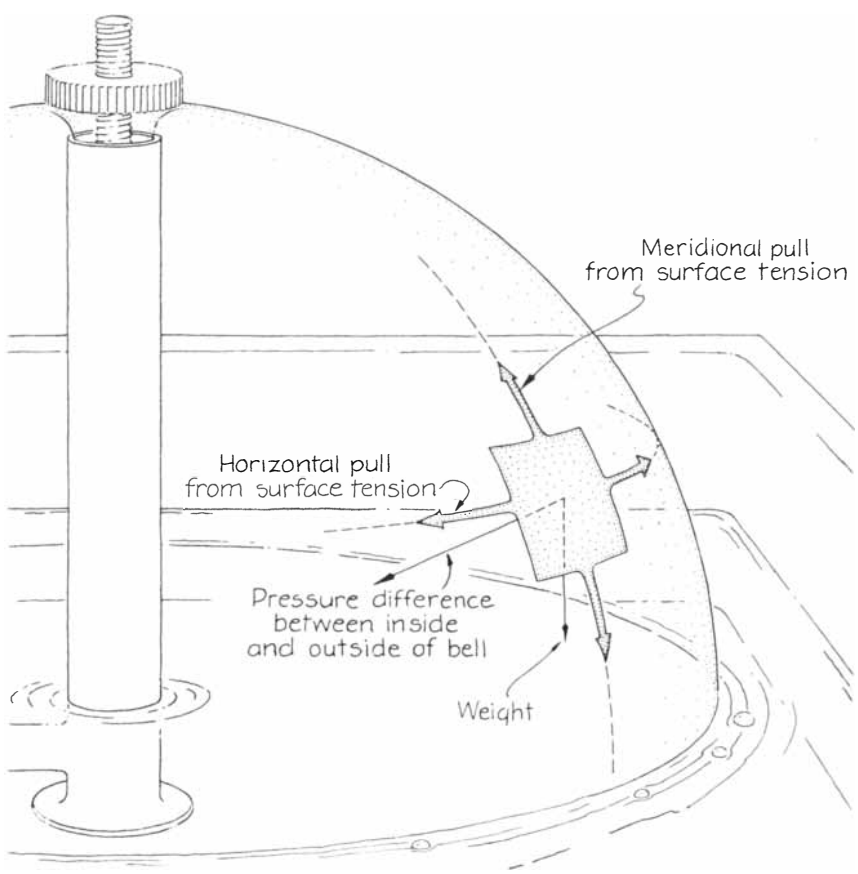
The water sheets can be made more readily but with far less control with my setup in which the water from a faucet is run through a *Y* connector and then the two streams are directed toward each other. I have little control over the diameter of a stream except by replacing the hose, but I can control the rate of flow by means of hose clamps. With this system water bells are easy to make.

The aspect of water sheets that most interests modern investigators is how the sheets disintegrate at their edges. The most intensive work on the subject was done by Lienhard and Huang on sheets they made with colliding streams. Some components of the disintegration can be seen clearly only with high-speed photographs or motion pictures, but others are visible in my homemade water sheets.



Change of the shape of a bell with increasing interior pressure

Lienhard and Huang investigated the disintegration in terms of a factor called the Weber number, which is a dimensionless number employed in studies of fluids. It is calculated by multiplying three quantities—the density and the diameter of the fluid stream and the square of its speed—and dividing the result by the surface tension. In general only the speed and the diameter can be



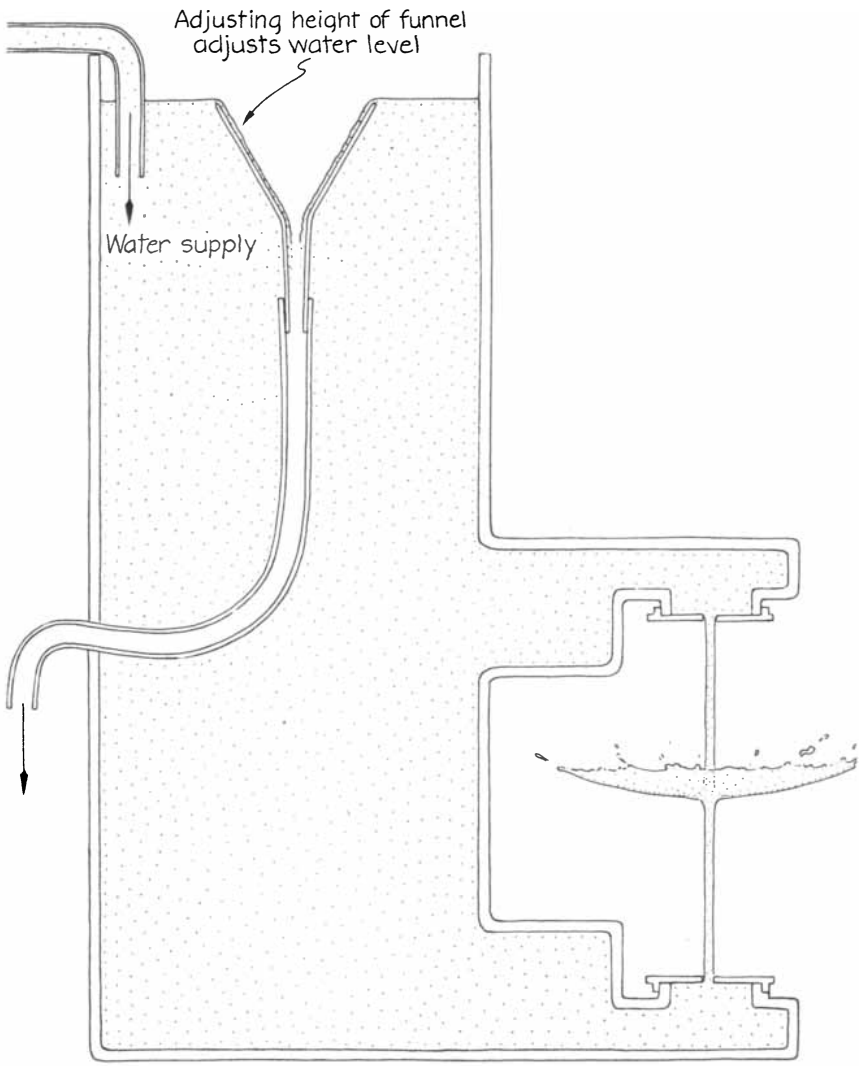
Forces on an area of the surface of a water bell

varied; with my kitchen water sheets the speed is varied by the water valve, and since a falling stream narrows, a diameter can be chosen at any point along the stream.

At low Weber numbers (values ranging from 100 to 500) the sheets are stable and have a nearly circular perimeter, with beads of water forming along the perimeter. High-speed photography reveals that tiny beads form on the edge and move along the perimeter, coalescing into larger beads that finally break away as drops. You can see the larger beads and the drops without any photographic aid, but the motion of the beads along the perimeter is too rapid to be perceived.

At somewhat larger Weber numbers (500 to 3,000) the perimeter of the sheets forms noticeable cusps that result from curved waves initiated near the impact point of the streams. The waves propagate outward over the sheet to set up the embroidery pattern along the perimeter. Sheets with the largest stable radii are in the range of Weber numbers from about 1,000 to 2,000. At the top end of the range the cusps diminish and the perimeter is again circular. Under the same conditions large waves can be seen moving radially outward from the collision point to the edge of the sheet. The waves are termed antisymmetrical because the fluid on opposite sides of the sheets moves in opposite directions, so that at some places along the wave the sheet thins and at other places it thickens. I can improve the visibility of these waves and the ones forming the cusps at lower Weber numbers by shining a high-frequency stroboscopic light on the water sheets. The antisymmetrical waves become larger in amplitude at still larger Weber numbers. By then the breakup of the edge is readily apparent because the sheet begins to flap like a flag in the wind.

With my simple sink setup for making streams collide I could change the angle of the streams to get a more glancing collision. As the two streams were angled closer together on one side the sheet formed on that side became steadily



Apparatus for producing colliding streams of water

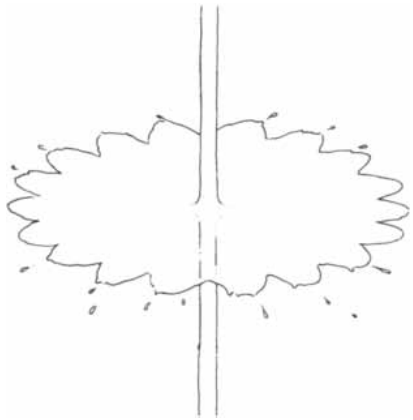
smaller because progressively more of the water was being thrown to the other side by the collision. When the angle between the two streams was reduced to 60 degrees or less, the sheet formed between them became quite short and the sheet on the other side grew considerably. The longer sheet was shaped like a leaf, being pointed at the far tip.

Returning to the water sheets produced by a stream directed onto an obstacle, I followed some work conducted by Sir Geoffrey Taylor (who has written several important papers on the dynamics of water sheets) by introducing a perturbation on the obstacle. With a knife I made a radial ridge on the bottom of one of my small plastic beakers. When the beaker is made to create water sheets in the middle range of Weber numbers, the ridge sends out curved waves similar to the ones that make the cusped perimeters in that range.

You might like to continue examining water sheets formed by obstacles or the collisions of streams. If you want to correlate the speed of the fluid with a sheet or a sheet's breakup mechanism, you will need to construct a fluid supply with a controlled pressure, as Lienhard and Huang have done. You may also want to investigate fluids other than water. Some of the non-Newtonian fluids I discussed in this department for November, 1978, may be particularly interesting, although you will have to avoid the ones with a viscosity so large that it prevents a flow adequate to produce a fluid sheet.

Roughly seven years ago Elizabeth Wood (known for her work with crystals and for her enchanting book *Science for the Airplane Passenger*) told me of another common sink phenomenon. With the water from the faucet flowing fairly slowly, hold the flat of a knife in the stream near the faucet. You may have to adjust the distance or the flow rate to see the effect, but usually the short length of the stream between the faucet and the knife exhibits ripples. The wavelength of the ripples depends on the speed of the water and the length of the falling stream.

For all seven years I occasionally searched for an explanation of the ripples. The answer finally came as Lienhard and I were discussing water bells for this article. He made an offhand remark about the effect and then pointed out that he had published the explanation in 1968 in the paper listed in the bibliography for this issue [*page 198*]. The explanation stemmed from work Lord Rayleigh had done in 1878 and 1879 on waves propagating along a liquid cylinder. What you see in the falling stream are the waves generated by the impact of the stream on the knife. They move upward along the water stream with the same speed at which the stream



A water sheet with a cusped perimeter

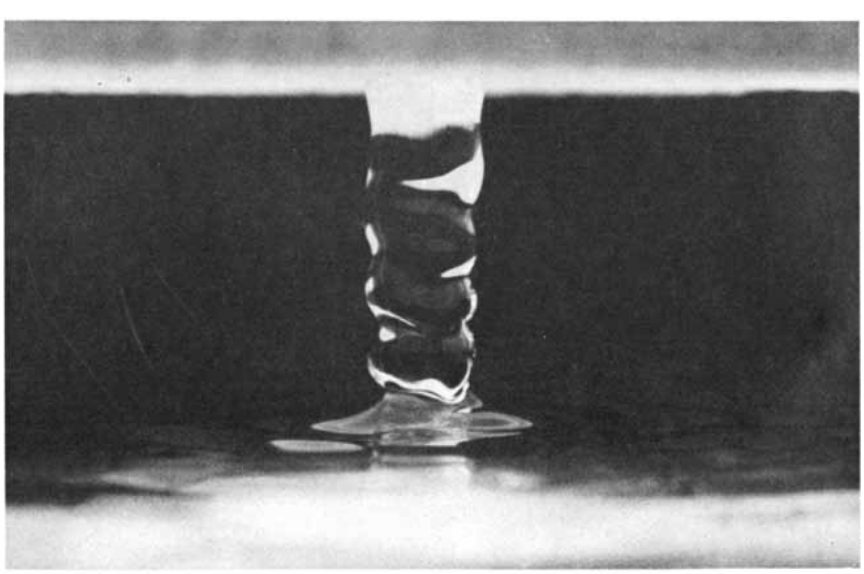
Waves on a falling stream

falls, and the result is that they look stationary to you.

To investigate these waves Lienhard arranged for a can full of water to leak through a small central hole onto the surface of water in a full martini glass. The diameter of the hole was .026 inch. The distance the water fell was both short and constant, but because the water level in the can slowly decreased, the pressure on the hole and the speed of the water also slowly decreased.

As the speed of the water draining from the can diminished, the correlation between the speed of the stream and the speed of the waves meant that the wavelength of the ripples had to increase. The wavelength finally reached an upper limit. Rayleigh showed that the water column would be unstable for any wave with a length exceeding the circumference of the column. The longest wavelength that can result in a stable wave in the falling stream is therefore equal to the circumference of the stream. When the speed of the water is low enough to result in instability, the ripples sometimes break up into droplets, sending some of them skimming across the surface of the water in the martini glass. (How such drops float above the bulk liquid was discussed in this department for June, 1978.) The greatest instability is in a wave with a wavelength about 4.5 times the diameter of the stream. It is therefore a larger wavelength than the largest one that gives a stable wave. Such a wave generates swellings in the stream that quickly grow larger until the stream breaks up into drops.

If you have access to a camera with a closeup lens, you can photograph the rippled stream as water slowly drains from a can. The wavelength of the ripples should be approximately inversely proportional to the distance between the water level in the can and the water level in the martini glass. With several closeup photographs you can measure the distance between adjacent ripples and plot their separation against the distance between the water levels. At the very least it is a good way to spend a hot summer afternoon.



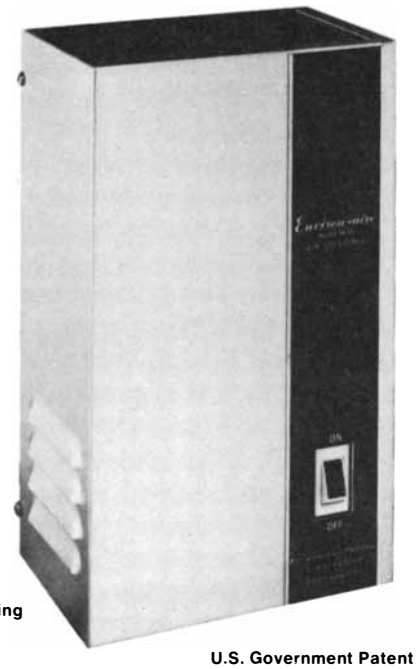
Waves on a stream emerging from a hole in a paper cup

Introducing . . .

The Fresh-Air Phenomenon

Guaranteed to Control Odors.

The *only* electronic air freshener listed by UL



Now, through this exclusive scientific breakthrough, *indoor* air can be purified in exactly the same way outdoor air is purified by lightning during a thunderstorm.

U.S. Government Patent No. 3,925,673

Think of that sweet "country fresh air" smell following a thunderstorm. That's actually the *absence* of odor. The electrical activity in the air created by lightning adds a small negatively-charged electron to each oxygen molecule in the air. Science has long recognized the miracle of this phenomenon known as *ionization*. Amazingly, these ionized molecules combine with any odor-bearing molecules in the air – destroying them.

The Environ-air unit's patented process electrostatically creates these same negatively-charged oxygen molecules that permeate any enclosed space and attract odor molecules like a magnet until they're totally "neutralized."

BEFORE ENVIRON-AIRE

There were two ways to control odors: *Dilute* – constantly recycle fresh air to dilute odor molecules or *deodorize* – constantly mask odor with a heavy perfumy scent. Now, with Environ-air, there's a third, more effective and highly scientific way. And that's to *destroy* all organic odor molecules electronically.

WHAT THE ENVIRON-AIRE IS NOT

It's not a one-shot cover-up. It doesn't mask odors – it destroys them electronically. And it kills tough odors *continuously* – not temporarily like sprays, wicks, stick-ons, etc. There are never any chemicals to buy, bulbs to burn out or filters to replace. The unit uses only as much electricity as a 25-watt light bulb.

WHAT IT IS

The Environ-aire is a *continuous* air-cleansing process so unique it's patented. It uses the latest in space-age electronic technology to create the first energy and cost-efficient air purification system for the home. Even in a smoke-filled room, you will be breathing only clean, fresh air all day long.

FORCED AIR – THE MOST IMPORTANT NEW FEATURE INNOVATION

One of the secrets of this new system is that it keeps the ionized air in *constant motion*. The Environ-aire is the *only* electronic air freshener that *continuously* "pumps out" ionized oxygen molecules into the air, permeating every square inch of space – ready to attack and deactivate any odor-causing molecules instantaneously. Keeping the ionized air circulating is such an important factor that so-called "space age" air fresheners without it are truly outdated. Without it, odor-killing molecules simply fall to the ground near the unit, rendering them useless against odors.

A DUST-FREE ENVIRONMENT: AN UNEXPECTED BONUS

The Environ-aire also de-activates dust particles suspended in the air. The ionized oxygen pumped into your room will attach itself to any impurities in the air, causing them to fall to the ground. Ahhhh . . . what's left is pure, clean air.

PROVEN EFFECTIVE IN HOSPITALS, FISH MARKETS, PET SHOPS & MORE

Prior to this special introduction to consumers, hundreds of Environ-aire units were used by businesses with their own peculiar odor problems. After using it for over 8 months they found it to be the only answer to annoying, persistent odors. Why? Because it outperforms other systems in what it does and how it does it.

TO USE IT IS TO LOVE IT

You will feel the difference immediately. The Environ-aire will cleanse any 20' x 20' room of odor-causing molecules within just 5 minutes! Imagine how exhilarated you'll feel with a fresh supply of ionized oxygen surrounding you day and night!

CLEAR THE AIR ONCE AND FOR ALL

In the kitchen . . . bathroom . . . basement . . . nursery . . . pet areas . . . smoke-filled offices. Use it anywhere stale, musty, offensive or pungent odors are a problem. The attractive wood-grain unit is compact – 10" x 6" x 4" deep – and lightweight – only 8 pounds. It can be wall mounted as an inconspicuous permanent fixture where annoying odors tend to accumulate. Or, it can be moved from place to place as needed, taking up little space on a shelf or floor. And installing it is simple – just plug it in. It uses regular household current.

LET YOUR NOSE PROVE ITS EFFECTIVENESS IN YOUR HOME OR OFFICE

The Environ-aire may sound too good to be true. That's why we offer a 30-day trial period and ask you to really give it a workout. For starters, simply turn the unit on, then cut up a big, juicy onion. No tears. No smell.

SOLIDLY BACKED

If anything goes wrong with your unit during the first year Environmental Electronics Corporation will repair it – without charge. Although the Environ-aire is built to last and be virtually maintenance free, it's still nice to know the manufacturer is service conscious.

SPECIAL INTRODUCTORY PRICE

The Environ-aire is manufactured for Cambridge International by EEC. We are offering this exciting new product directly to our customers exclusively through the mail for only $119.95 during our national introduction. Order one at no obligation today.

CAMBRIDGE INTERNATIONAL, Inc.

8700 Waukegan Road Morton Grove, Illinois 60053 (312) 966-5510

Order Now — Call Toll Free
800-621-5559
(In Illinois call 800-922-5858)
24 Hours A Day — 7 Days A Week

5 UNBREAKABLE PROMISES WE MAKE TO YOU

We promise that each exciting product we introduce to you has been carefully evaluated and judged by an independent panel of consumers and found to be:

1. A genuine value at a special price
2. Made of only the finest quality components
3. Covered by the manufacturer's own warranty *plus* our added 30-day money-back guarantee
4. Excellent enough to earn government approvals or UL listings (when applicable)
5. Developed from the most advanced technology available

Clip and mail to:
Cambridge International, Inc. Dept. A8
8700 Waukegan Rd., Morton Grove, Ill. 60053

☐ Please rush me _____ Environ-aire unit(s) at just $119.95 each plus $6.95 shipping & handling. (Ill. residents add 5% tax.) I understand that if I am not absolutely satisfied with my purchase, I may return it within 30 days for a prompt and courteous refund.

☐ Check or MO enclosed. ☐ Charge to: ☐ Visa ☐ Master Charge ☐ American Express

Account # _____ Exp. Date _____

Name _____

Address _____

City _____ State _____ Zip _____

Signature _____ 131

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

CONVERGENCE ON THE ARGAND DIAGRAM. L. W. H. Hull in *The Mathematical Gazette,* Vol. 43, No. 345, pages 205–207; October, 1959.

COMPLEX NUMBERS IN GEOMETRY. I. M. Yaglom. Academic Press, 1968.

THE HISTORICAL DEVELOPMENT OF COMPLEX NUMBERS. D. R. Green in *The Mathematical Gazette,* Vol. 60, No. 412, pages 99–107; June, 1976.

BACTERIAL TESTS FOR POTENTIAL CARCINOGENS

METHODS FOR DETECTING CARCINOGENS AND MUTAGENS WITH THE SALMONELLA/MAMMALIAN-MICROSOME MUTAGENICITY TEST. Bruce N. Ames, Joyce McCann and Edith Yamasaki in *Mutation Research,* Vol. 31, No. 6, pages 347–364; December, 1975.

PROPHAGE λ.INDUCTION IN *ESCHERICHIA COLI* K12 *ENVA UVRB:* A HIGHLY SENSITIVE TEST FOR POTENTIAL CARCINOGENS. Patrice Moreau, Adriana Bailone and Raymond Devoret in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 73, No. 10, pages 3700–3704; October, 1976.

ULTRAVIOLET MUTAGENESIS AND INDUCTIBLE DNA REPAIR IN *ESCHERICHIA COLI.* Evelyn M. Witkin in *Bacteriological Reviews,* Vol. 40, No. 4, pages 869–907; December, 1976.

PROGRESS TOWARD A TOKAMAK FUSION REACTOR

TOKAMAK DEVICES. L. A. Artsimovich in *Nuclear Fusion,* Vol. 12, No. 2, pages 215–252; March, 1972.

TOKAMAK RESEARCH. H. P. Furth in *Nuclear Fusion,* Vol. 15, No. 3, pages 487–534; June, 1975.

A PHYSICIST'S ABC ON PLASMA. L. A. Artsimovich. Mir Publishers, 1978.

RECENT PROGRESS IN TOKAMAK EXPERIMENTS. Masanori Murakami and Harold P. Eubank in *Physics Today,* Vol. 32, No. 5, pages 25–32; May, 1979.

A PRE-NEOLITHIC FARMERS' VILLAGE ON THE EUPHRATES

PAPERS IN ECONOMIC PREHISTORY. Edited by E. S. Higgs. Cambridge University Press, 1972.

FARMING IN PREHISTORY. Barbara Bender. St. Martin's Press, 1975.

THE EXCAVATION OF TELL ABU HUREYRA IN SYRIA: A PRELIMINARY REPORT. A. M. T. Moore in *Proceedings of the Prehistoric Society,* Vol. 41, pages 50–77; December, 1975.

THE NEOLITHIC OF THE NEAR EAST. James Mellaart. Charles Scribner's Sons, Inc., 1976.

CORALS AND CORAL REEFS

PROBLEMS OF GROWTH AND CALCIUM DEPOSITION IN REEF CORALS. T. Goreau in *Endeavor,* Vol. 20, No. 77, pages 32–39; January, 1961.

THE BIOLOGY OF CORAL REEFS. C. M. Yonge in *Advances in Marine Biology: Vol. 1,* edited by F. S. Russel. Academic Press, 1963.

REEF CORALS: AUTOTROPHS OR HETEROTROPHS? Thomas F. Goreau, Nora I. Goreau and C. M. Yonge in *The Biological Bulletin,* Vol. 141, No. 2, pages 247–260; October, 1971.

CORAL REEFS AND MOLLUSCS. C. M. Yonge in *The Royal Society of Edinburgh Transactions,* Vol. 69, No. 7, pages 147–166; 1974.

SUPERNOVAS AND STAR FORMATION

SUPERNOVAE. D. N. Schramm. D. Reidel Publishing Co., 1977.

DARK NEBULAE, GLOBULES, AND PROTOSTARS. Bart J. Bok in *Publications of the Astronomical Society of the Pacific,* Vol. 89, No. 531, pages 597–611; October, 1977.

OBSERVATIONAL EVIDENCE FOR SUPERNOVA-INDUCED STAR FORMATION: CANIS MAJOR R1. William Herbst and George E. Assousa in *The Astrophysical Journal,* Vol. 217, No. 2, Part 1, pages 473–487; October 15, 1977.

PROTOSTARS AND PLANETS. Edited by T. Gehrels. University of Arizona Press, 1978.

THE MATHEMATICS OF PUBLIC-KEY CRYPTOGRAPHY

THE CODEBREAKERS: THE STORY OF SECRET WRITING. David Kahn. The Macmillan Company, 1968.

NEW DIRECTIONS IN CRYPTOGRAPHY. Whitfield Diffie and Martin E. Hellman in *IEEE Transactions on Information Theory,* Vol. IT-22, No. 6, pages 644–654; November, 1976.

A METHOD FOR OBTAINING DIGITAL SIGNATURES AND PUBLIC-KEY CRYPTOSYSTEMS. R. L. Rivest, A. Shamir and A. Adleman in *Communications of the ACM,* Vol. 21, No. 2, pages 120–126; February, 1978.

PRIVACY AND AUTHENTICATION: AN INTRODUCTION TO CRYPTOGRAPHY. Whitfield Diffie and Martin E. Hellman in *Proceedings of the IEEE,* Vol. 67, No. 3, pages 397–427; March, 1979.

PROBLEMS IN PHYSICS WITH MANY SCALES OF LENGTH

THE RENORMALIZATION GROUP AND BLOCK SPINS: THE BOLTZMANN MEDAL ADDRESS. Kenneth G. Wilson. Publishing House of the Hungarian Academy of Sciences.

CRITICAL PHENOMENA IN 3.99 DIMENSIONS. K. G. Wilson in *Physica,* Vol. 73, No. 1, pages 119–128; April 1, 1974.

THE RENORMALIZATION GROUP IN THE THEORY OF CRITICAL BEHAVIOR. Michael E. Fisher in *Reviews of Modern Physics,* Vol. 46, No. 4, pages 597–616; October, 1974.

RENORMALIZATION GROUP METHODS. Kenneth G. Wilson in *Advances in Mathematics,* Vol. 16, No. 2, pages 170–186; May, 1975.

THE RENORMALIZATION GROUP: CRITICAL PHENOMENA AND THE KONDO PROBLEM. Kenneth G. Wilson in *Reviews of Modern Physics,* Vol. 47, No. 4, pages 773–840; October, 1975.

INTRODUCTION TO THE RENORMALIZATION GROUP AND TO CRITICAL PHENOMENA. Pierre Pfeuty and Gérard Toulouse. John Wiley & Sons, 1977.

THE HORMONAL CONTROL OF BEHAVIOR IN A LIZARD

PSYCHOBIOLOGY OF REPTILIAN REPRODUCTION. David Crews in *Science,* Vol. 189, No. 4208, pages 1059–1065; September 26, 1975.

BIOLOGICAL DETERMINANTS OF SEXUAL BEHAVIOR. Edited by John B. Hutchinson. John Wiley & Sons, Inc., 1977.

THE PARENTAL BEHAVIOR OF RING DOVES. Rae Silver in *American Scientist,* Vol. 66, No. 2, pages 209–215; March–April, 1978.

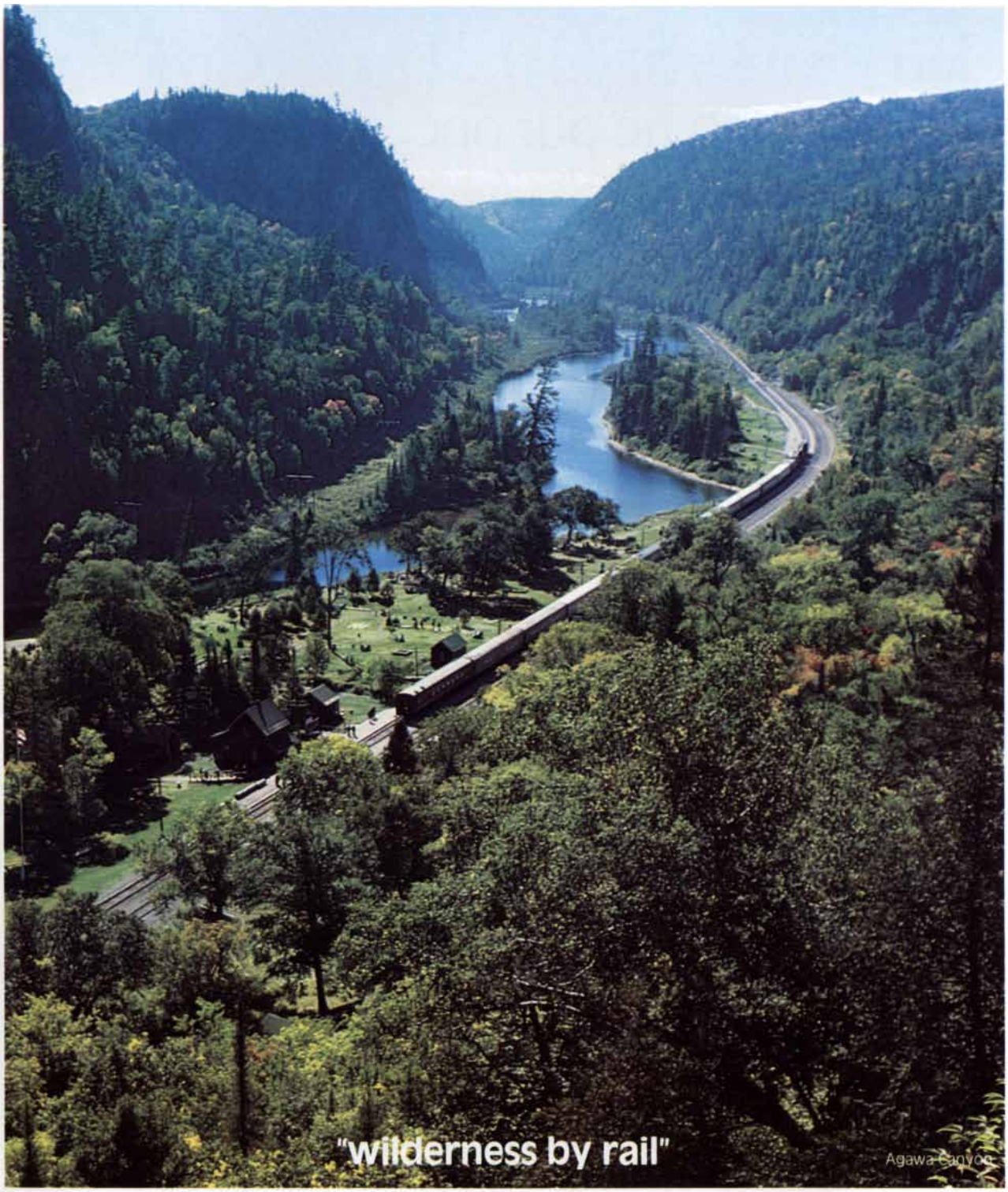
ENDOCRINE CONTROL OF REPTILIAN REPRODUCTIVE BEHAVIOR. D. Crews in *Endocrine Control of Sexual Behavior,* edited by Carlos Beyer. Raven Press, 1979.

THE AMATEUR SCIENTIST

WATER BELLS. F. L. Hopwood in *The Proceedings of the Physical Society,* Section B, Vol. 65, Part 1, No. 385B, pages 2–5; January 1, 1952.

CAPILLARY ACTION IN SMALL JETS IMPINGING ON LIQUID SURFACES. John H. Lienhard in *Transactions of the American Society of Mechanical Engineers, Series D, Journal of Basic Engineering,* Vol. 90, pages 137–138; March, 1968.

THE BREAK-UP OF AXISYMMETRIC LIQUID SHEETS. J. C. P. Huang in *Journal of Fluid Mechanics,* Vol. 43, Part 2, pages 305–319; August 28, 1970.



Take in the terrific terrain from the civilized comfort of one of our wilderness rail excursions. Pack up a picnic lunch and treat the family to a unique day-trip adventure. For all you need to plan your vacation, call us COLLECT (416) 965-4008, or write: Ontario Travel, Dept. W.R., Queen's Park, Toronto, M7A 2E5.

Ontario Canada

We treat you royally

Jerry Kibler liked the Honda Civic engine so much he put one in his airplane.

Jerry Kibler of Bellflower, California, has a refreshing hobby. He builds airplanes. Real airplanes. And after he builds them he flies around in them.

Naturally, we were pleased to learn that Jerry thought enough of our Honda Civic®1200 engine to use one of them in three of his airplanes, including the BD-5 you see here.

"It's basically a simple engine," he told us. "It's light and compact, but it has enough power to give me a top speed of 212 in the BD-5. In my three planes I've logged 230 hours so far, and the engine works just great."

Actually, the Civic 1200 is one of our earlier Civic engines.

Later on, when we introduced our Honda Civic CVCC® models, we gave them a new 1488cc engine. And more recently we designed a larger 1751cc engine for our larger cars, the Accord® and the Prelude.

Jerry Kibler modified his Civic 1200 engine in several ways in order to save weight and make it more suitable to an aircraft. He also added a turbocharger to increase its power. Which brings us to another point. Honda engines are designed strictly for cars, and we recommend that you do not put them in airplanes.

Nevertheless, we feel that Jerry Kibler's faith in our simple engine amounts to an impressive testimonial. Especially when you consider that Jerry was so pleased by the Honda Civic engine in his plane that he went out and bought the whole Honda Civic.

HONDA

We make it simple.

Jameson Irish

"Scotch and water."

If you like fine Scotch, you'll love light, imported Jameson Irish.

Try a glass of Jameson Irish the way you would your favorite Scotch. With water. Soda. On the rocks.

You'll notice how much it tastes like fine Scotch—only lighter and more delicate.

The dedicated Scotch drinker will instantly appreciate this flavor difference.

Though it may take a little time getting used to saying, "Jameson Irish and water, please."

Jameson. World's largest-selling Irish Whiskey.

80 & 86 PROOF • CALVERT DIST. CO. N.Y.C.