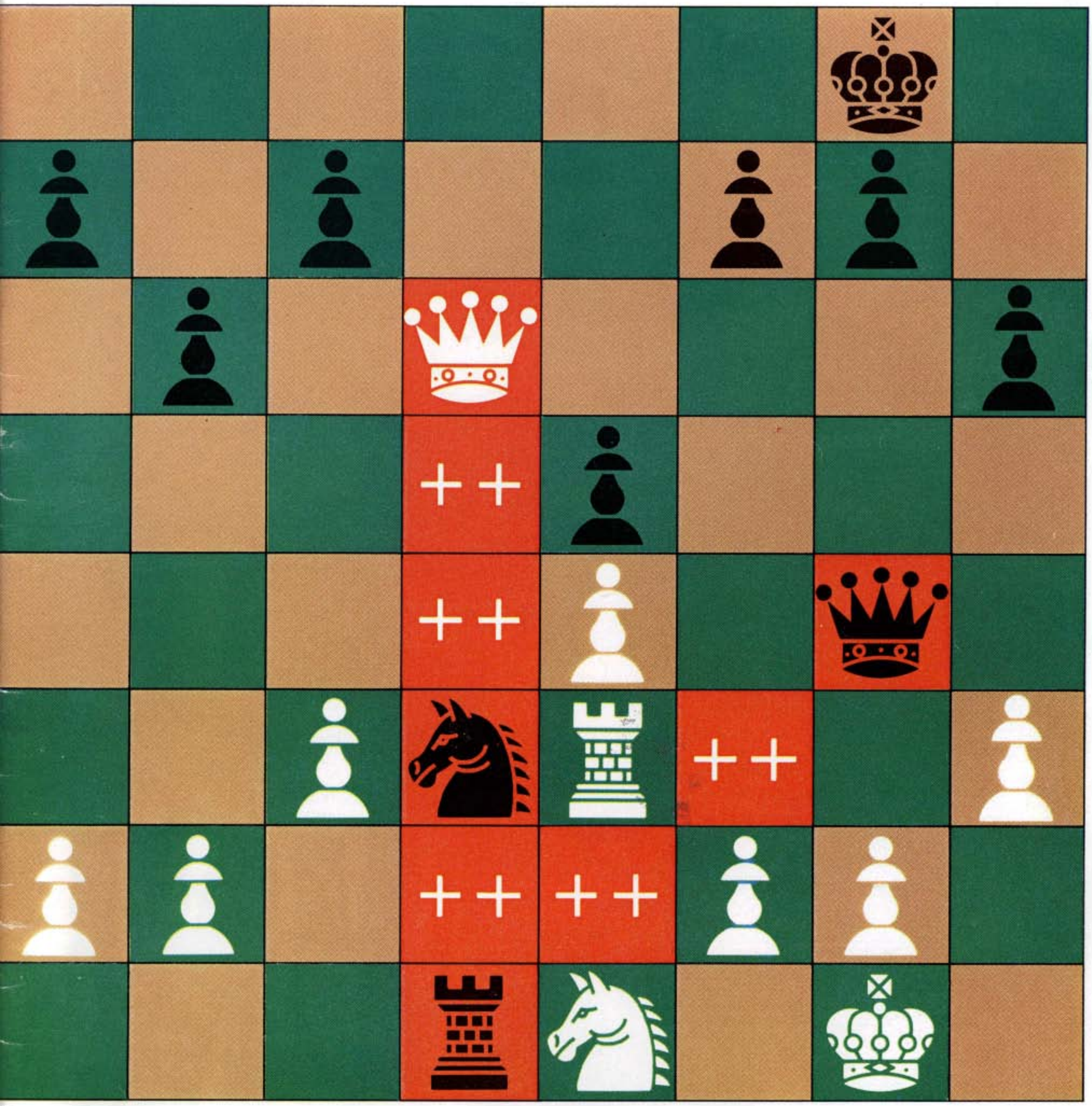
SCIENTIFIC AMERICAN



COMPUTER-CHESS SNAPSHOT

ONE DOLLAR

June 1973


"We don't push our luck."
UNIROYAL
THRILL DRIVERS
Uni: Maybe you think leaping off a ramp and flying 40 feet through space is daring. But for thrill show drivers like us, it's more a matter of confidence. And we feel confident on Uniroyal Steel Belted Radial tires.
UNI


©Uniroyal, Inc.
Roy: Actually, we scare as easy as anyone. But with Uniroyal's traction and responsiveness, the way they handle on the track, the way those steel belts hold up under stress, we figure the odds are in our favor.
ROY
AL
Al: So if you have some rough driving to do, like getting on the turnpike at rush hour, or stopping short in traffic, buy Uniroyals. Don't push your luck.
Uniroyal Steel Belted Radials

"A RECONSTRUCTED WILDERNESS."

Highlights of a dialog between John O. Logan, President of Universal Oil Products Company and Robert W. Johnson, President of UOP Realty Development Company.

Mr. Johnson: . . . so actually, John, we're making available 234,000 acres of our copper country land for this community project. That's an area six times larger than the District of Columbia!

Mr. Logan: I can see how this will stimulate the economy and growth of Michigan's entire Upper Peninsula, Bob. These plans are really exciting!

Mr. Johnson: The potential of this area is tremendous. Actually, all *we've* done is to provide a master plan that we think explores this potential to the fullest. Its development is up to the community. It will be *their* undertaking.

Mr. Logan: Apart from year-round recreational facilities, what will go into these other areas?

Mr. Johnson: There will be land for residential construction, resorts, retirement homes, camping grounds, probably a small airport, too.

Mr. Logan: You know, this whole area could one day be another Williamsburg or New Salem, with its exhibition center, craft shops, quaint stores.

Mr. Johnson: And notice, too, that huge tracts of virgin timberland have been permanently set aside for conservation and ecology. They'll remain absolutely untouched.

Mr. Logan: A reconstructed wilderness! One that will enrich the lives of generations to come. I'm delighted that we're a part of it!

Write for additional information. Ten UOP Plaza, Des Plaines, Ill. 60016

Established 1845 SCIENTIFIC AMERICAN June 1973 Volume 228 Number 6

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1973 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. 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: $10 PER YEAR, U.S., POSSESSIONS AND CANADA; $12.50 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O INTERNATIONAL DISTRIBUTORS, KOEPOORTBRUG 1, B 2000 ANTWERP, BELGIUM.

THE COVER

The painting on the cover is based on one of some 1,500 "snapshots" of hypothetical chess positions analyzed by an IBM 370/155 computer programmed to play chess by Albert L. Zobrist and Frederic R. Carlson, Jr., of the University of Southern California (see "An Advice-taking Chess Computer," page 92). The computer, playing White, is using a special routine for discovering "pins" in trying to select its 24th move against its "tutor," U.S. senior master Charles I. Kalme, a member of the U.S.C. mathematics department. Chess players will see immediately that the two queens occupy improbable positions. The computer, however, is concerned only with the disposition of pieces on the squares in red. The snapshot was generated on the basis of the board position after Black's 23rd move (. . . R-Q8). The computer hypothesizes a situation several moves ahead in which the white queen could pin Black's knight against the black rook, but it also sees that if the black queen could safely move to the square indicated (N5, in the usual chess notation), the pin would be foiled, or, as chess enthusiasts like to say, refuted. The double-plus symbol indicates hypothetical lines of attack. In other snapshots the computer sees that the black queen is not really needed to refute the pin, since the black knight could simply capture White's king bishop pawn (or queen knight pawn), thus occupying a square from which it could protect the rook. The computer uses such snapshots to make a preliminary selection of its most favorable moves. Eventually, after studying the situation for 40 seconds, it concluded that 24 Q-B1 (from QB4) was the best move. An analysis of the entire game begins on page 101.

THE ILLUSTRATIONS

Cover illustration by Jerome Kuhl

Page	Source
16–17	U.S. Bureau of Reclamation
20	Chemical Construction Corporation
23	C. P. Idyll
24–26	Tom Prentiss
27	Peruvian Ministry of Fisheries
28	Tom Prentiss
29	American Museum of Natural History
30–32	Hale Observatories
33–35	Allen Beechel
36	Vera C. Rubin
42	R. R. Alfano and S. L. Shapiro (*top*), F. W. Goro (*bottom*)
44–48	Jim Egleson
49	Jim Egleson (*top*), Bell Laboratories (*bottom*)
50–58	Jim Egleson
64	Milford Small
66	Burroughs Corporation (*top left*), Dan Todd (*top right*), Burroughs Corporation (*bottom*)
67	Dan Todd (*top*), Burroughs Corporation (*bottom*)
68	Milford Small
69–70	Dan Todd
71	Milford Small (*top*), American Micro-systems, Inc. (*middle*), North American Rockwell Microelectronics Company (*bottom*)
72	RCA Corporation (*top*), Wagner Electric Corporation (*bottom*)
73	Sigmatron, Inc.
74	William C. Denison
76–79	Eric Mose
80	William C. Denison
83–84	Curtis Wilson
85	Joseph Feldman
86–90	Bunji Tagawa
92	Albert L. Zobrist and Frederic R. Carlson, Jr.
94–105	Jerome Kuhl

The new Wang 2200. Some people call it a mini; others call it a calculator.

For under $7,000, call it a miracle!

Plug in the new Wang 2200 and you're in business. If your data handling has been constantly delayed or reshuffled by crowded, expensive time-sharing . . . here's a cost/efficient alternative. If you are thinking of time-sharing to meet growing needs; here's a way to beat high terminal costs and programming time. If you run a smaller engineering or technical/scientific group; the Wang 2200 may be all the computing power you will ever need.

BASIC Programming With Hard-Wired R/O Memory: You get 4K (field expandable to 32K) of unimpaired memory. And BASIC programming; the easiest to work with language yet developed. And it's *pure* BASIC, not some manufacturer's version that really requires learning a new language.

Programming Doesn't Byte Off Much Storage: In the Wang 2200, program statements are a single keystroke and single byte. Example: The statement "Print Using" takes only one byte of memory; not ten! This almost doubles capacity compared to the usual system.

Big 16 Lines (of 64 Characters) CRT Display: It's hard to crowd this tube. You'll get ample space for monitoring programs and data . . . it "rolls" forward or backward on command . . . remaining storage is automatically displayed as the last line on every program. It tells you where you've been, where you are and how far you can go. You can go a long way on the 2200.

Read/Write/Modify Programs On Cassette Tape. This could put big tape consoles to shame. With our optional cassette drive, you can search programs and data files by name and read, write and up-date in place. This usually requires two tape drives to carry off. Additional Tape Drives may be added to increase the file handling capability.

Two Keyboards, Two Printers and Other Goodies: Our 2200 lets you tailor a system to your needs; not vice versa. Take either an Alpha keyboard or a traditional calculator (with alpha) arrangement. Your choice. We have two fast printers and both handle carbon forms. There's even a plotter typewriter and a graph plotter that prints letters too.

For under $7,000, Try To Match The Wang 2200 For Price/Performance and Flexibility. Even if you call the 2200 a small system, you've got to admit it's a big idea. For the price, you can't match the storage or its power in handling tough problems. Compared with anything on the market today, it's cheap, fast, powerful and flexible.

A Low Cost Alternative To Time-Sharing: Since the 2200 can tackle big problems, many users are installing them to replace more expensive time-sharing terminals. And, they are eliminating the waiting, scheduling and priority problems of time-sharing. They're getting more done at lower cost.

Faster Problem Solving: "Basic" programming on the 2200 is almost literal. You get to the heart of problems and examine alternatives faster. And, the Wang *special function* feature gives you 32 additional keys that allow you to "tailor" the system to facts, figures and constants that are particular to your business or your company . . . or your department.

Drives Wide Range of Peripherals: Our new 2200 gives you more ways to solve problems. We have 15 (more soon) peripheral devices like flat bed plotters, additional mag tape cassette drives, a typewriter that makes graphs and a 3 million byte high speed disc.

It's Got The Price, Performance and Delivery: If you need more calculating and computing power now or if you want to examine cost/efficient alternatives to time-sharing, the new Wang 2200 system could be your answer.

For More Information

☐ Please send me full details on the Model 2200.
☐ Please ask my representative to call.

Name__________

Title__________

Company__________

Street__________

City_____ State_____ Zip_____

Wang Laboratories, Inc.
836 North Street
Tewksbury, Mass. 01876 SA 6
Tel: (617) 851-4111 Twx 710-343-6769

WANG

LETTERS

Sirs:

May I, through the courtesy of your columns, thank those who have commented on my article "Bicycle Technology" [SCIENTIFIC AMERICAN, March]. I hope they will forgive me if I do not reply to them individually or quote them in full but rather pick out those points that have aroused particular interest.

The chart showing the energy consumption of various modes of transport with respect to body weight has prompted several comments. Thomas D. Moder and others have suggested that teaching a horse to ride a bicycle would produce the most economic mode of transport. A similar idea occurred to a Mr. Brandreth of Liverpool in 1829 at the time of the Rainhill trials of locomotives for the Liverpool & Manchester Railway. His "Cycloped" was powered by a horse driving a moving belt geared to the wheels, but alas the "Rocket" was chosen in preference.

Other correspondents have suggested other modes of transport of greater energy efficiency than the .15 calorie per gram per kilometer I suggested for a man on a bicycle. W. L. Kitterman and F. C. Schwenk have calculated .1 for an airship at a speed of 150 kilometers per hour. James B. Calvert has stated .05 for a railway train and suggested even lower figures for large, slow ships and, of course, sailing vessels.

One point to bear in mind in all such comparisons is the restriction on body weight, so that although a railway train and a ship may be efficient on a basis of total weight or per ton of cargo, they may not prove so efficient per kilogram of passenger weight. I believe there is need for much further study and comparison of the energy efficiency of different transport systems, and it is hoped that such a study will be started at Oxford in September. Any references or data on this topic would be welcome.

Jules Finker has queried the efficiency of the rotary pedaling action as compared with the to-and-fro action used by Bob Bundschuh of Windsor, Conn., in a pedal car. Many of the early bicycles used treadles or ratchets, including Macmillan's of 1839 (page 82 of my article), the Coventry lever tricycle of 1876 (page 84) and the U.S.-built "Star" of 1885 (page 85). Still, the ubiquitous sprocket-and-chain drive triumphed and has remained unchallenged. Comparative tests ought to be able to quantify any differences. I believe the rotary pedaling is probably superior because of the uniform motion of the feet, thereby obviating the "shadowboxing" type of energy wastage that occurs in any reciprocating mechanism at either end of the stroke.

Lee Dembart has pointed out that Glenn Curtiss, the pioneer U.S. airplane builder, started with bicycles and motorcycles. I have since found that a pioneer in a different field was Dan Albone of Biggleswade in England, who in 1886 made the successful "Ivel" cross-frame racing safety bicycle and in the same year the first tandem bicycle. In 1902 Albone introduced the first successful motorized tractor, also called the "Ivel."

I am indebted to Professor E. D. Hirsch, Jr., of the University of Virginia for drawing my attention to a letter of John Keats dated 1819, in which Keats describes a hobbyhorse under the name "Velocipede," the name adopted much later by the Michaux brothers for their machines.

S. S. WILSON

Oxford, England

Scientific American, June, 1973; Vol. 228, No. 6. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

Advertising correspondence should be addressed to Harry T. Morris, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 25 cents.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25
All other countries:
☐ 1 year, $12.50 ☐ 2 years, $23 ☐ 3 years, $32

NAME

NEW ADDRESS

OLD ADDRESS

Conversation Pieces

Technically intriguing items from TRW, guaranteed to add luster to your conversation and amaze your friends.

Is anyone out there? The question of whether life is unique to planet earth has long fascinated the mind of man. As early as the 4th Century B.C., the Epicurean philosopher Metrodoros said, "To consider the earth as the only populated world in infinite space is as absurd as to assert that in an entire field of grain sown with millet, only one grain will grow." In the sixteenth century the heretic Giordano Bruno announced that "innumerable suns exist" and "innumerable earths revolve about these suns . . . Living beings inhabit these worlds."

Harvard astronomer Harlow Shapley approached the problem statistically. Of the 10^{23} stars in the universe, said Shapley, it is probable that 100 million have planets similar enough in composition and environment to earth to support life.

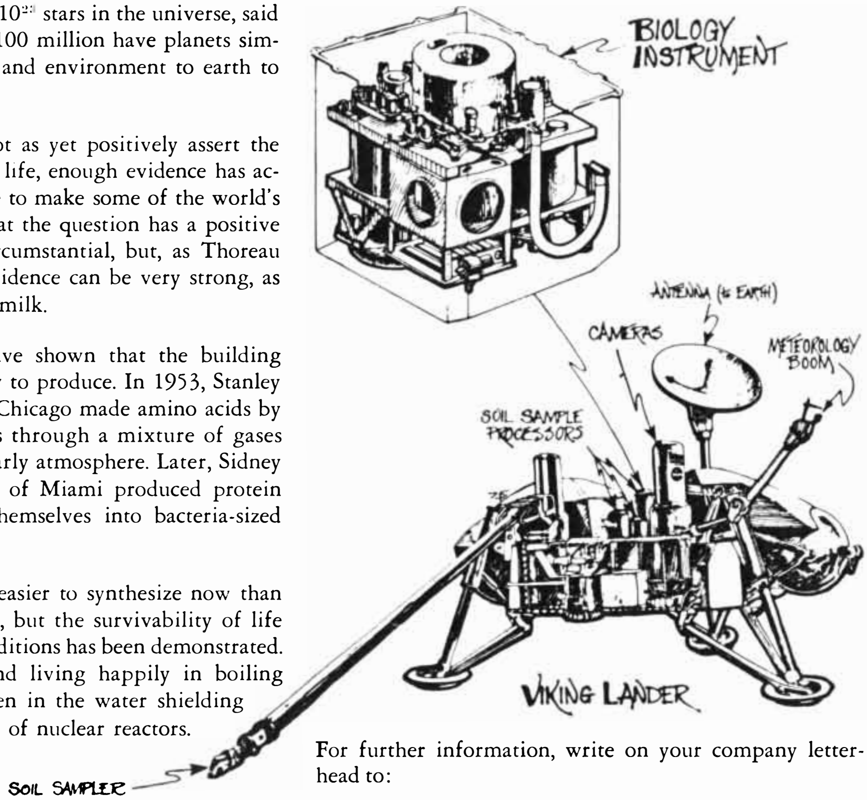
Although scientists cannot as yet positively assert the existence of extra-terrestrial life, enough evidence has accumulated in the last decade to make some of the world's leading scientists suspect that the question has a positive answer. The evidence is circumstantial, but, as Thoreau said, some circumstantial evidence can be very strong, as when you find a trout in the milk.

First of all, scientists have shown that the building blocks of life are rather easy to produce. In 1953, Stanley Miller of the University of Chicago made amino acids by passing electrical discharges through a mixture of gases that simulated our earth's early atmosphere. Later, Sidney Fox of Florida's University of Miami produced protein fragments which formed themselves into bacteria-sized spheres.

Not only does life seem easier to synthesize now than it did twenty-five years ago, but the survivability of life under presumably lethal conditions has been demonstrated. Organisms have been found living happily in boiling water, strong acids, and even in the water shielding the highly radioactive cores of nuclear reactors.

Recently, biologists have simulated the biologically rigorous conditions of the Martian environment in "Mars jars." Some of the organisms placed in the jars readily adapted themselves to the extremely cold carbon dioxide atmosphere.

As a subcontractor to Martin-Marietta, TRW Systems is at work on a NASA project which may shed some light on the question of extraterrestrial life. We are building the Viking Lander Biology Instrument, three tiny, fully automated laboratories, which will be landed on the Martian surface in 1976. On earth, these laboratories would occupy several rooms and a full crew of scientists. We are shrinking them into a single foot of space. We hope you're interested as we are in what Viking will find.



For further information, write on your company letterhead to:

TRW
SYSTEMS GROUP

Attention: Marketing Communications, E2/9043
One Space Park • Redondo Beach, California 90278

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1923: "As long as prohibition was not in force there was comparatively little danger of careless people obtaining rapidly poisonous alcohol in the place of the slowly poisonous variety in drink, commonly known as grain alcohol or merely alcohol. But since the prohibition law has been in effect, and since a great commerce in intoxicating liquors has arisen through the efforts of the bootleggers, each and every individual who imbibes such liquors is open to very grave danger of poisoning, blindness and death. The manufacturer of grain alcohol has always been under governmental control. Intoxicating drinks were formerly highly taxed, but industry cannot afford to pay the tax on grain alcohol. In order to avoid being subjected to the high tax, industries denatured the alcohol, rendering it unfit for human consumption. This was accomplished by adding various substances, formerly consisting of wood alcohol and benzol for the most part. Since prohibition came into effect, it was almost inevitable that attempts would be made by bootleggers, unscrupulous manufacturers and chemists to remove the denaturants from the alcohol and make intoxicants from it by coloring it with a little burnt sugar or caramel. But it is impossible to remove them in their entirety from the alcohol. There is not a single denaturant that does not have a very bad effect on the human system. The person who drinks bootleggers' concoctions is open to all these dangers, and the only sure way to avoid them is to avoid the concoctions."

"The consumption of electrical energy per capita is greater than 125 kilowatt-hours each year for 10 nations. Two small nations, Norway and Switzerland, and one large one, Canada, use current more liberally than we do in the United States, and it will be observed that all three of them are countries in which hydroelectric power is abundant. Niagara's power alone, divided on a basis of approximate equality between Canada's nine millions and our 110 millions, would make a huge difference in per capita consumption of the electric current between ourselves and our northern neighbor. In 11 countries more than 15 per cent of the population lives in electrically lighted homes. Again Canada is in the lead with 38.3 per cent of its population, and the United States is close behind with 36.8 per cent; next runners-up are Japan and Denmark with just under 30 per cent."

"In a lecture delivered before the Geological Society of London, Prof. Arthur S. Eddington says, regarding the question of the evolution of the earth-moon system: 'After the evolution of the solar system, we naturally turn to consider the evolution of the earth-moon system. My impression is that nothing in recent progress suggests any doubt that the beautiful theory of Sir George Darwin is correct. The main features are that the moon at one time formed part of the earth and broke away. At that time the rotation period of the earth was between three and four hours. That period has since lengthened to 24 hours, owing to frictional dissipation of energy by lunar and solar tides, and the back-reaction of the lunar tides on the moon has caused the moon to recede to its present considerable distance. Modern research has enabled us to calculate the magnitude of this tidal friction at the present time, and to estimate the date of the moon's birth to some 1,000 million years ago.' "

SCIENTIFIC AMERICAN

JUNE, 1873: "Various sanitary reform projects are being undertaken to utilize refuse so as to transform it from a source of expenditure to one of valuable profit. In large cities the question of obtaining the fertilizing matter from the material is subordinate to the problem of disposing of the noxious substance in such a manner that it shall not breed disease or nurture pestilence in the narrow and confined limits of thickly populated districts. The death rate from zymotic diseases alone averages 9,000 a year within the corporate limits of New York City, and the cases of sickness aggregate from the same period at least 100,000. It requires but a casual stroll through certain portions of the city to determine the reasons for this fearful mortality. Old marshes that, in the overwhelming desire to raise grades and make streets, were filled up with sand and stone, have asserted their existence and converted the land into a sponge, absorbing the filth which flows from the faulty sewers. Burying a nuisance is not abating it. It is an urgent necessity to thoroughly overhaul and, if need be, entirely alter our sewerage."

"Professor Charles A. Young of Dartmouth College has recently made some interesting experiments in substituting a finely ruled metallic plate for the prisms in a solar spectroscope designed for the observation of the solar prominences. The grating was ruled by Mr. Rutherfurd, the lines being 1/6,480 inch apart and the ruled surface covering something more than a square inch. Professor Young says that the dispersion is nearly the same as would be given by four prisms. The outline of the chromosphere and the forms of the prominences were as well seen, both in spectra of the first and third order, as with the ordinary instrument. President Morton of the Stevens Institute suggests that ruled plates could be duplicated by electrotyping and thus could be readily furnished by opticians at a price far below the expense of a train of prisms."

"Various methods are resorted to for the purpose of adulterating milk. The adulterating substances are principally water, flour or starch, gum arabic or dextrin, cerebral matter, chalk or whiting, turmeric or annatto, gum tragacanth, carbonate of magnesia, arrowroot, sugar, emulsions of almonds or hempseed, carbonate of soda, eggs and salt. It is no secret to milkmen that the specific gravity of milk is not much changed if 4 per cent of water be added for every 1 per cent of cream abstracted. Cerebral matter is usually some villainous mixture of the brains of sheep, employed to counteract the blue tinge of impoverished milk. Chalk or whiting is sometimes employed to neutralize the acidity of soured milk. Turmeric or annatto is used to give a rich cream color. Sugar, in the form of caramel or brown sugar, is used to add to the color and develop the flavor of impoverished milk. It is estimated that the yearly supply of milk for the city of Boston was 8,763,285 gallons, the cost of which to consumers may be reckoned at $2,979,517. If we assume the average amount of water, fraudulently added, to be but 12 per cent, and this is putting it at a low figure, the amount expended by the citizens during this year for water, apart from the legitimate water rates, amounted to $357,542."

THE AUTHORS

NOEL DE NEVERS ("Enforcing the Clean Air Act of 1970") is professor of chemical engineering at the University of Utah, where he joined the faculty 10 years ago after working for five years with the research arms of the Standard Oil Company of California. He recently worked as a staff engineer of the air pollution office of the U.S. Environmental Protection Agency during a year's leave of absence from the university. He writes that the experience gave him "one more data point for my favorite generalization: The only simple objects are the ones you don't know much about." De Nevers obtained his bachelor's degree at Stanford University in 1954 and his master's and doctor's degrees from the University of Michigan in 1956 and 1959 respectively.

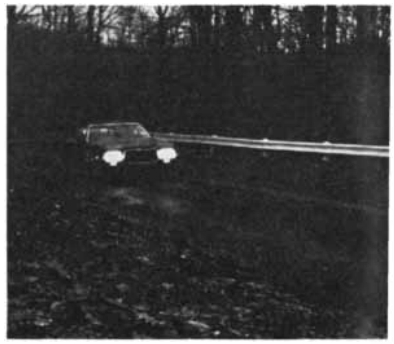
C. P. IDYLL ("The Anchovy Crisis") is fishery research adviser with the department of fisheries of the Food and Agriculture Organization of the United Nations. Born in Canada, he took his bachelor's and master's degrees in zoology at the University of British Columbia in 1938 and 1940 respectively; his Ph.D. (in fisheries) was obtained from the University of Washington in 1951. Idyll was a biologist with the International Pacific Salmon Fisheries Commission from 1941 to 1948 and then taught at the University of Miami until 1972, when he took up his present work. He is the author of *Abyss: The Deep Sea and the Creatures That Live in It* and *The Sea against Hunger* and editor and coauthor of *Exploring the Ocean World: A History of Oceanography.*

VERA C. RUBIN ("The Dynamics of the Andromeda Nebula") is a member of the staff of the Department of Terrestrial Magnetism at the Carnegie Institution of Washington. Her degrees are in astronomy: a bachelor's degree from Vassar College in 1948, a master's degree from Cornell University in 1951 and a Ph.D. from Georgetown University in 1954. "My principal interest," she writes, "has always been optical spectroscopy to determine the structure and motions in galaxies. My outside interests are those one acquires in raising four children (a geologist, an astronomer, a mathematician and a violinist) and running a busy, exciting, maidless household. Five or six times a year I disappear from Washington for a week or two to observe at Lowell Observatory, Kitt Peak Observatory or the Cerro Tololo InterAmerican Observatory in Chile. I integrated the Palomar Observatory by being (in 1965) the first woman allowed to observe there. I have found time twice in the past few years to teach an astronomy course at a local high school. All of this is made possible and fun by the support and encouragement of Robert J. Rubin, a physicist at the National Bureau of Standards."

R. R. ALFANO and S. L. SHAPIRO ("Ultrafast Phenomena in Liquids and Solids") are respectively at the City College of the City University of New York and the Los Alamos Scientific Laboratory of the University of California; Alfano is professor of physics and Shapiro is a member of the Los Alamos laboratory's laser division. They were together at the General Telephone and Electronics Research Laboratories from 1969 to 1972. Alfano obtained his bachelor's and master's degrees at Fairleigh Dickinson University and his Ph.D. from New York University. He enjoys painting landscapes and gourmet cooking. Shapiro received his bachelor's degree at the University of California at Berkeley and his Ph.D. from the University of California at Santa Barbara.

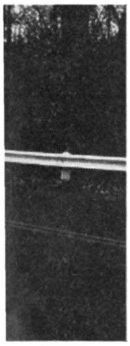
ALAN SOBEL ("Electronic Numbers") is a member of the research department of the Zenith Radio Corporation. After receiving his bachelor's and master's degrees in electrical engineering at Columbia University in 1947 and 1949 respectively, he worked in New York as an electronics engineer, usually taking one or two physics courses a year toward his doctorate. "I finally decided," he writes, "that the part-time route to the Ph.D. was too arduous, and so with the assistance of a working wife (who had received her master's degree in chemistry) I went back to school full time and eventually ground out a Ph.D. in physics from the Polytechnic Institute of Brooklyn in 1964."

WILLIAM C. DENISON ("Life in Tall Trees") is associate professor of botany at Oregon State University. He writes: "I grew up in Rochester, N.Y., where my father taught school. At his suggestion I took shop courses in lieu of language in high school. I also worked as an apprentice millwright for the Eastman Kodak Company. After three years at Oberlin College I was thrown out for marrying a girl whom I now view, from the perspective of 25 years of marriage, as the best of many fine things I found



The one-minute long distance rate. An inexpensive way to say you got home okay.

When you've something short to say, like "I'm home," we've got a way for you to say it for very little money.



It's our late-night and early-morning one-minute rate for long distance calls you dial direct without operator assistance.

You can use it any time after 11 at night all the way till 8 the next morning.

*Dial-direct rates apply on all interstate calls (excluding Alaska and Hawaii) completed from a residence or business phone without operator assistance. They also apply on calls placed with an operator from a residence or business phone when direct dialing facilities are not available. They do not apply to person-to-person, coin, hotel-guest, credit-card or collect calls, or to calls charged to another number, because an operator must assist on such calls.

Any night of the week.

And it won't cost very much. Even coast-to-coast. New York to Los Angeles, for instance, is only 35¢, plus tax (additional minutes are 20¢ each).*

Obviously, most of your calls will take longer, so they'll cost more. But you'll still save money when you dial direct.

At AT&T and your local Bell Company, we want you to know all the ways to use your phone so you can save money. That includes using our late-night one-minute rate and dialing all long distance calls direct.

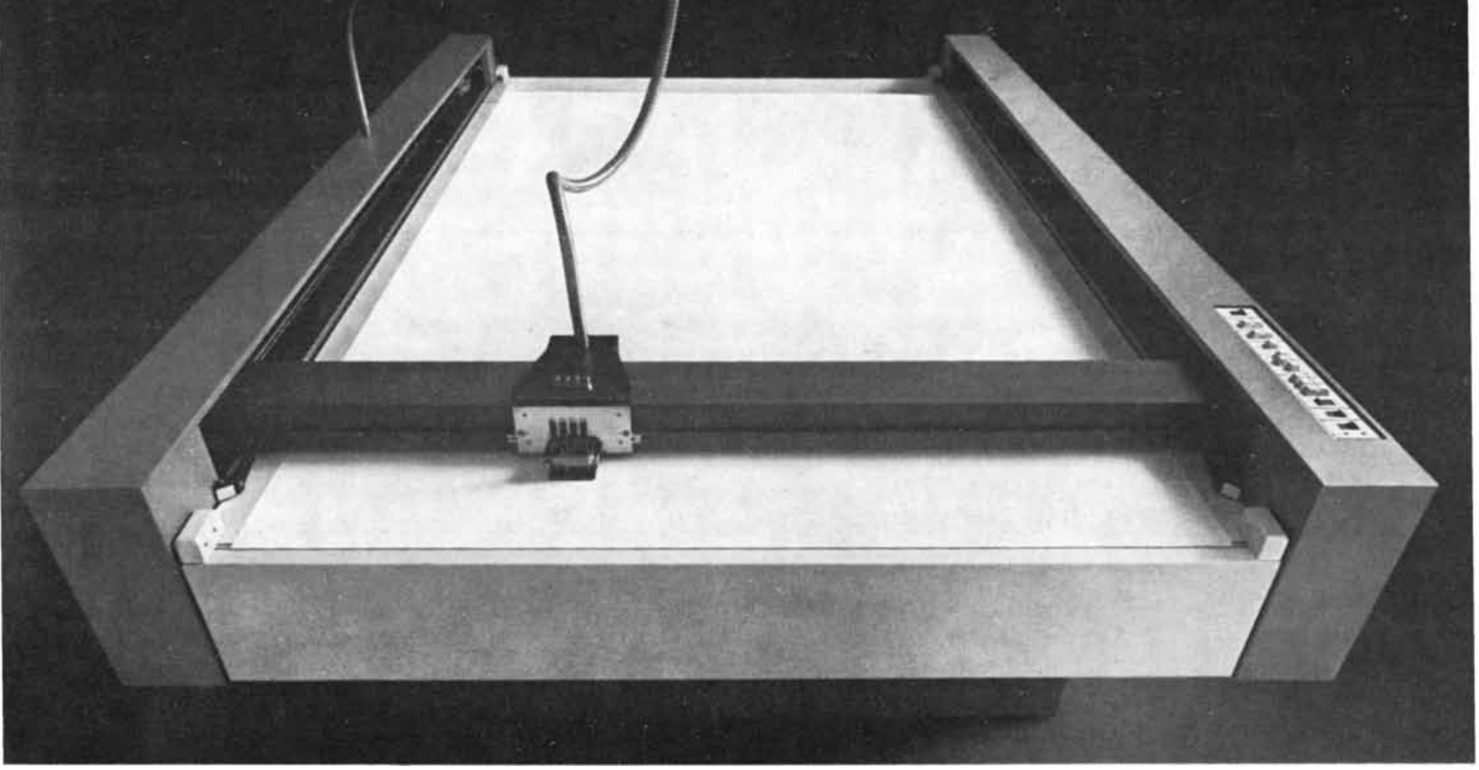
Because we hear you.

We hear you.

at Oberlin. Eventually I returned and received both bachelor's and master's degrees in botany. I obtained a Ph.D. in mycology from Cornell University and spent 11 years teaching at Swarthmore College." Denison wishes to acknowledge the support of the Western Coniferous Forest Biome of the International Biology Program.

RICHARD A. LERNER and FRANK J. DIXON ("The Human Lymphocyte as an Experimental Animal") are at the Scripps Clinic and Research Foundation: Lerner is associate member of the department of experimental pathology and Dixon is chairman of that department and of the biomedical research departments. Lerner obtained his bachelor's degree and his M.D. at Stanford University; Dixon's are from the University of Minnesota. Lerner has been at Scripps since 1965 except for two years at the Wistar Institute. Dixon was at the Harvard Medical School, the Washington University Medical School and the School of Medicine of the University of Pittsburgh before going to Scripps in 1961. Lerner writes that he and Dixon "are addicted to jogging along the beach in La Jolla; we have been running together almost daily for six years, and many of our scientific discussions take place as we run."

ALBERT L. ZOBRIST and FREDERIC R. CARLSON, JR. ("An Advice-taking Chess Computer"), are at the University of Southern California; Zobrist is assistant professor of electrical engineering and computer science and a member of the research staff of the university's Information Sciences Institute and Carlson is assistant professor of electrical engineering and computer science and assistant professor of education. Zobrist obtained his bachelor's degree (in mathematics) at the Massachusetts Institute of Technology in 1964 and his advanced degrees at the University of Wisconsin: his master's degree in mathematics in 1966 and his Ph.D. in computer science in 1970. "Computer chess is supposed to be a side interest," he writes. "My main areas of study are pattern recognition, robot vision and computer analysis of natural language. I enjoy living in the West and head for the mountains or the desert as often as I can." Carlson received his bachelor's degree at the University of Rochester and his Ph.D. from the University of Southern California. "My main interest is in applications of computers in education," he writes. "I enjoy recreational flying and have my pilot's license."



The four-fingered fast-drawing kid is coming to town.

We call it the 748. It's big. And it's fast.

Its 4 inking pens move at speeds over 40 inches a second. That's faster than ink flows, so we had to figure out a pressure inking system that lets the ink catch up to the pens.

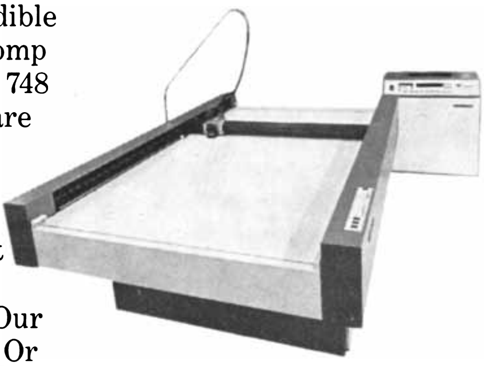
If you make integrated circuits or maps. Or if you have precision drafting needs, this is the new tool.

The 748's plotting area is 48″x 82″. It can scribe coated materials and cut strippable film. At peak speeds!

The quality of its line is not impaired by its incredible speed. And like all CalComp flatbed plotters, the new 748 offers easy-to-use software for most computers.

We service it in 29 countries. But the way we've made it, you won't be calling us very often.

Do call us to see it. Our number is (714) 821-2011. Or write California Computer Products, Inc., SA-M6-73, 2411 West La Palma Avenue, Anaheim, California 92801.



CALCOMP

Established 1845 **SCIENTIFIC AMERICAN** June 1973 Volume 228 Number 6

Enforcing the Clean Air Act of 1970

The act calls for the substantial reduction of pollutants by 1975. Its provisions are now the subject of many adversary proceedings. In the long run a new kind of cost-benefit analysis may be needed

by Noel de Nevers

Some two and a half years have passed since Congress enacted and the President signed the Clean Air Act of 1970, setting in motion a vigorous attack on air pollution in the U.S. The act has figured prominently in the news in recent months, mostly because it requires that emissions from new automobiles be reduced by 90 percent during the period from 1971 to 1976. The law also deals with emissions by new industrial installations such as electric power plants, oil refineries and copper smelters and with a special category of "hazardous pollutants."

The Clean Air Act is beginning to take effect at precisely the time when its pressures to improve air quality bring it into conflict with efforts to deal with shortages of energy. For example, to meet air-pollution standards many plants are changing from high-sulfur coal, of which the U.S. has a plentiful supply, to low-sulfur oil, which is in short supply and is chiefly imported. In this situation the Clean Air Act was bound to evoke controversy. Indeed, the history of activity on air pollution since the act was passed is largely the story of efforts by some groups to implement it and by other groups to prevent its implementation from causing them hardship. I have followed the debate closely, both as a member of the chemical engineering faculty of the University of Utah and (during the academic year 1971–1972) as a staff engineer of the U.S. Environmental Protection Agency, which is charged with putting the Clean Air Act into effect. In what follows I shall describe the major provisions of the act and summarize how matters stand in implementing them.

Air-pollution laws of some kind have existed in many areas for a long time. Until the late 1940's, when Los Angeles began to break new ground in an effort to deal with the photochemical smog resulting principally from automobile emissions, the laws were mostly smoke ordinances, requiring large users of fuel to minimize their emissions of smoke. Since visible smoke normally signifies wasted fuel, such ordinances could be enforced without great difficulty once the emitter recognized that a smoking stack was wasting money.

The Federal Government's involvement in air pollution began in 1955 with the enactment of a law that authorized the U.S. Public Health Service to conduct research on air pollution and to give technical assistance to state and local governments. Later laws extended the Federal role, but none was as comprehensive or as strict as the 1970 act, which reflected the growing public concern over the environment and also a recognition by Congress that most of the Federal, state and local agencies concerned with air pollution had been unable to accomplish much. Congress decided that the only way to press the attack on air pollution effectively was by means of a strong law that established deadlines for accomplishing results and authorized large fines for violators.

The basic philosophy of the act is most easily discussed in terms of the concept of the threshold value. In the act Congress assumed that for most of the major air pollutants (sulfur dioxide, carbon monoxide, nitrogen oxides, particulates and oxidants) there is some concentration below which the pollutant has no measurable effect. That concentration is termed the threshold value. Congress based the principal set of regulations on the threshold concept. It also recognized, however, that there are or may be other pollutants that are harmful in any amount and thus have no threshold. They are the ones covered in the separate set of regulations on hazardous pollutants.

For threshold-value pollutants Congress directed the administrator of the act to determine on the basis of medical evidence what the threshold value is for each pollutant and then to use that value, minus "an adequate margin of safety," as a basis for setting primary and secondary standards, which are called national ambient-air quality standards. The primary standards are designed to make sure that the pollutants cause no damage to human health; the secondary standards relate to human welfare and are supposed to ensure that the pollutants do not damage property, vegetation or animals. Once the standards were set the states were directed to prepare state implementation plans showing how they would make certain that the standards were never exceeded anywhere in any state after June, 1975.

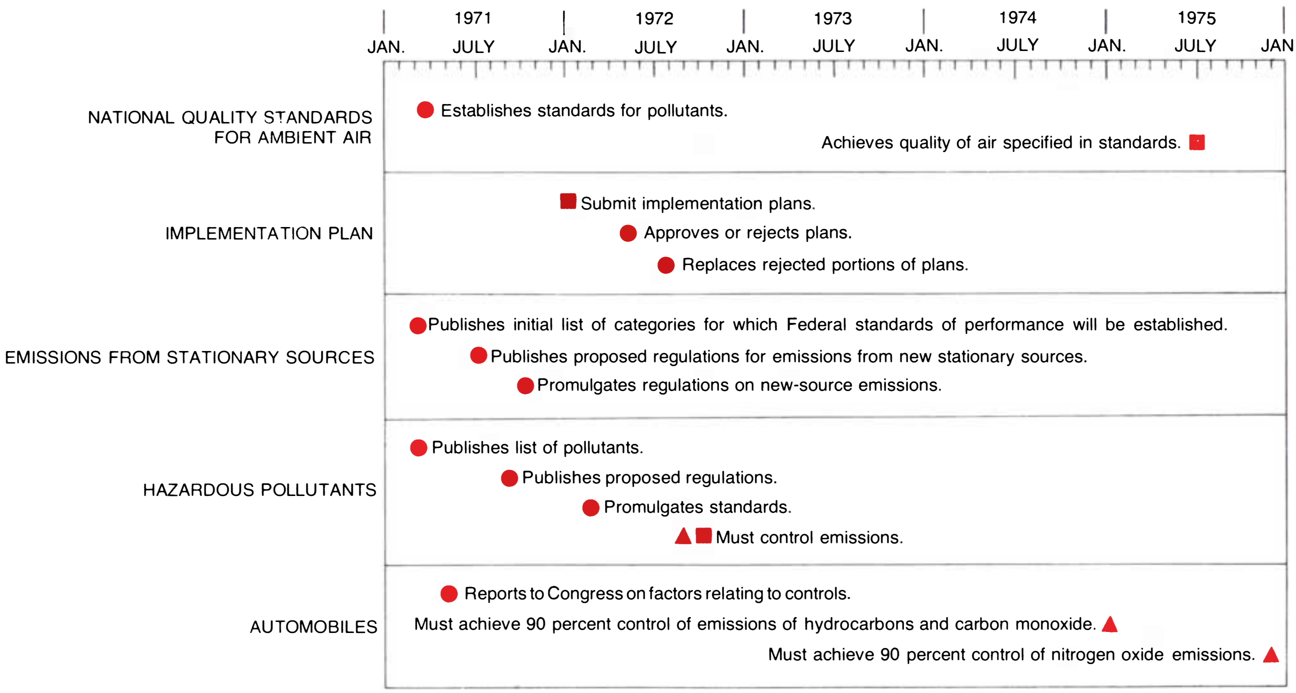
State plans were to be submitted to the Environmental Protection Agency. The agency can approve a plan or replace all or any part of one with regulations that have the force of law. The act called for the entire planning process to be completed by the beginning of 1973, a goal that has been substantially achieved.

Hazardous pollutants are defined as pollutants that have no threshold value and are deemed likely to cause "an increase in mortality... or... irreversible, or incapacitating reversible, illness." The Environmental Protection Agency was directed to prepare a list of such pollutants and then to issue regulations limiting their emission by both new and existing sources. The pollutants listed so far include beryllium, asbestos and mercury. The regulations affecting them go into effect this year. A state can participate in the enforcement of the regulations affecting these pollutants if it wants to, but the primary responsibility is lodged with the Federal Government.

In providing for controls on the emission of air pollutants from new industrial plants Congress attempted to prevent promoters and local governments in sparsely populated areas from attracting industry with promises that no undue fuss would be made over air pollution. A case in point is the Four Corners power plant, which was built in a remote section of northwestern New Mexico. The plant originally had only rudimentary devices for controlling air pollution and so was cheaper to build and run than plants built to the higher pollution-control standards routinely observed in most cities. When plants are built as the Four Corners plant was, the effect is to transfer air pollution from heavily populated areas to sparsely populated ones. The Clean Air Act of 1970 therefore directed the administrator to set "new source performance standards," which limit the emission of pollutants by all new plants. In this way the opportunity to hold out the prospect that a plant can be built with little or no outlay for control of air pollution is eliminated.

Although the act leaves the states free to impose whatever regulations they deem necessary on industries, commercial establishments and homes, Congress preempted the regulation of motor vehicles and aircraft for the Federal Government on the ground that it would be impractical for the manufacturers of such mobile vehicles to face different air-pollution regulations in different states and cities. Only in the case of the automobile regulations did Congress itself set standards instead of delegating the responsibility to the Environmental Protection Agency. Concerned that the industry would put political pressure on the agency for weak standards, Congress wrote into the act requirements that automobiles of the 1975 model year must be built to emit 90 percent less carbon monoxide and hydrocarbons than the 1970 models and that the 1976 automobiles must achieve the same decrease from the 1971 models in the emission of nitrogen oxides. (In April the then administrator of the agency, William D. Ruckelshaus, granted the automobile industry's request for a one-year extension of the deadline for meeting the standards for hydrocarbons and carbon monoxide.) Since by 1971 the manufacturers were already forced by existing law to make automobiles with significantly lower emission rates than those of the uncontrolled cars of the late 1950's and the early 1960's, the requirements of the act for further reductions of 90 percent by the middle of the 1970's present the industry with a formidable technological challenge.

Although Congress gave the Environmental Protection Agency the responsibility for implementing the Clean Air



MAJOR DEADLINES established by Congress in the Clean Air Act of 1970 are summarized. The circle symbolizes actions required of the U.S. Environmental Protection Agency, which has responsibility for putting the act into effect. The square symbolizes action that must be taken by the states, and the triangle shows deadlines for industry. In April the Environmental Protection Agency granted the request of the automobile industry for a year's delay in meeting the standards for hydrocarbons and carbon monoxide.

Act and the power to do many things, it did not grant unlimited power or provide the agency with a clear field in which to operate. As in the case of many Federal activities, several agencies have an interest in what is done about air pollution. They reflect the views of different interests in the population and tend to defend those views within the Government. The final action in a field such as air pollution usually represents a compromise of the positions assumed by the various adversary branches of the Government. Three adversaries that the Environmental Protection Agency has faced in trying to implement the Clean Air Act are the Department of Commerce, the Bureau of Mines of the Department of the Interior and the President's Office of Science and Technology (which is being disbanded).

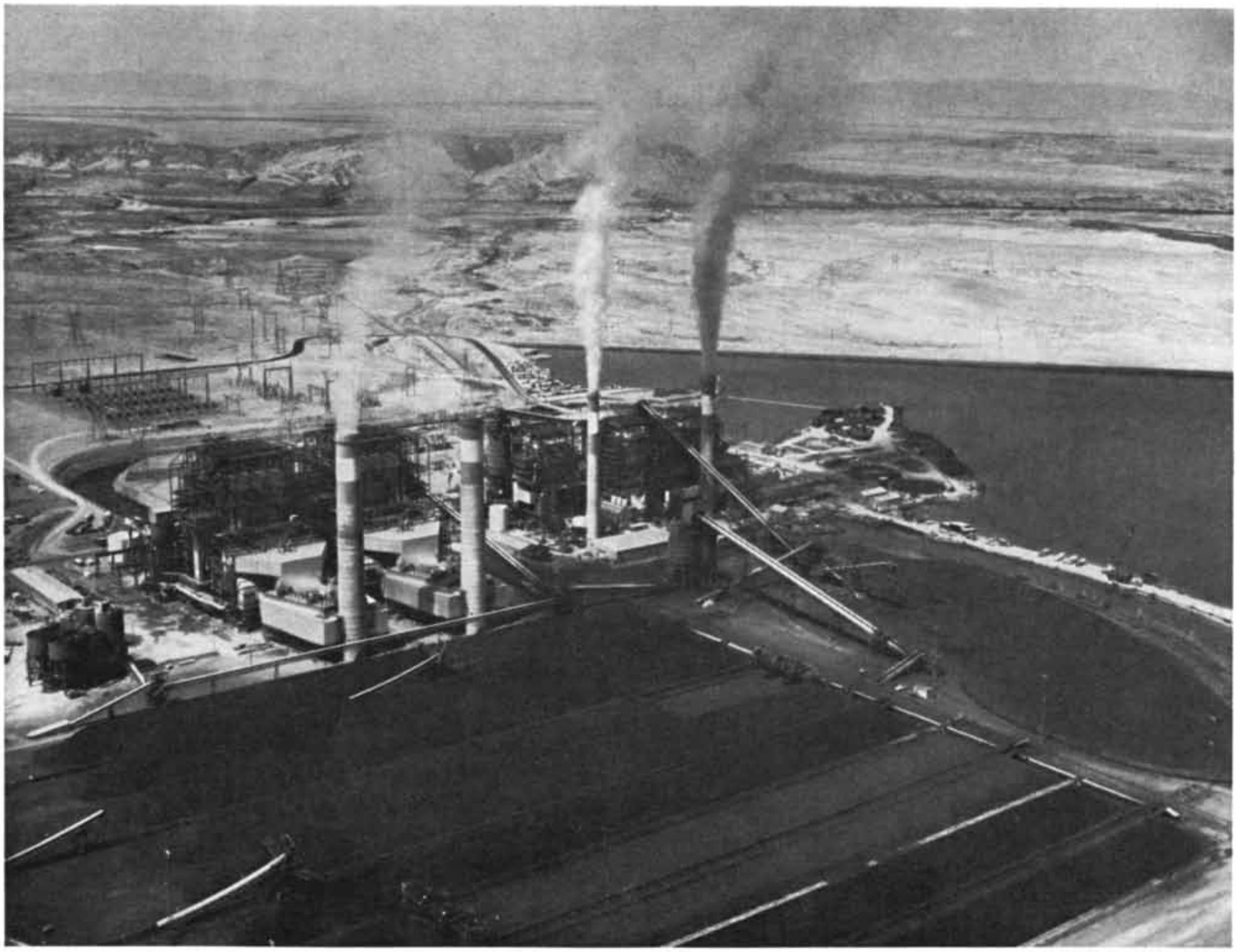
The Department of Commerce represents the views of industry within the Government. Industry, with few exceptions, is opposed to any vigorous pollution-control program because of the costs involved. The department has therefore aided industry in an effort to prevent effective implementation of the Clean Air Act. Former Secretary of Commerce Maurice H. Stans attacked the act in speeches and articles. The department also issues under its imprimatur the publications of the industry-controlled National Industrial Pollution Control Council, which seeks to explain and defend industry's position.

The Bureau of Mines represents within the Government the views of the mining and fuel industries. It has regularly commissioned and published reports attacking proposed actions by the Environmental Protection Agency as being likely to harm the industry. The Office of Science and Technology issued a detailed report calling on Congress to repeal the provisions of the act that were most objectionable to the automobile industry.

The role of judge and jury in the intragovernment adversary system is held by the Office of Management and Budget. It reviews the proposed actions of the Environmental Protection Agency and clears them with other agencies. When there are conflicts, the budget office resolves them.

The Clean Air Act required the states to submit to the Environmental Protection Agency by the beginning of 1972 their implementation plans for meeting the national air-quality standards. The planning process gave rise to several controversies, of which the most colorful is a contest involving the Environmental Protection Agency, the copper-smelting industry and the eight Western states where the nation's copper smelters are located.

In its guidelines to the states for preparing their implementation plans the



FOUR CORNERS POWER PLANT in northwestern New Mexico had few pollution-control devices when it first went into operation. In this photograph, which was made in August, 1971, the stacks are emitting conspicuous plumes of particulates into what had previously been unpolluted air. Scrubbers for removing particulate matter were later installed on the two smaller stacks at right.

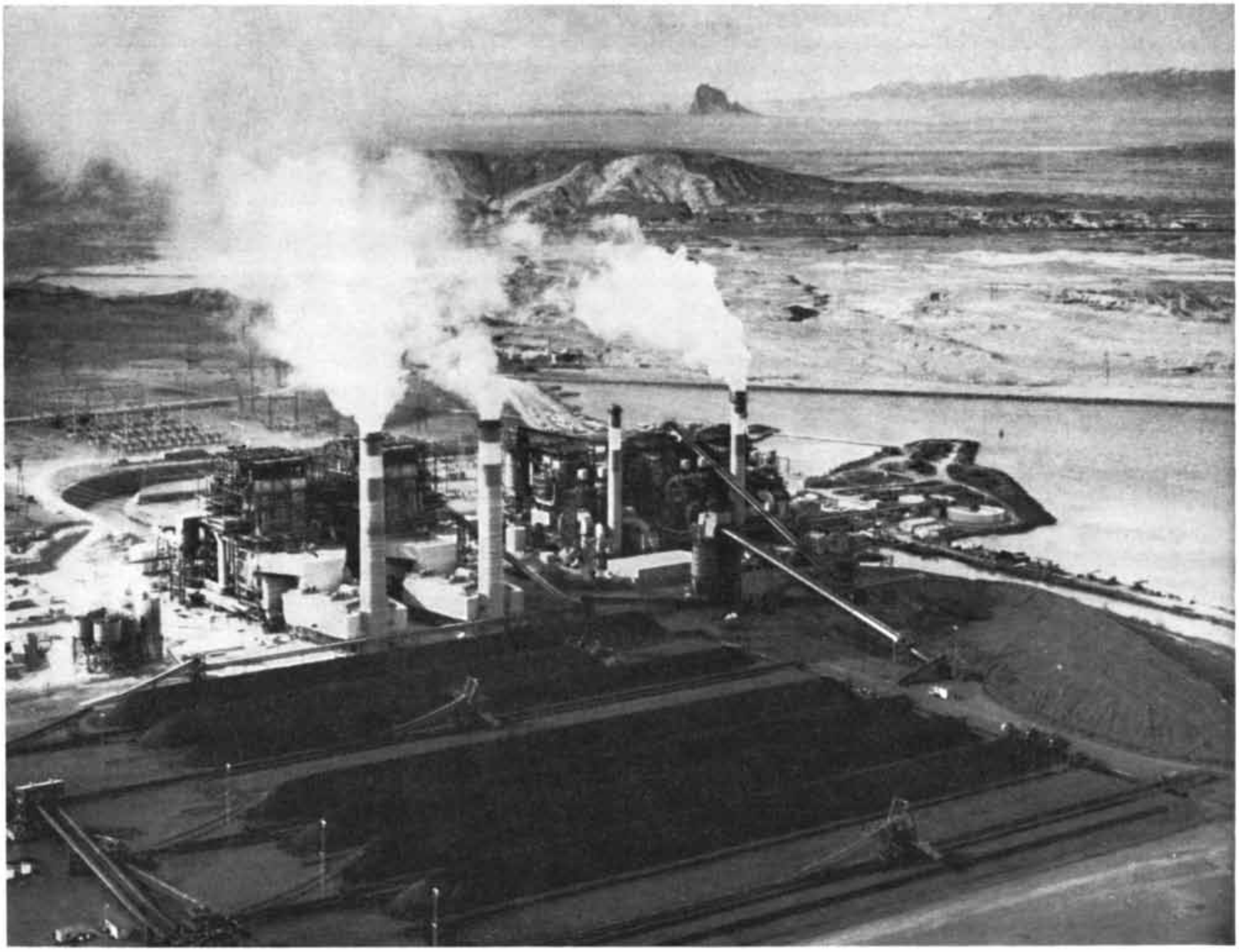
Environmental Protection Agency provided examples of reasonable degrees of control to be expected of various industries. For copper smelters it suggested that recovery of 90 percent of the sulfur dioxide emitted in the smelting process was a reasonable objective. Several states adopted the standard in the first draft of their plans.

Then the smelter operators began to protest. In each state they argued that the cost of the control measures was exorbitant and warned that enforcement would put the smelters out of business in that state, forcing them to move to some other state with more lenient regulations. The argument was effective. Arizona, which is the largest producer of copper, included in its implementation plan only one page on controlling sulfur dioxide. The page said that the plan for sulfur dioxide would be submitted later.

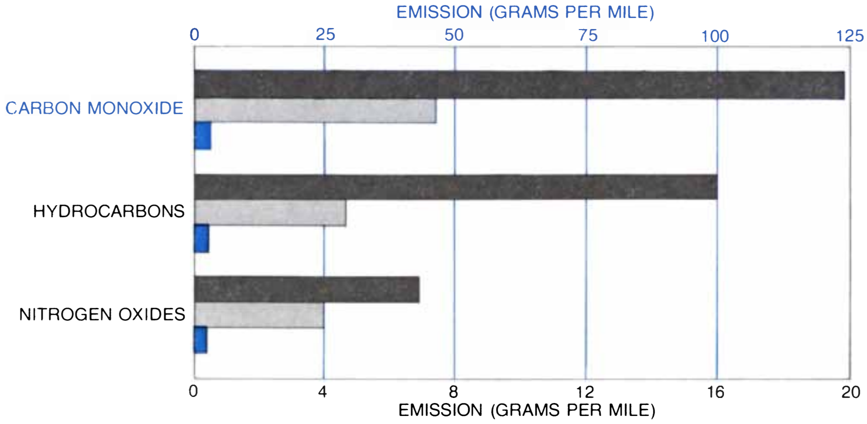
The governor of Montana sent to the Environmental Protection Agency a plan from which he had deleted the section that the state's air pollution control board had proposed on controlling sulfur dioxide from copper smelters. The board thereupon sent the original plan to the Federal agency. The agency accepted the plan with the governor's signature as the state's official submission. The state board sought by court action to have its plan accepted as official. In the end the Environmental Protection Agency decided to write the regulations for all copper smelters. The industry is now challenging the regulations in court.

The group within the Environmental Protection Agency that is responsible for issuing the new-source performance standards has also encountered legal challenges. The agency's mandate from Congress was to choose categories of industry and prescribe allowable emission rates for them. In the first group the agency chose fuel-burning electric power plants, municipal incinerators and plants that make sulfuric acid, nitric acid or cement. When the regulations were issued, the agency was sued by associations representing three of the five industries.

The challenge by the electric power industry hinges on the words in the Clean Air Act requiring the agency to base the standards on the "best system of emission reduction" that has been "adequately demonstrated." The issue in litigation is what constitutes adequate demonstration, in this case of controlling the emission of sulfur dioxide from power plants that burn coal or oil. The Environmental Protection Agency points out that several plants built in England some 35 years ago have exhaust systems containing apparatus for controlling sulfur dioxide and that various suppliers in the U.S. will install and guarantee such equipment today. The power industry contends that the British plants and



EFFECT OF SCRUBBERS installed at the Four Corners plant is evident in the emission from the stack at the right. The emission is steam, which quickly dissipates. The other small stack was not operating when the photograph was made in January, 1972. The two large stacks at the left, which had not been equipped with scrubbers, were still emitting large amounts of particulate matter.



EMISSIONS BY AUTOMOBILES will be greatly reduced under the Clean Air Act. The chart shows emissions of the three major automotive pollutants by a typical car in 1960 (*black*), before any controls were required, in 1970 (*gray*), when cars were subject to certain controls, and in 1976 (*color*) under the standards of the Clean Air Act. The values are for specified driving cycles, which are intended to approximate typical urban usage, but since the cycles have been changed several times the values are not directly comparable.

plants built in this country with sulfur dioxide controls have been unreliable, expensive and troublesome to operate, so that the technology has not been adequately demonstrated. In the end the issue will be resolved in the courts.

A second lawsuit against new-source standards has been filed by the cement industry. In trying to establish what had been "adequately demonstrated" for controlling the emission of fine particles from cement plants the Environmental Protection Agency tested several plants and found some that were doing a better job of removing fine particles from their exhaust gas than any electric power plant was doing. On the basis of the "best system of control" to have been adequately demonstrated the agency issued a more stringent standard for cement plants than for power plants. The cement industry is challenging the standard in court. Apart from the legal issue the case poses a moral question: When one industry has done a better job than another in a type of pollution control, should the less diligent industry be allowed to persist in its ways or should both industries be required to meet a common standard?

While under attack from industry on the ground that some of the standards are too stringent, the Environmental Protection Agency has been sued by groups of conservation organizations contending that some of the standards are not as stringent as the law requires. In one suit the conservationists pointed out that the stated purpose of the Clean Air Act is to "protect and enhance the quality of the Nation's air" and that the agency in issuing its standards had specified that the standards "shall not be considered in any manner to allow significant deterioration of existing air quality in any portion of any state." The conservationists argued that the agency's state planning guidelines, which allowed increases in air pollution up to the secondary standards, did not protect and enhance air quality and did allow significant deterioration. The conservationists succeeded in their suit in a Federal district court, which enjoined the Environmental Protection Agency from approving any state plan that failed to protect areas with air cleaner than the standards from having their air degraded to the level of the secondary standards. The decision was upheld by a circuit court of appeals and is now under review by the U.S. Supreme Court.

As the decision now stands, it presents a serious dilemma for the Environmental Protection Agency. If the decision is interpreted in the strictest way, it would mean that areas now lacking in industry and population can never have them, because there is no known way of converting open land to industrial or residential use without some increase in the rate of emission of air pollutants. It seems unlikely that the law will be thus interpreted, but the issue still awaits a judicial determination of what is meant by "significant deterioration."

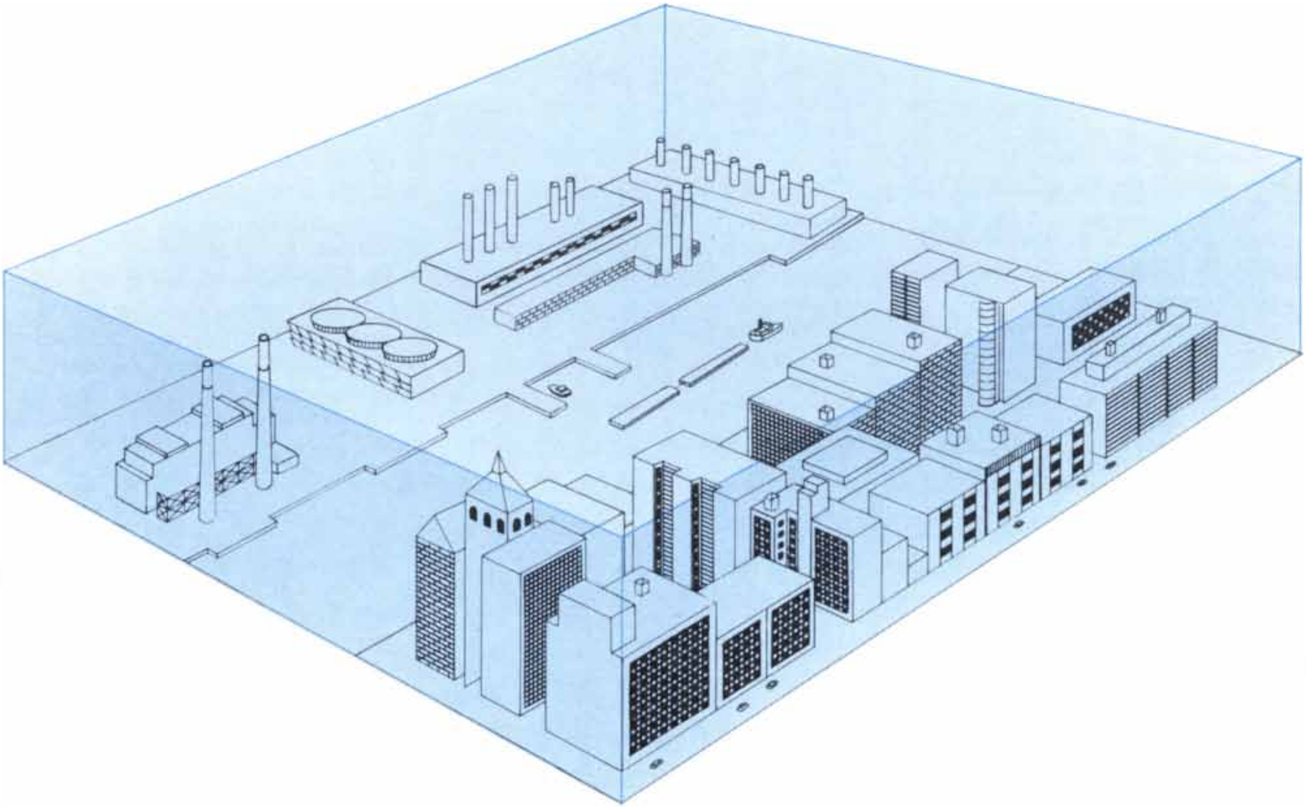
In other lawsuits conservation groups have sought to hold the Environmental Protection Agency to the letter of the law and to force it to meet the deadlines written into the act. Conservation groups succeeded in having the courts overturn most of the delays that the Environmental Protection Agency granted under a provision of the act allowing the agency to extend a state's deadline for meeting air-quality standards if the state shows adequate reasons for a delay. The courts held that the agency had granted the delays without thorough review of the reasons.

Another controversy involving enforcement of the Clean Air Act concerned the use of mathematical models of air pollution. Plainly one cannot duplicate in a laboratory the conditions of air pollution in a state or a large metropolitan area; the task would require reproducing on a reduced scale the emission pattern, the climate and the mixing of pollutants in the entire atmosphere for the entire year. Yet the Environmental Protection Agency had to have some basis for meeting its statutory mandate to decide whether a state's laws would clean up the air to the extent required by the Clean Air Act or whether, as the act specified, the agency should promulgate regulations overriding any inadequate state provisions. The Environmental Protection Agency informed the states that they could employ either of two mathematical models to compute the concentration of pollutants to be expected after a state's regulations went into effect.

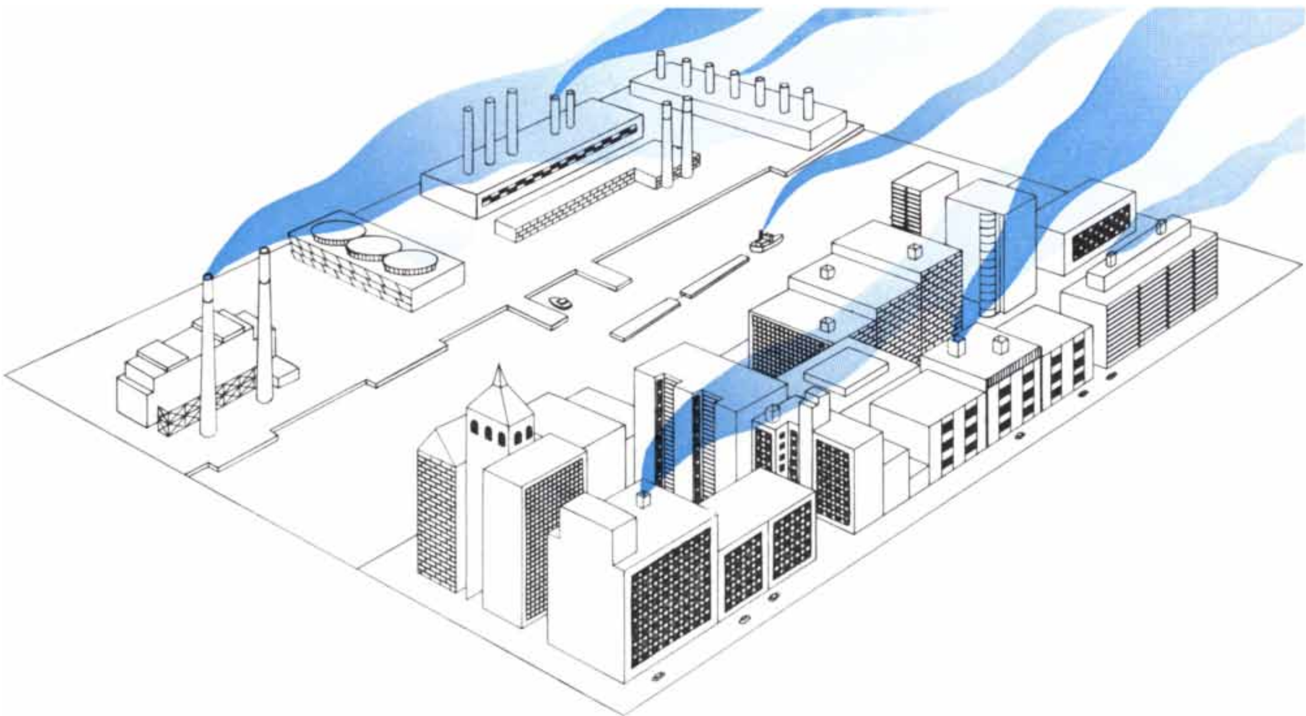
One model is relatively simple. It assumes that the concentration of pollutants at any point is a linear function of the total rate at which pollutants are emitted in the region. On this basis one can readily calculate the required reduction in the emission rate. The problem is that the model is quite unlikely to give accurate predictions of the effects of reducing emissions from certain sources, such as power plants, without corresponding reductions in all others, such as household heating systems.

The second model takes into account the variabilities of the atmosphere and the locations and emission rates of all the sources of air pollution in the region. This model is quite complicated, requiring large amounts of expensive computer time. Moreover, it requires much more information than is usually at hand, including a detailed census of emissions. It can, however, be expected to yield more accurate predictions than the simple model, although neither model has been thoroughly tested by comparing its predictions with full-scale results in real locations.

Most of the state plans sent to the Environmental Protection Agency for approval employed the simple model and were easily checked. The agency was aware of the limitations of the model



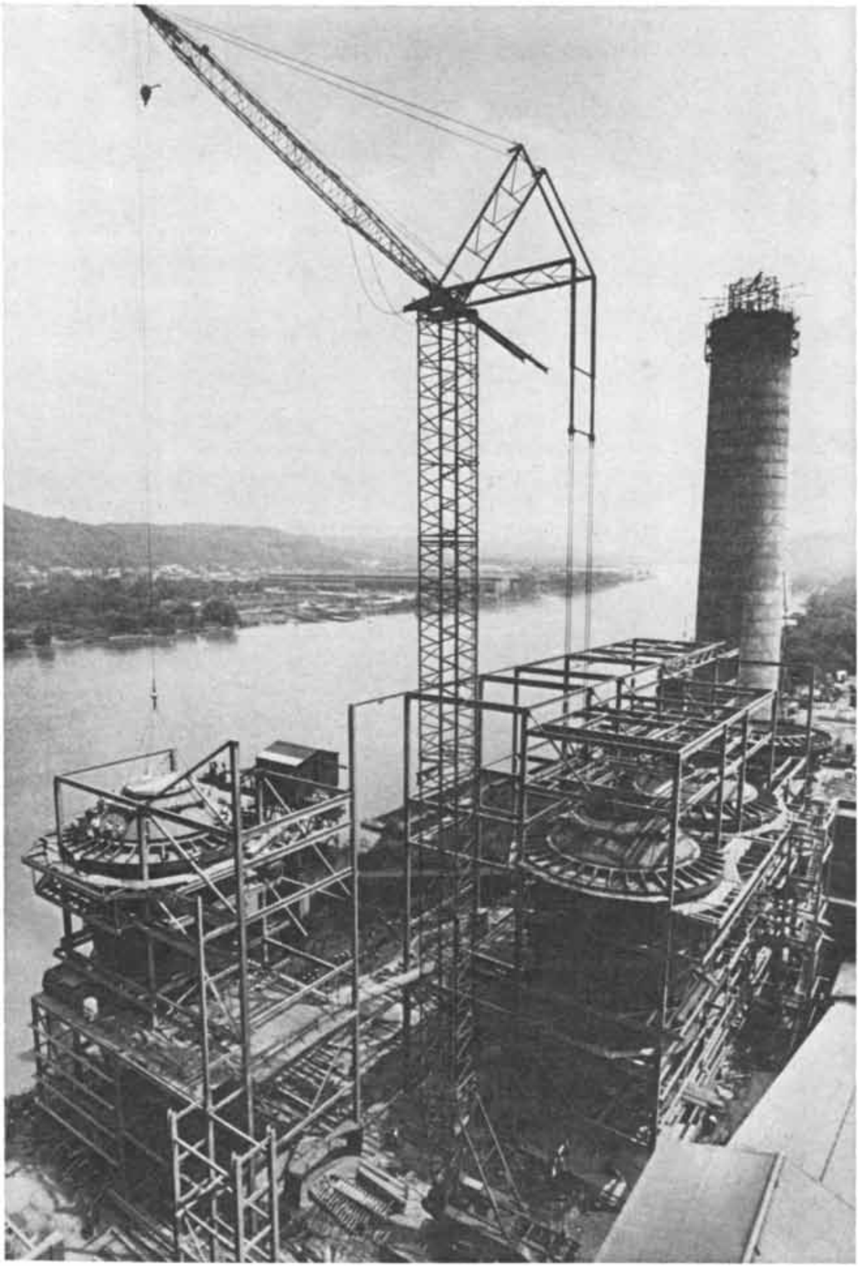
SIMPLE MATHEMATICAL MODEL, which is one of two employed by the Environmental Protection Agency to evaluate state plans for implementing the Clean Air Act, assumes that all the air over a city is well mixed, as represented by the rectangular block of color, so that the concentration of pollutants is linearly dependent on the total rate of emissions from all sources in the city. Although the model has limitations of which the Environmental Protection Agency was aware, the agency accepted it for most plans.



COMPLEX MATHEMATICAL MODEL is more satisfactory for cities with severe pollution problems. It seeks to follow pollutants from each major source and from groups of minor sources by taking into account the dispersion of the plumes, which is a function of wind direction, wind speed and atmospheric stability. With this model one usually considers the frequency of each kind of weather, computes the concentration of pollutants at each source for a given kind of weather and multiplies the concentration by the frequency, summing overall weather conditions to find the annual average concentration. Model has been employed for large cities.

but accepted the plans based on it. For cities with really difficult air-pollution problems, however, the simple model would have required emission reductions that probably would have been impossible to achieve. In such cases the more complex model was employed and the control regulations were based on its predictions. This move caused a controversy within the Environmental Protection Agency because some of the officials were opposed to any kind of modeling and were particularly suspicious of a model as complicated as the one used. The internal controversy was not resolved in any satisfactory way, because the opponents of modeling had no practical alternative to propose.

The models also gave rise to external controversy. When a model predicted concentrations of pollutants that were in excess of standards and the Environmental Protection Agency accordingly issued regulations requiring certain industries to make large expenditures for pollution control, the result was legal action. The basic issue in these cases is whether or not the Environmental Protection Agency can force a company to make large expenditures based on calculations of what their effect will be. Industry's position is that such requirements must be based on something firmer than calculated effects. If the courts uphold this view, Congress will have to provide the Environmental Protection Agency with a new basis for ensuring compliance with the standards mandated by the Clean Air Act.

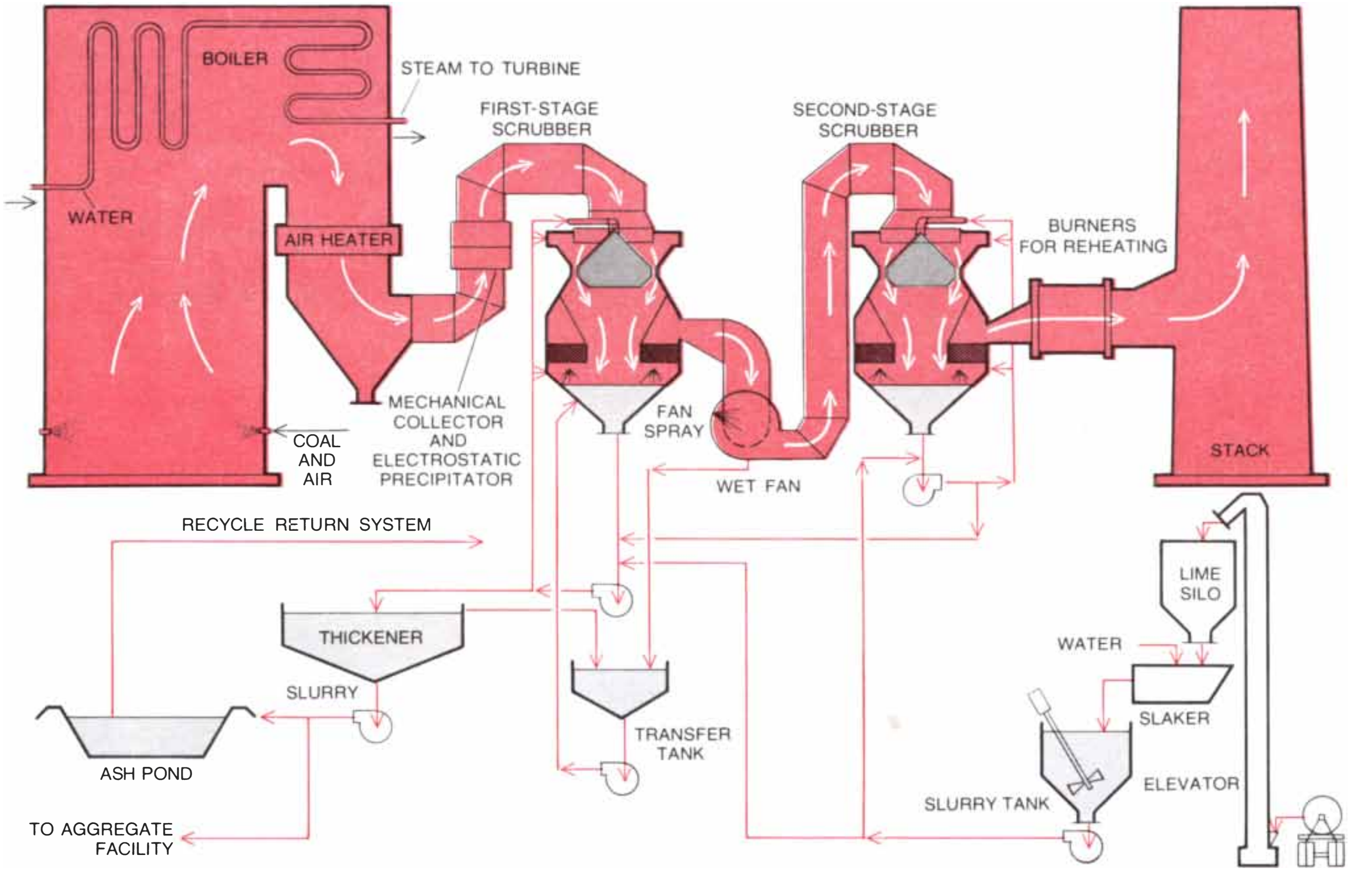


POLLUTION-CONTROL EQUIPMENT is installed at the Phillips Power Station of the Duquesne Light Company near Pittsburgh. The installation was in progress when the photograph was made. The three cylindrical structures at right are fly-ash scrubbers. The cylindrical structure at left is a sulfur dioxide scrubber designed to operate in conjunction with one of the fly-ash removers. The coal-fired plant has six steam generators and a total electrical output of 400 megawatts. The river flowing past the power plant is the Ohio River.

A case illustrating many of the dilemmas of the Clean Air Act is provided by the large power plant now under construction in Huntington Canyon in Utah. The plant will burn low-sulfur coal; if it were located in a flat area, it would meet the standards for sulfur dioxide emission. The fact that it is in a canyon, however, complicates matters. For a site in a canyon the mathematical models I have described are not applicable, and special models are required. The Environmental Protection Agency employs a model which predicts that the plume from the plant will strike the canyon wall, resulting in levels of pollution higher than the standards for ground level. The agency therefore issued regulations requiring the installation of equipment to control sulfur dioxide emissions.

The owner of the plant argues that a number of competent meteorologists believe that the special model for canyon sites is incorrect and that according to more reliable models the plume will never run directly into the wall. The owner also agrees with the electric power industry that sulfur dioxide controls have not been adequately demonstrated. A final argument is that people seldom, if ever, visit the steep canyon walls, so that a large expenditure to meet air-quality standards there is unwarranted. This controversy, like many others involving the Clean Air Act, appears to be headed for the courts.

A special and apparently unique situation still exists in the Los Angeles area, where the atmospheric conditions are highly conducive to photochemical smog and the motor-vehicle population is large. According to the Clean Air Act of 1970 every part of the U.S. must meet the air-quality standards by 1975 or soon afterward, no matter how draconian the necessary control measures may be. In January of this year the Environmental Protection Agency proposed regulations to meet the standard for oxidants in Los



FLY-ASH AND SULFUR DIOXIDE REMOVAL SYSTEM employed in the Phillips Power Station is depicted schematically. It is a two-stage system in which particulate matter and some of the sulfur dioxide are removed in the first stage, that is, in the fly-ash scrubbers, and more sulfur dioxide is removed in a limestone solution in the second stage. The drawing is based on a diagram of the two-stage installation made by the Chemical Construction Corporation, which supplied several of the system's components.

Angeles. The regulations include severe restrictions of motor-vehicle travel, to be achieved by gasoline rationing. The proposed restriction would allow only 18 percent of the present amount of use of motor vehicles. Ruckelshaus has acknowledged that such a reduction would probably be politically impossible, but he was forced by court order to obey the Clean Air Act and propose regulations that would, if they were enforced, enable Los Angeles to meet the standard. The motor-vehicle problem in the Los Angeles area will represent a crucial test of whether or not the air-quality standards can be attained in areas with severe pollution problems.

The catalogue of the troubles the Environmental Protection Agency has had enforcing the Clean Air Act might leave the reader with the impression that the act is an example of poor legislation. In my view it is good legislation. It sets definite goals and establishes deadlines. The fact that difficulties have arisen shows only that it is hard to write a far-reaching law that does not require study and amplification.

In dealing with air pollution everyone is acting on the basis of inadequate information. The relation of dose and response is simply not known for the major pollutants. For this reason it is difficult to resolve the basic argument between the conservationists, who maintain that if the emitter of a pollutant cannot show that the pollutant is harmless, he should be made to halt or control the emission, and the promoters of industrial growth, who maintain that if a planned or existing emission cannot be proved to be harmful, no valid ground exists for preventing it. In this situation Congress sought a reasonable middle ground and decided that the principle of air-quality standards was the best one before it. It probably was the best approach available at the time. As more information emerges about the true effects of air pollution, however, a more refined and cost-effective scheme for regulating pollution is likely to be developed.

Probably in the long run the cost-benefit method of deciding what measures to adopt for the control of air pollution will prove best. A cost-benefit analysis assigns numerical values to such effects of air pollution as impairment of health, damage to buildings and crops and degradation of the quality of life. It will not be easy to develop a cost-benefit approach because hard factual and moral questions must be answered. Facts are needed on the cost of the various kinds of damage done by the various pollutants. Judgments must be made about the relative values of these damages. Moral issues such as whether it is acceptable to clean the air of many by dirtying the air of some must be faced. Decisions of this kind are not part of the standard-setting philosophy of the Federal Government and would be difficult to make with the present level of knowledge about air pollution and the means of controlling it. Until they can be made, the Clean Air Act's doctrine of establishing air-quality standards probably is the best approach to the problem of air pollution.

THE ANCHOVY CRISIS

Over the past decade the world's largest fishery has been in the Peru Current. A periodic ecological disturbance, combined with the heavy fishing, now threatens to destroy the industry

by C. P. Idyll

The world of the Peruvian anchovy is the sweep of a great cold ocean current. In a slow northward drift the current carries the little fishes along in company with countless tiny plants and animals that the anchovies avidly devour and with larger fishes, squids and a host of other marine animals. The anchovies form thronging legions that wheel and dart in the current. Their world is often entered by aliens: birds that plunge from above, snatching up the anchovies by the hundreds of thousands, and men who cast great net enclosures around the fishes, carrying them off to shore by the millions.

In the brief life-span of the anchovy, rarely longer than three years, its cold-current environment usually changes only within narrow limits. During the lifetime of some generations, however, their world may be put out of joint. The slow northward drift of the current, only two-tenths to three-tenths of a knot (compared with the six knots of the Gulf Stream off Florida), becomes still slower and may even reverse itself. The water grows warmer and less salty; the makeup of its populations changes and many of the usually abundant microscopic plants and animals dwindle in number. Finding their world poorer, the little anchovies scatter; many may die prematurely and many in the successor generation may not be born at all. The sea change known as El Niño has arrived.

Coastal Peru is normally a cool and misty land, quite unlike the steamy Tropics that occupy the same latitudes on the eastern coast of South America. The Peru coast is kept that way by the temperature of the ocean current, which in the south can be as cool as 10 degrees Celsius (50 degrees Fahrenheit) and only reaches about 22 degrees C. (71.6 degrees F.) in the north. Although the northernmost part of Peru is a mere three degrees of latitude below the Equator, the average air temperature is a moderate 18 to 22 degrees C. (64.4 to 71.6 degrees F.). Along the 1,475 miles of Peruvian coastline, a distance 100 miles longer than the Pacific coast of the U.S., there are no marshes, mud flats and estuaries—only arid desert, most of it a treeless, monotonous, barren brown. This bleak strip of sand extends a short distance inland, rarely more than 40 miles or so, to the upthrust Andes, one of the most awesome mountain ranges in the world. The high rain shadow of the Andes robs the prevailing southeast winds of their moisture and keeps the coastal region arid. A few streams that rise in the mountains cross the desert; their narrow valleys support what little agriculture exists along the coast.

The Peru Current consists of four components, the interaction of which creates, molds and changes the world of the anchovy. Two of the components travel in a northerly direction: the Coastal Current, flowing next to the shore, and the Oceanic Current, located farther out to sea. Between the two, on and near the surface, runs the Peru Countercurrent. Beneath all three runs the Peru Undercurrent. Both the Countercurrent and the Undercurrent flow southward.

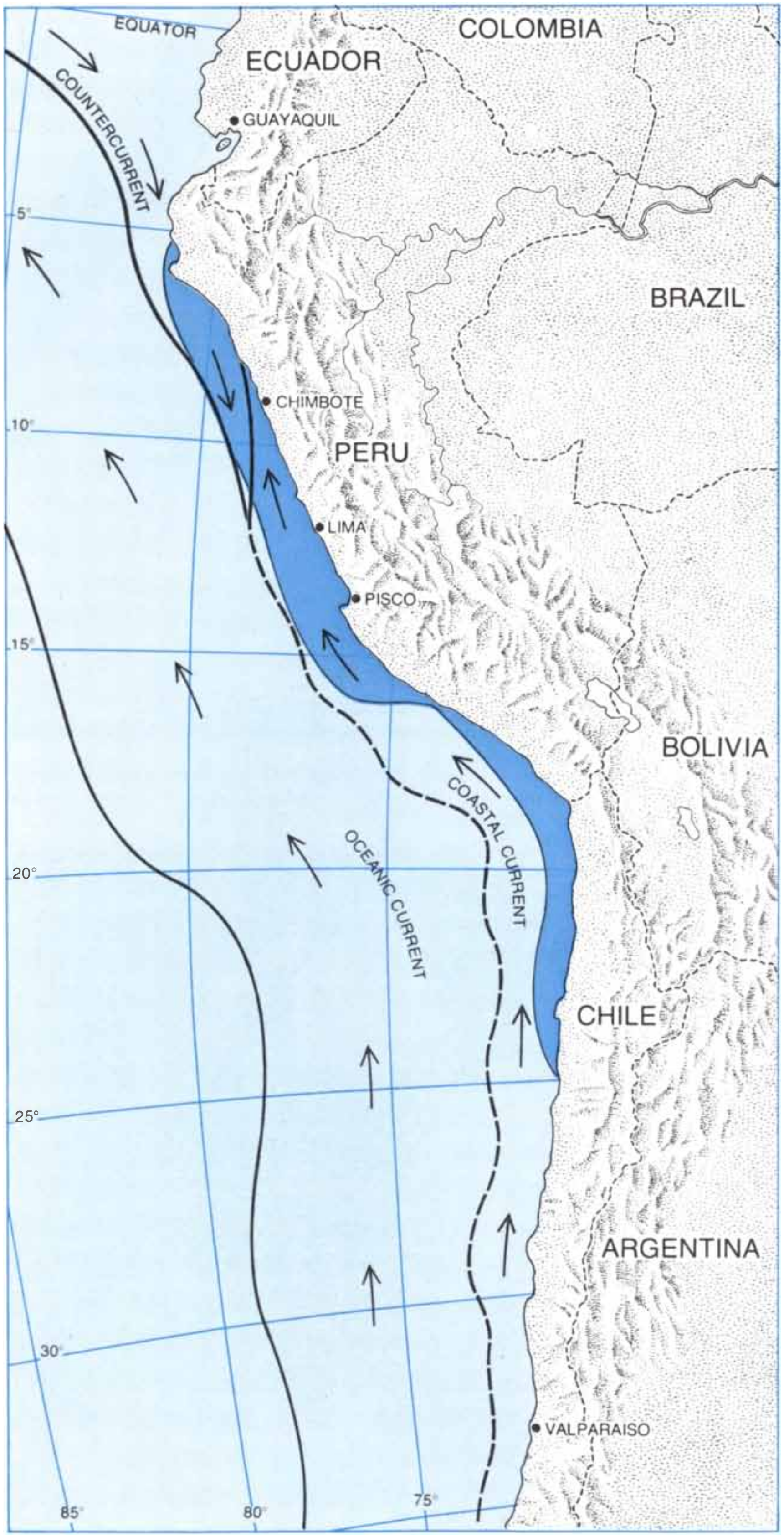
The Coastal Current runs deep and hugs the land from about Valparaiso in Chile in the south to north of Chimbote in Peru, a stretch of some 2,000 miles. The anchovies live mostly in the northern part of this great band of water, which constantly changes shape and size, becoming wider or narrower, deeper or shallower, altering and twisting like an elongated amoeba. The Oceanic Current is longer than the Coastal Current. It is often several hundred miles wide, and it runs as deep as 700 meters. It flows north to a point about opposite the Gulf of Guayaquil before bending west.

The great northward sweep of water is often called the Humboldt Current, after the German naturalist who described the phenomenon following a visit to South America in 1803. Humboldt thought that the cold of the water was the chill of the Antarctic. His conjecture was partly right: the current does include subantarctic water. Much of the cold, however, is the cold of subsurface water. As the water on the ocean surface is swept away by the prevailing winds, deeper low-temperature water wells up slowly to replace it. The trade winds in this part of the world, channeled and bent by the Andes, blow from the south and southeast, mostly parallel to the shore. This prevailing wind urges the surface water northward at the same time that another influence, the Coriolis force, deflects it to the west. As the resulting steady offshore drift skims off the surface layer the cold subsurface water rises with stately slowness to replace it, traveling vertically at a rate ranging from 20 to 100 meters per month, depending on the location and the season.

The biological effect of the upwelling is enormous. That stretch of water, only a tiny fraction of the ocean surface, produces fully 22 percent of all the fish caught throughout the world. Its richness springs from a constantly renewed supply of the chemical nutrients—principally phosphates and nitrates—that stimulate plant growth. Accumulated gradually in the deep layers of the ocean as the debris of dead marine plants and animals sinks to the bottom, the nutrients travel with the upwelling water to the top levels. There the light is sufficient to drive photosynthesis, and the nutrients help the marine plants to flourish. The concentration of nutrients in the Peru upwelling is many times greater than that in the open ocean. In terms of

ANCHOVY TRAWLERS by the score lie at anchor in the fishing port of Pisco in Peru. Structures in the foreground are part of a fish-meal factory. Tonnage of anchovies alone landed in Peru in a normal year outweighs the combined fish catch of any other nation.

CASCADE OF ANCHOVIES spills from a draining grate into the cargo hold of a trawler as crewmen keep watch. The fish gather by the billions near shore and close to the surface, particularly in summer months; trawlers can then net anchovies 100 tons at a time.



TWO NORTH-FLOWING COMPONENTS of the Peru Current are the deep, narrow Coastal Current that hugs the land from Valparaiso in Chile to north of Chimbote in Peru and the deeper and wider Oceanic Current that reaches the latitude of the Gulf of Guayaquil. The anchovies (*dark color*) are normally found in the Coastal Current between 25 and five degrees south latitude; they may consist of a northern and a southern population.

the amount of carbon fixed photosynthetically per cubic meter of water per day, the range in the upwelling region is from 45 to 200 milligrams, compared with less than 15 milligrams in the waters immediately adjacent. Perhaps only one other part of the world ocean is richer: the Benguela Current off the southwestern coast of Africa.

The Peru upwelling sustains an enormous flow of living matter. The food chain begins with the microscopic diatoms and other members of the phytoplankton that comprise the pasturage of the sea. The plants absorb the nutrients and grow in rich profusion, providing fodder for billions of grazing animals, principally minute crustaceans such as copepods but including arrowworms and a wide variety of other small marine herbivores. The food chain can then go on to several more links, progressing from the small fishes that eat the herbivores to the larger fishes and squids that prey on the small fishes and perhaps continuing to include one or more further advanced levels of marine predation. The food chain in the Peru Current does go on in this fashion to some degree, but most of its energy flow stops with the anchovies. This single fish species has succeeded in capturing an extraordinarily high proportion of the total energy available in the ecosystem and in converting it into enormous quantities of living matter. At the height of the anchovies' annual cycle the total bulk of the species is probably of the order of 15 to 20 million metric tons.

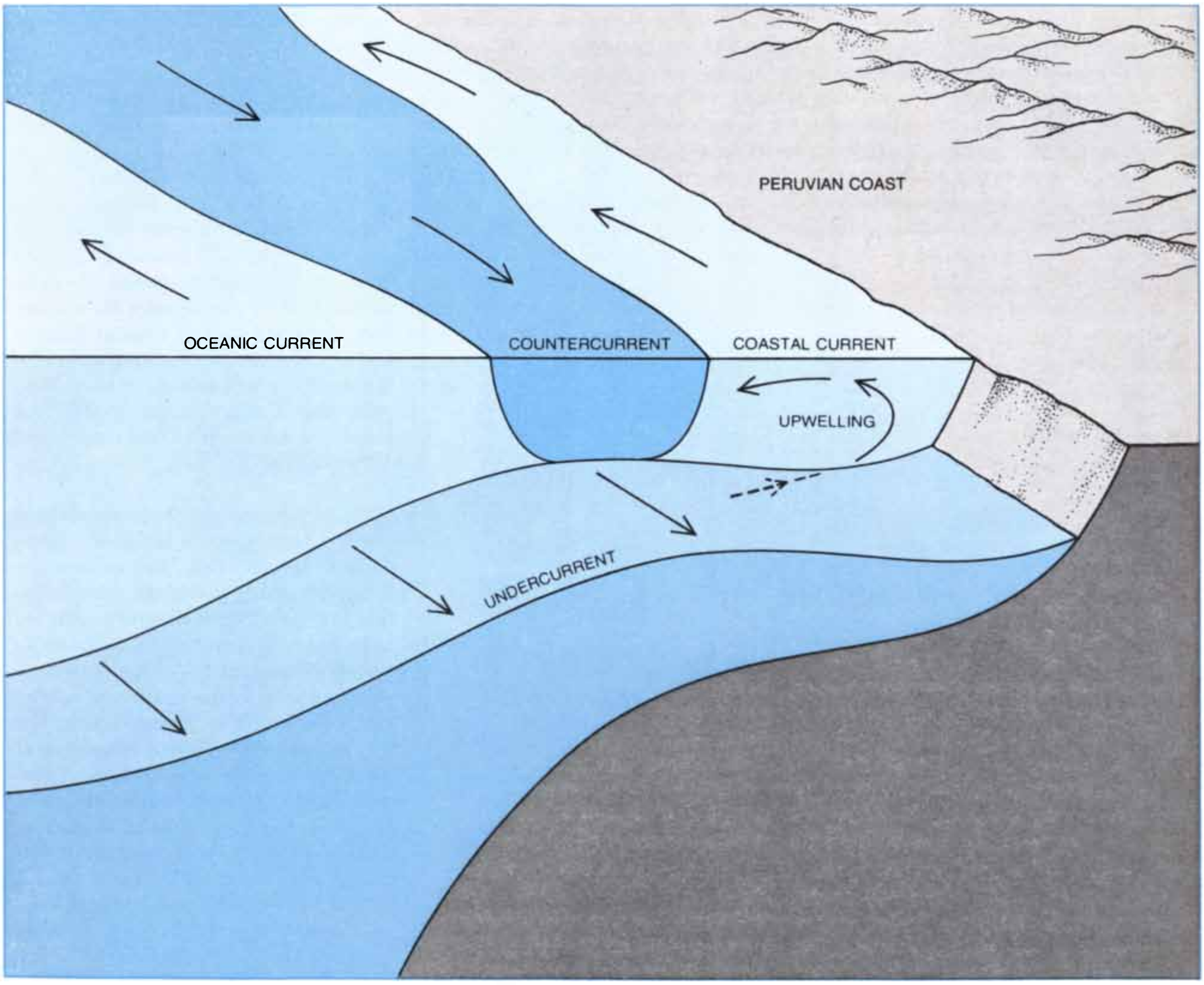
The Peruvian anchovy belongs to the same genus (*Engraulis*) as the common anchovy of the eastern Atlantic and the Mediterranean (*E. encrasicolus*), but it comprises a separate species (*E. ringens*). Its life begins in the form of an egg, a tiny oval spot of nearly transparent protoplasm adrift in the sea. Eggs can be spawned at almost any time of the year but there are two periods when the anchovies' reproductive activity is highest. The major spawning occurs in August and September, during the southern winter, and it is repeated on a lesser scale in January and February. Anchovies are precocious: most females are capable of spawning when they are a year old. By then each female, a little over four inches long, may produce 10,000 eggs. If she survives to the age of two and reaches a length of six inches, her output increases to some 20,000 eggs.

The delicate larvae that hatch from the eggs lead a perilous existence. Many species of fish produce eggs that contain a considerable store of yolk; the reserve of nutrient helps to sustain the newly hatched young until they adjust to finding their own food. The anchovy egg has a negligible yolk store, and so the larva must locate food quickly or starve. To make matters worse, the larva has limited swimming powers and a high rate of metabolism. If more than a few wiggles are required to obtain the food it needs, it will not survive. Because every larva consumes plankton in substantial amounts and because the peak hatches produce larvae numbering in multiples of billions, only enormous swarms of microscopic plants and larval crustaceans can sustain the anchovy stock. Nor is starvation the only peril: the larval anchovies feed swarms of predators. They are eaten by the same copepods that will, if the little fish survive, be the anchovies' own main sustenance. Arrowworms also devour them, and so do their own parents.

One month after the time the anchovy larvae are hatched more than 99 percent of them have perished. Even with such a high mortality rate, a process that begins with billions of spawning fish, each casting 10,000 to 20,000 eggs, produces enormous quantities of anchovy larvae. The little fish grow rapidly; in the course of the first year they attain a length of 4.2 to 4.3 inches. The year-old fish are so slender, however, that they weigh a scant third of an ounce.

The anchovy schools do not move at random; apparently because of a strong preference for the cold water of the Coastal Current they remain within a comparatively restricted zone. The Coastal Current is at its narrowest during the southern summer, running close to shore and seldom exceeding 200 meters in depth. Within this shrunken world the anchovies press together in enormous concentrations near the shore and close to the surface. It is now that predators fare best. The larger fishes and the squids feed well; the several species of guano birds have to fly only short distances from their island nesting grounds and need not dive deep to reach their prey. The greatest of the predators, the fisherman, finds summer work the easiest. He can often set his purse seine within sight of port and gather in anchovies 100 tons at a time.

El Niño occurs at irregular intervals. There is said to be a seven-year cycle, but in actuality the phenomenon is far



TWO SOUTH-FLOWING COMPONENTS of the Peru Current are seen schematically in relation to the two north-flowing components in this diagram. The Countercurrent (*dark color*), a surface or near-surface stream of water, intrudes between the north-flowing cold Coastal and Oceanic currents. Normally the Countercurrent does not extend much south of the Equator, but when the wind that moves the north-flowing currents along falters or changes direction, its warm water pushes far to the south with disastrous biological consequences. Deep below all three currents is the second, far larger south-flowing component, the Undercurrent (*light color*).

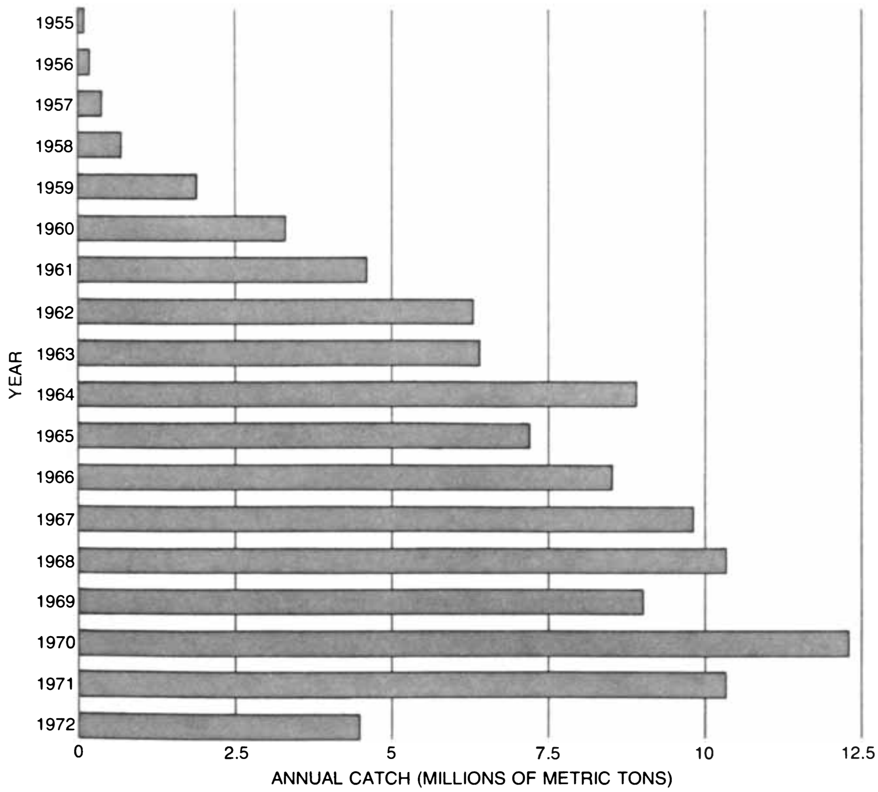
less precise in its appearance. The severe environmental dislocations may be repeated for two or more years in a row or may not recur for a decade or longer. Another kind of regularity, however, has given El Niño its name. The change usually begins around Christmastime and so is given the Spanish name for the Christ Child. The complicated chain of events in a year of El Niño disturbs the anchovies, sometimes profoundly. The wind now comes from the west rather than from the southeast, and it is laden with moisture from the Pacific. With no mountains to rob the air of its burden of water the arid coast is often subjected to torrential rains and severe windstorms. In a desert region where even a heavy mist can cause problems the floods of El Niño are often devastating. Oddly enough, however, in some Niño years no rain falls.

A warning that the sea change may be on its way is given when the temperature of the coastal water begins to rise. If the increase in temperature persists and spreads, the delicately adjusted world of the anchovy tilts. With the warm water come unfamiliar inhabitants of the northern Tropics: the yellowfin tuna, the dolphinfish, the manta ray and the hammerhead shark. Some of them feed on the anchovies. A greater threat to the anchovies' survival, however, is a slowing of the northbound Coastal Current and a decline or even a halt in the usual upwelling of subsurface waters. As the supply of nutrients diminishes, the planktonic plant life that provides the base of the ocean food chain becomes less abundant. As a result herbivorous planktonic animals become scarcer, and so it goes link by link up the chain. Furthermore, the water temperature is now too high to suit the anchovies themselves. Even if the shortage of food has not yet greatly reduced their numbers, the fish scatter, no longer forming the enormous schools that normally afford the guano birds and fishermen such rewarding targets.

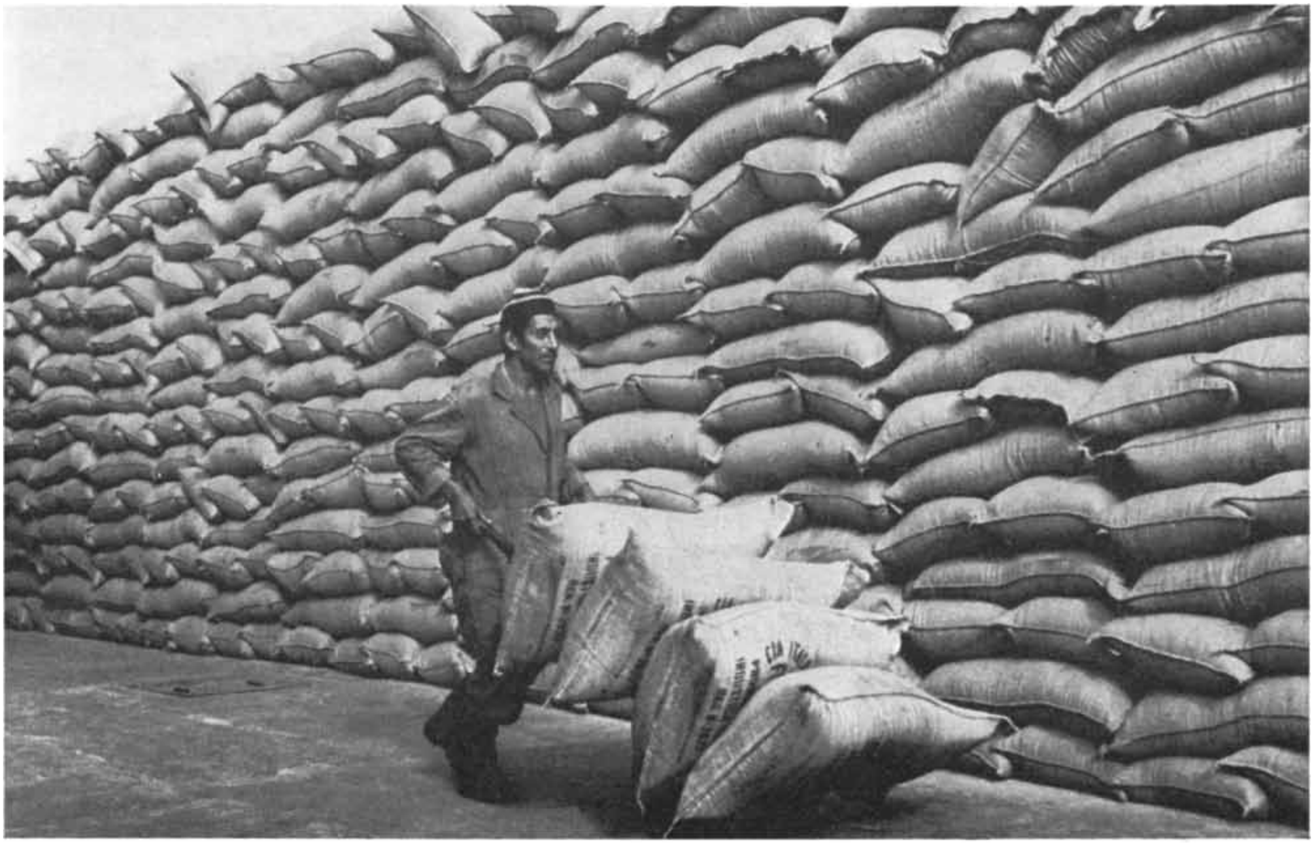
The effect on the guano birds and marine animals is among the most serious of the changes wrought by El Niño. The birds starve or fly away, deserting their nestlings. Fishes, squids and even turtles and small sea mammals die. Their decaying bodies release evil-smelling hydrogen sulfide that bubbles up through the water and blackens the paint on the boats in the harbors. This unpleasant phenomenon is called El Pintor (The Painter). Patches of reddish, brownish or yellow water similar to the "red tides" that upset the Florida tourist industry become relatively common. They are caused by prodigious blooms of dinoflagellates: microscopic planktonic plants that are toxic in high concentrations. The greater frequency of the blooms during Niño years may be because the nutrient composition of the seawater suits the organisms better then or because the less vigorous currents fail to disperse accumulating clusters of dinoflagellates as quickly as usual.

The causes of El Niño are wind changes and sea changes on a very large scale. When the steady southeast trade winds weaken or when the wind blows from the west, the ocean currents that run to the northwest are no longer pushed along with the same vigor. Under normal circumstances the south-flowing Peru Countercurrent is weak, but when the prevailing winds fail or are reversed, the Countercurrent thrusts a tongue of warm water into the cleft between the now less vigorous Coastal and Oceanic currents. As it meets less resistance the Countercurrent penetrates farther south, pushing the weak north-flowing currents aside and covering their cold waters with a 30-meter layer of warm tropical water. The water may have come from as far away as the Panama Bight, north of the Equator, and part of it may even have originated as land runoff in Central America. It can be as much as seven degrees C. warmer than the north-flowing Coastal Current and is lower in salinity, deficient in oxygen and poor in nutrients. In some Niño years the Countercurrent pushes the tropical water as far as 600 miles south of the Equator.

The organisms most obviously affected by a Niño year are the guano birds, various species that are colloquially lumped together under the same name that is applied to the droppings that accumulate in large quantities on the rocky islands where they nest. There are three principal species: the guanay, or cormorant (*Phalacrocorax bougainvillii*); the piquero, or booby (*Sula variegata*), and the alcatras, or pelican (*Pelicanus thagus*). Over the millenniums the bird droppings have accumulated in piles as high as 150 feet on some islands. Because guano is perhaps the finest natural fertilizer known, the guano islands have provided the foundation for a valuable industry. Of the guano birds' diet between 80 and 95 percent is made up of anchovies, and the coastal waters of Peru support what is probably the largest population of oceanic birds anywhere in the world. In recent years estimates of the birds' total number have gone as



ANCHOVY CATCH remained below two million metric tons until the 1960's. Exploitation of the fishery skyrocketed thereafter; the annual rate approached or exceeded nine million tons for six of eight consecutive years. The peak year, 1970, saw a catch of 12.3 million tons.



SACKS OF FISH MEAL awaiting shipment abroad line a wharf at a Peruvian port. The meal is used to enrich feeds for poultry and other livestock. In 1970 the export of fish meal and fish oil earned some $340 million, a third of Peru's total export revenue.

high as 30 million. The five million individuals that inhabited one particular guano island are believed to have consumed 1,000 tons of anchovies a day. The guano birds' annual anchovy catch in recent years is calculated to average 2.5 million metric tons, or between a fourth and a fifth of the commercial-fishery catch.

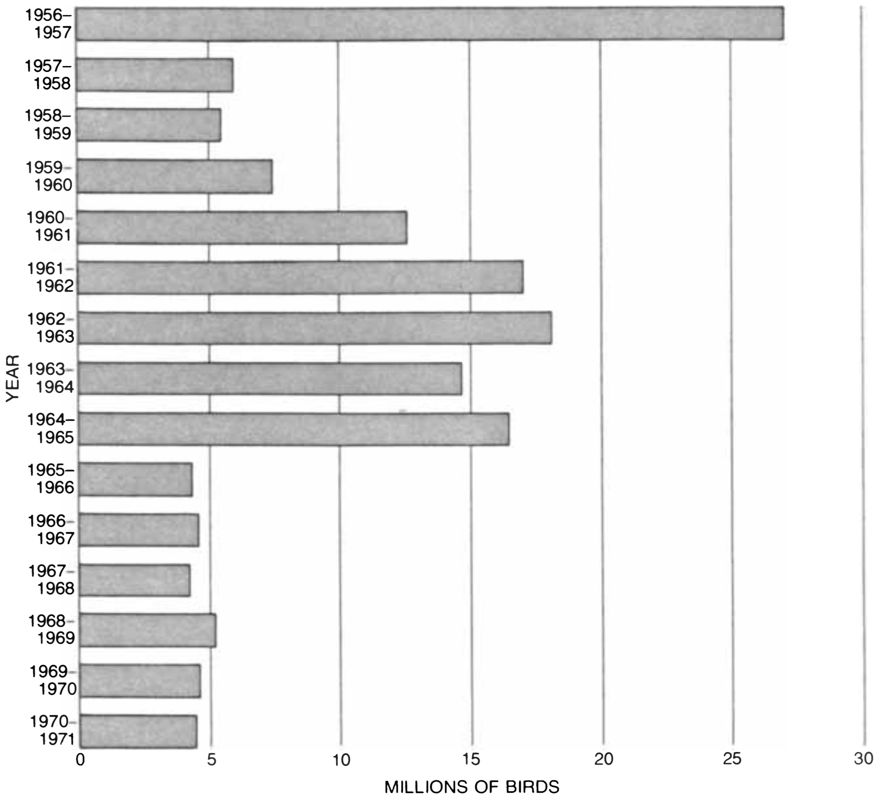
Following every Niño year of any consequence the bird population declines just as the anchovy population does. When the warmer water scatters the dense surface schools of fish, the birds find it harder to feed themselves, let alone their nestlings. Adult birds fly to other areas. Juvenile birds, less efficient fishers than their parents, perish in large numbers. The deserted nestlings are doomed to starvation. After the severe Niño year of 1957 the guano-bird population, then estimated to be 27 million, plummeted to six million and dropped to a low of 5.5 million the following year. Numbers slowly increased thereafter, so that there were 17 million birds when the Niño year of 1965 arrived. That year the population fell to 4.3 million.

Since then the guano birds have failed to recover at the normal rate. There is concern that the commercial anchovy fishery, which has expanded greatly in the same period, is depriving the birds of so much food that their numbers may fall below the level that is critical to their survival as social species. The late Robert Cushman Murphy of the American Museum of Natural History devoted some years to the study of these populations, and it was his opinion that the birds and the fishermen were essentially incompatible. It seems likely, however, that in spite of Murphy's contrary view the two competitors will be able to coexist at some suitable level of commercial fishing. At the same time it may well be that the size of the commercial catch in recent years has prevented the bird population from regaining its former numbers.

It is not commonly known that in the past few years the anchovy fishery has made Peru the world's leading fish-producing nation. Until recently Peru was harvesting anchovies at a rate of 10 million metric tons or more per year. This is a greater weight than that of all the species of fish being caught by any one nation in the Old World, and is twice the tonnage of the combined all-species catch of all the nations of North and Central America. The fish meal made from the Peruvian catch is sold around the world to enrich feeds for poultry and other livestock; the fish oil goes into margarine, paint, lipstick and a score of other products. In 1970 the export of fishery products brought Peru some $340 million, nearly a third of the nation's foreign-exchange earnings. In addition to this the tax revenues from the industry and the domestic employment it provides have become major elements in Peru's economy.

The anchovy industry began in earnest in 1957. Within 10 years the profits that could be made from catching and processing the fish attracted hundreds of fishing boats and led to the construction of dozens of fish-meal factories. No fish stock, however, can stand unchecked exploitation. Government authorities and fishery biologists became concerned about the future of the resource. Peru was a newcomer to large-scale commercial fishing and had neither fishery scientists nor administrators with experience in the complexities of fishery research and management. The Peruvian government turned to the United Nations for help.

In 1960, with a grant from the UN Development Programme and a matching amount in Peruvian funds, the Insti-



GUANO-BIRD POPULATION, hard hit by El Niño in 1957, had regained more than half its former numbers when El Niño reappeared in 1965. The population did not recover at the normal rate thereafter, possibly because the large commercial anchovy catches cut down the birds' food supply. Only a million or so guano birds may have survived the 1972 Niño.

tuto del Mar del Peru was set up to conduct research on the anchovy stocks and to advise the government on management of the fishery. The Food and Agriculture Organization of the UN (FAO) recruited experienced fishery scientists from around the world to work at the institute, conduct research on the anchovy stocks and train a Peruvian staff. Located in Lima, the institute is now a firmly established fishery-research center where more than 50 young Peruvian scientists are conducting the studies needed to establish conservation regulations for the anchovy fishery. The Peruvian staff is advised by a few resident FAO scientists and by a panel of distinguished experts from around the world, organized by the FAO, that meets twice a year.

It has taken biologists nearly a century to unravel the intricacies of fish populations and the complexities of their response to the dual stresses of environmental change and human exploitation. Until last year fishery biologists could point to the management of the Peruvian fishery as an exemplary application of this hard-won knowledge. A major stock had been put under rational control before exploitation had depleted it, and conservation measures seemed to be ensuring an enormously high yield at the limit of the biological capacity of the Peru Current ecosystem. Then in 1972 such pride was chastened, if not utterly humbled.

After the Niño year of 1965 the fishery had enjoyed several very successful seasons, culminating in 1970 with an anchovy catch of 12.3 million metric tons. Then, toward the end of April, 1972, and only a few weeks after the start of the season, fishing suddenly faltered. By the end of June catches had dwindled to almost nothing, and at the close of the 1972 season only 4.5 million tons of anchovies had been harvested. The catch this year threatens to be even poorer. Indeed, there is some reason to fear that the world's greatest stock of fish may have been irreversibly damaged, in which case the Peru fishery would be destined to collapse altogether.

That is the gloomiest outlook. It is based on two disturbing circumstances. First, the size of the "standing stock" (the total anchovy population) now appears to be far smaller than normal. It may be as low as one or two million metric tons, compared with an average of 15 million tons in recent years and an estimated 20 million tons in 1971. Second, "recruitment" (the numbers of fish grown big enough to enter the commercial fishery in any year) has been by far the smallest ever observed. It is scarcely 13 percent of the recruitment in a normal year.

What is causing the trouble? Very possibly El Niño has been a major factor, but some puzzling circumstances have made the scientists closest to the subject uncertain about the extent of the relation between the sea change and the reduction in the anchovy stock. It is clear, however, that the Niño year of 1972 was one of the most severe ever observed. Instead of remaining at the normal level of 22 degrees C. the surface temperature of the Coastal Current rose to 30.3 degrees in February. Although the temperature fluctuated thereafter, it remained high for the rest of the year and was still above normal in January, 1973. It did not fall to near-normal temperature until March of this year. At the same time that the temperature rose the salinity of the water declined from a normal 35 or 36 parts per 1,000 to 32.7 parts per 1,000.

Tropical and subtropical marine plants and animals began to appear far south of their usual limits: dolphinfish, skipjack tuna and the tropical crab *Euphilax.* The guano birds fled from their nesting islands, abandoning their young; their population may now number no more than a million. The warm Countercurrent forced the cold-seeking anchovies so close to the shore that the fishermen often found the water too shallow for their nets. Moreover, the crowded fish did not spawn as abundantly as usual and the eggs and larvae, already reduced in numbers, did not survive at the usual rate because of greater predation by their own close-packed parents.

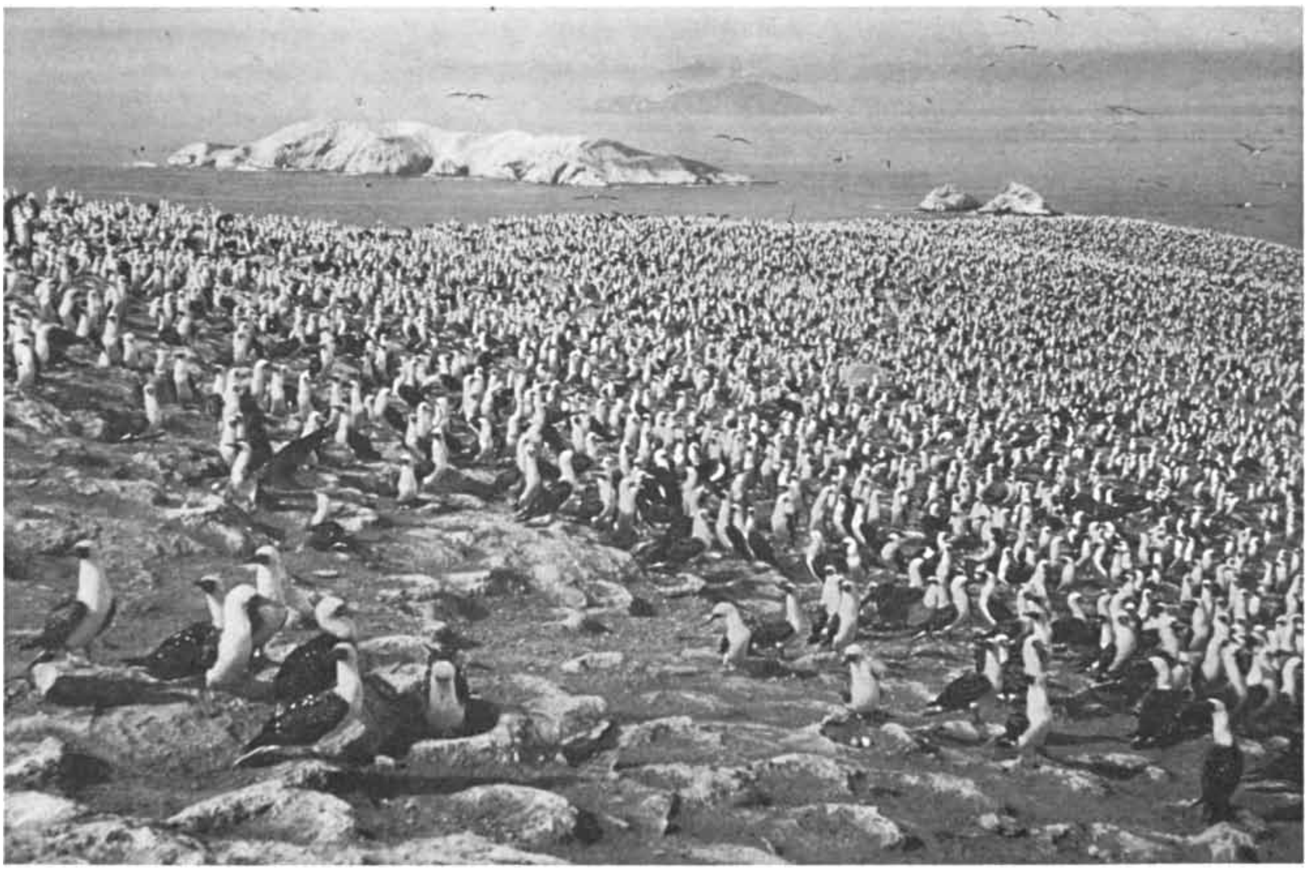
Biologists are nonetheless unwilling to blame El Niño for all these occurrences. For example, with respect to the anchovies they note that recruitment of young fish to the adult population was observed to fall below normal levels before it became obvious that the surface temperature of the Coastal Current had risen. In seeming contradiction to this, however, the tropical crabs too made their appearance before the surface temperature rose, apparently indicating that a body of tropical water had by then already invaded the world of the anchovies. In the light of such oddities the experts frankly admit that they do not know how much influence El Niño exerted on the anchovies in 1972.

Quite apart from the sea change, however, the Peruvian commercial fishery must accept a share of the blame. The 1970 catch of 12.3 million tons considerably exceeded the 10-million-ton level that fishery biologists had estimated to be the maximum sustainable yield of the Peruvian stock. Several economic and political stresses were responsible for the excessive catch. Foremost among these harsh realities is that there are many more fishing boats and fish-meal factories in Peru than are needed to harvest and process the catch. The anchovy fleet is so large that it could harvest the equivalent of the annual U.S. catch of yellowfin tuna in a single day or the annual U.S. salmon catch in two and a half days. The fleet could be reduced by more than 25 percent and still comfortably harvest a rational quota of 10 million tons of anchovies a year. Moreover, the record 1970 catch figure does not measure the full toll the fishery took of the anchovy stock that year. Conservative estimates of losses from spoilage at sea and unloading and processing ashore raise the commercial total to some 13 or 14 million tons.

At this writing the future of the Peruvian anchovy is uncertain. The gloomy forecasts based on biological sampling in 1972 have been confirmed by additional observations early this year and by the results of trial fishing allowed at that time. For three weeks in March the anchovy fleet went to sea and about one million tons of fish were caught. During that brief period the catch per unit of fishing effort (a statistic that provides a measure of the size of the stock) declined rapidly. This suggests that the fleet had caught a significant proportion of all the fish that were available. Most of the catch consisted of fish recruited since July, 1972. There will not be any substantial additions to the stock until this coming October, when the progeny of the present population, much reduced by the fishing in March, have grown big enough to enter the fishery. Even so, fishing was authorized again in April with the quota set at 800,000 tons. Only 400,000 tons were taken. At present the 1973 catch is forecast at no more than three million tons.

If things are as bad as the worst prognostications indicate, the anchovy fishery may, like the California sardine industry and the Hokkaido herring industry, collapse forever. Many aspects of the history of the Peru fishery bear a disturbing resemblance to the events that brought about these earlier disasters.

Nature being what it is, the Peruvian coast will sooner or later once again have normal winds and ocean currents. If the anchovy population has not been too severely reduced, the fishery will then begin to recover. On the other hand, human nature being what it is, difficulties may arise in enforcing soon enough and strictly enough the moderate catch quotas required to avoid overexploitation of the diminished population. Unless such a policy of moderation is achieved, not only will the fish stock suffer but also the world of the Peruvian anchovy will be permanently changed. The guano-bird population will be further reduced and perhaps even eliminated. There will also be enormously complex effects among the many other animals that depend to a greater or lesser degree on the presence of the little fish. Finally, if the anchovies' world is allowed to go awry, the biggest loser will be man. He will have lost not only a rich natural resource but also some of the quality of his own world.



GUANO ISLAND off the coast of Peru is a nesting ground for a vast population of piqueros, or boobies. Cormorants, pelicans and boobies are the principal guano birds. Although now reduced in numbers, they once ate some 2.5 million tons of anchovies per year.

The Dynamics of the Andromeda Nebula

The stars, dust and gas of this spiral galaxy are all in motion. Spectrographic observations show that they do not simply wheel around the galactic center but move in a quite complex pattern

by Vera C. Rubin

All the stars that can be seen with the unaided eye from the earth belong to our galaxy. They are members of a flat spiral system that rotates around a massive center some 10,000 parsecs from the sun. (One parsec is 3.26 light-years.) It is difficult to study the internal motions of the galaxy directly because we are located in its central plane, which is clogged with interstellar dust and gas that at visible wavelengths obscure the galactic center and most of the more distant stars. Thus although we can study some of the motions of the galaxy from the observation of relatively nearby stars and from the radio waves emitted by distant clouds of hydrogen, in order to learn more about the dynamical behavior of galaxies we must turn to other systems. The nearest galaxy closely resembling our own is the Great Nebula in Andromeda, and its internal motions have recently been studied in considerable detail. Some of the results of these studies are quite unexpected.

GREAT NEBULA IN ANDROMEDA, also known as M31, is the nearest spiral galaxy that resembles our own. It was photographed with the 48-inch Schmidt telescope on Palomar Mountain. Rectangle outlines region shown in greater detail on opposite page.

On a clear night away from city lights in the Northern Hemisphere the Andromeda nebula is just barely visible to the unaided eye as a faint elongated patch of light. It was described by the Persian astronomer Umar al-Sufi Abd-al-Rahman in the 10th century; it appeared on Dutch star charts in 1500. It was first observed with a telescope in 1612 by Simon Marius of Germany, who described it as resembling the light of a candle flame seen through translucent horn. In 1781 Charles Messier of France listed it as No. 31 in his catalogue of nebulous objects, and to this day the Andromeda nebula is also commonly known as Messier 31, or M31.

William Parsons of England, better known as the Earl of Rosse, began observing M31 in 1848 with his 72-inch speculum-metal reflecting telescope. His journal of observations was published some 40 years later, in 1885. That same year a brilliant new star—a supernova—appeared near the center of M31. Ultimately this star served as a link in a chain of reasoning that established that spiral nebulas were not nearby clusters of stars or clouds of gas but were stellar systems outside our own. Another link was provided by Sir William Huggins, who obtained the first spectrogram of M31 in 1890, and by Julius Scheiner, who first discussed the spectrum of M31 in 1899. Scheiner recognized that the spectrum arose from the light of many stars rather than from a glowing cloud of gas.

Modern observations of M31 date from 1914, when V. M. Slipher, using the 24-inch refracting telescope at the Lowell Observatory in Flagstaff, Ariz., determined that the solar system and the center of M31 are approaching each other at a speed of 300 kilometers per second. It is now known that most of this observed velocity reflects the motion of the sun around the center of our galaxy. The sun is rotating around the galactic center at a velocity of about 250 kilometers per second in the direction of M31. If we could make observations from the galactic center, they would show that our galaxy and M31 are actually approaching each other at a rather modest speed: some 50 kilometers per second.

Soon after Slipher had made his observations F. G. Pease obtained two spectra of M31 with the 60-inch reflector on Mount Wilson. For the first spectrum he lined up the slit of the spectrograph along the long axis of the tilted galaxy; for the second one he lined up the slit along the short axis. Each spectrum was the product of some 80 hours of exposure time spread out over a period of three months! The absorption lines of the stars in M31 were inclined at an angle in the spectrum taken along the long axis, but they were not inclined in the spectrum taken along the short axis [*see illustrations on page 33*]. Pease correctly inferred that the shift of the lines was due to the fact that the galaxy was rotating.

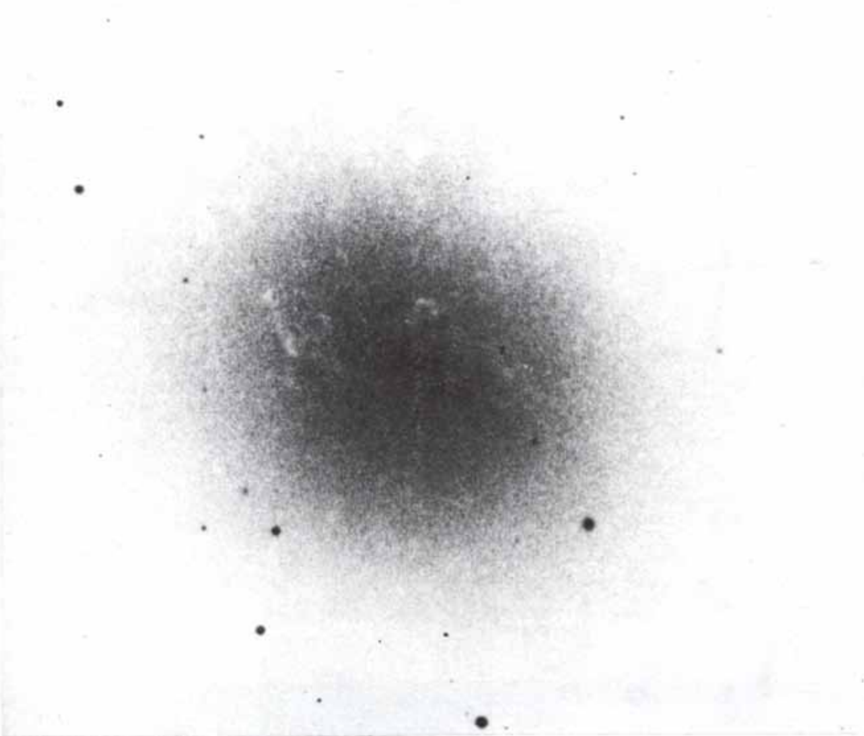
Over the next three decades further observations of the rotation of M31 were conducted by Horace W. Babcock and Nicholas U. Mayall at the Lick Observatory, but the extreme faintness of the individual stars made the exposure times almost prohibitively long and made it impossible to determine velocities very far from the center of the galaxy, where the density of stars is low. In the 1950's and 1960's, however, the situation changed.

First, photography had advanced to the point where photographic plates sensitive in the red region of the spectrum had been developed. Walter Baade, working with the 100-inch reflector on Mount Wilson, detected faint nebulous patches on a red-sensitive plate during a long exposure he was making of M31 for another purpose. He inferred that these nebulous patches, now known as H II regions, were clouds of gas ionized by the ultraviolet radiation of hot stars. The red color of the patches arises from the fact that the gas and dust in M31 absorb the blue light of the galaxy more than red light. Moreover, a significant portion of the light from the H II regions is emitted at two lines in the red region of the spectrum: the line designated hydrogen alpha and a "forbidden" line of ionized nitrogen. (Forbidden lines arise when an electron within an atom drops from a state of higher energy that is metastable, or long-lived, to the ground state. In the diffuse gas of interstellar space an oxygen atom spends approximately 10^4 seconds, or some three hours, in such a metastable state before radiating a strong forbidden line in the visual region of the spectrum. Under normal conditions on the earth the atom would be deexcited by many collisions during that length of time and would never lose its energy by radiating a forbidden line.) During Baade's lifetime he identified 688 H II regions in M31.

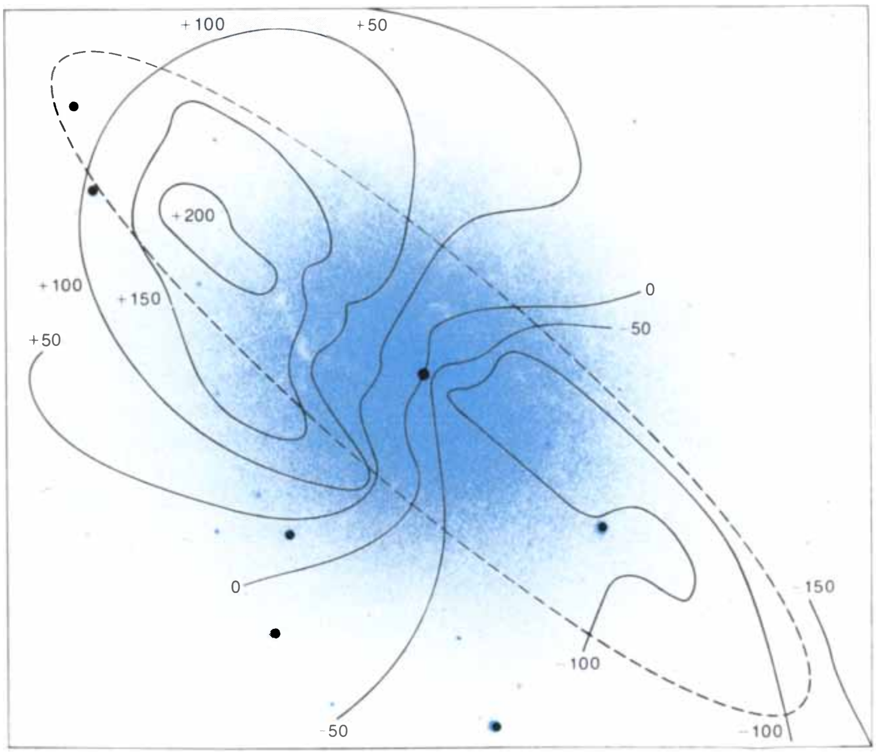
We now know that young, hot Type O and Type B stars and their surrounding ionized H II regions are a major constituent of the spiral arms of galaxies. In the middle and late 1930's, however, Edwin P. Hubble of the Mount Wilson Observatory had been unable to detect any nebulous regions in M31 with photographic plates sensitive in the blue region. He had concluded that the absence of H II regions meant the absence of Type O and Type B stars in the galaxy. It was the availability of photographic plates sensitive in the red that made Baade's later observations successful.

A second change in the situation was that in the early 1960's electronic image-intensifiers, or image tubes, had come into their own. When photons of starlight enter an image-intensifier, they strike a photoemissive surface that ejects electrons. The electrons are accelerated and multiply, and finally they produce an amplified image of the star field or spectrum. Image-intensifiers in the early 1960's were comparable to photographic plates in their resolution and their capacity to distinguish a signal from noise, and they could produce an image 10 times faster. Moreover, their relative gain was

SOUTHERN REGION OF M31 is resolved into individual stars in this photograph taken with the 100-inch reflecting telescope on Mount Wilson. The interstellar dust and gas and the stars in this portion of the galaxy are rotating toward the earth as they orbit around the nucleus; in the northern region of M31 (*not shown*) they are rotating away from the earth. Their velocities can be determined by amount by which lines in their spectra are Doppler-shifted, owing to their motion, relative to their normal positions.



NUCLEUS OF M31, photographed by Walter Baade with the 200-inch Hale reflecting telescope on Palomar Mountain, contains clouds of dust and gas. In this negative print the clouds of dust show up as wispy white tendrils. They show that there are lanes of dust in the center of the Andromeda galaxy as well as in the spiral arms, a discovery that is unexpected and surprising. The observed velocities and directions of the dust clouds (*see illustration below*) fall into a pattern that suggests that they may be falling into the galactic plane.



PATTERN OF OBSERVED VELOCITIES (*black*) of the excited gas in the nucleus of M31 is shown superimposed on the photograph of the nucleus (*color*). The contours represent lines of equal velocity in kilometers per second. Positive values indicate that the gas is receding in the line of sight; negative values indicate that it is approaching. The gas is rotating and expanding asymmetrically. In northern quadrant of the galaxy (*top*) positions of excess positive velocities are correlated with positions of the dust clouds.

even greater in the red, where photographic plates are relatively insensitive and where the H II regions emit their principal radiation. Hence the study of H II regions in M31 with image-intensifiers was full of promise.

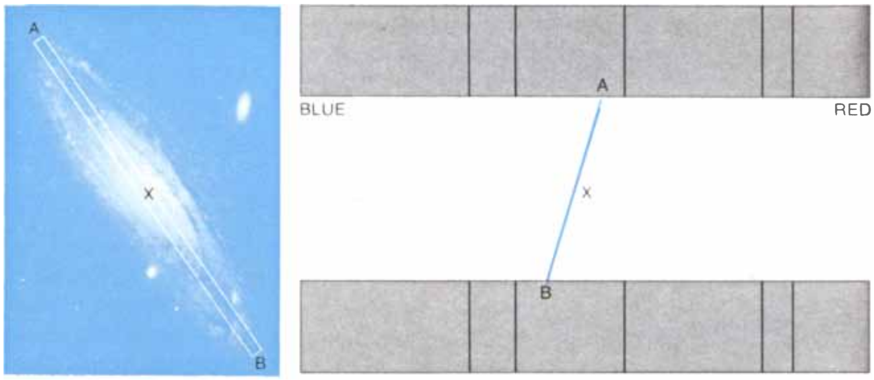
The spectrum of an H II region consists primarily of bright emission lines. Some are recombination lines of hydrogen and helium, which arise as an electron combines with an ionized atom and drops to lower energy levels. Also present in the spectrum are forbidden lines of un-ionized oxygen and lines of singly and doubly ionized oxygen (oxygen stripped of one or two electrons) and singly ionized nitrogen and sulfur. With a telescope of moderate size and an image-tube spectrograph (an image-intensifier attached to a spectrograph) it is possible to record the bright lines in the spectrum of an H II region in M31 in about an hour.

The velocity of each H II region along the line of sight results in a Doppler shift of the emission lines from their normal unshifted position. By measuring the location of the hydrogen-alpha line on the photographic plate to the nearest micron (thousandth of a millimeter) with respect to comparison lines put on the photographic plate at the telescope, one can obtain a velocity for each H II region. From the observed velocity one can compute the H II region's circular velocity around the center of M31. In principle one could observe the spectra of individual stars and obtain their velocities instead of the velocities of the H II regions. However, because the spectrum of a star is continuous, that is, spread out over all wavelengths, a star whose magnitude over all these wavelengths is equal to the magnitude of an H II region will require an exposure that is many times longer. H II regions emit almost all their light in just a few spectral lines. My colleague W. Kent Ford, Jr., and I exploited this fact when we chose to obtain spectra of these regions.

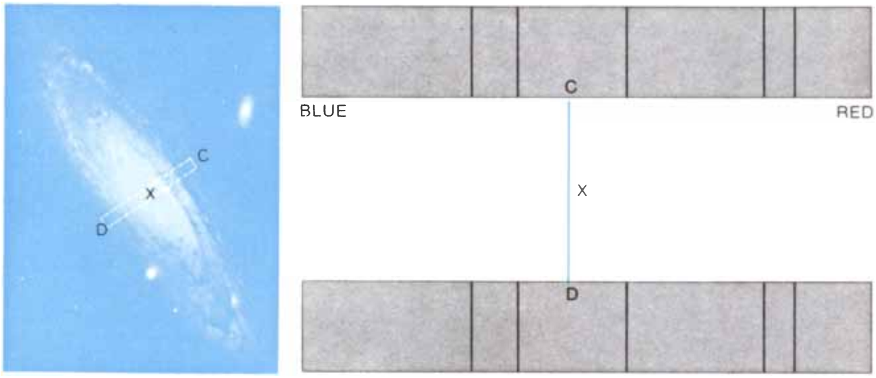
For the past six years Ford and I have been obtaining spectra from selected areas within M31: from 70 individual H II regions that define the spiral arms, from the integrated starlight of the bulge of the galactic nucleus and from a disk of diffuse excited gas within the nuclear region. From these spectra we have mapped regions of differing velocity within M31 and have learned about the variation of the abundance of the chemical elements as a function of distance from the center of the galaxy. We have been working on the spectra of the H II regions since 1966, using a spectrograph that was designed and built in the Department of Terrestrial Magnetism of the Carnegie Institution of Washington.

The spectrograph has been attached to the 72-inch Perkins reflector of Ohio Wesleyan University and Ohio State University at the Lowell Observatory and to the 84-inch reflecting telescope at the Kitt Peak National Observatory near Tucson, Ariz. All the H II regions are too faint to be visible in telescopes of this size. In order to obtain a spectrum of each H II region we had to offset the telescope from brighter visible stars on the basis of positions that had been determined beforehand from long-exposure photographs.

We have not identified any emission regions closer to the galactic nucleus than 3,000 parsecs. On spectra taken of regions near the nucleus, however, a weak forbidden line of singly ionized nitrogen emitted by the diffuse gas does appear superimposed on the spectrum of the integrated starlight. For the gas we have been able to measure velocities to within a few parsecs of the galactic center. We plotted the velocities of the gas in the nucleus and of the H II regions against their distance from the center of M31. The most striking feature of the resulting rotation curve is a deep minimum in the gas velocities at about 2,000 parsecs from the center [*see illustration on next page*]. If we assume that the motions we observe arise from particles of dust and gas moving in circular orbits in the gravitational field of M31, then it is possible to map the distribution of mass within the galaxy from the velocities. In this way we determined the distribution of mass within M31 with respect to distance from the nucleus. That curve too displayed a deep minimum at a radius of 2,000 parsecs from the galactic center. The minimum implies that there is a region of very low mass at that distance.



SCHEMATIC SPECTRUM ALONG LONG AXIS of M31 (*left*) yields spectral lines that are inclined at an angle because of the rotation of stars and gas within the galaxy. White rectangle represents the slit of the spectrograph. With respect to the observer, the northern portion (*A*) of M31 is receding and the southern portion (*B*) is approaching. The resulting spectrum (*right*) shows an emission line from the gas in the northern portion Doppler-shifted toward longer wavelengths, or the red region of the spectrum (*A*) with respect to the undisplaced center of the galaxy (*X*); the same line from the southern portion is shifted toward the shorter-wavelength, or blue, region of the spectrum (*B*). By measuring the displacement of the emission line with respect to the position of comparison lines placed on the spectrum at the telescope, one can determine the velocity of the galaxy's rotation. In actual practice spectrograph slit covers about a hundredth of the long axis of M31. Only for more distant (and hence smaller) galaxies can a single spectrum cover the entire galaxy.



SCHEMATIC SPECTRUM ALONG SHORT AXIS of M31 (*left*) yields undisplaced spectral lines. The western portion (*C*) and the eastern portion (*D*) of the galaxy move across the field of view and have no motion along the line of sight toward or away from the observer. Thus the emission line (*from C to D*) in the spectrum (*right*) shows no Doppler shift with respect to the undisplaced center of the galaxy (*X*) and is not inclined at an angle.

The total mass of M31, out to the last observed H II region some 24,000 parsecs from the galactic center, is 1.8×10^{11} times the mass of the sun. The form of the rotation curve near the nucleus does not affect the determination of the total mass or the distribution of the mass outside the nuclear region of M31. Even if the observed velocity minimum is due to some local disturbance, and does not indicate the overall normal circular gravitational velocity at that point, there is definitely a peculiarity in the velocities of the gas at a radius of 2,000 parsecs from the center.

If the only velocities observed are from stars or gas moving in normal circular orbits around the nucleus of the galaxy, then all the velocities we see along the short axis should be equal to the velocity of the center of M31; there would be no component of the rotational velocity along the line of sight. This is not, however, what the observations show. We have measured the velocities of the gas in the nucleus along the short axis. These measurements indicate that at some points very close to the nucleus, gas is flowing out of it at velocities that range as high as 135 kilometers per second. Only a very small amount of mass, however, is leaving the central region: less than 1 percent of the mass of the sun per year.

Ford and I were intrigued by the presence of this diffuse ionized gas near the nucleus of the galaxy, and so we extended our observations to nearby regions. The emission line of singly ionized nitrogen was difficult to detect above the background radiation from the stars. We attempted to increase our system's sensitivity by increasing the dispersion of the spectra from 135 angstroms per millimeter to 28 angstroms per millimeter. The result of this procedure is that the light from the background stars is spread out by a factor of five and is therefore dimmed, while the sharp forbidden emission line of singly ionized nitrogen is left unaltered. The technique was so successful that we were able to detect five emission lines from the gas where formerly we had seen only one. The only penalty for working at a higher dispersion is one of exposure time: even with our image-tube spectrograph, exposures as long as six hours were necessary.

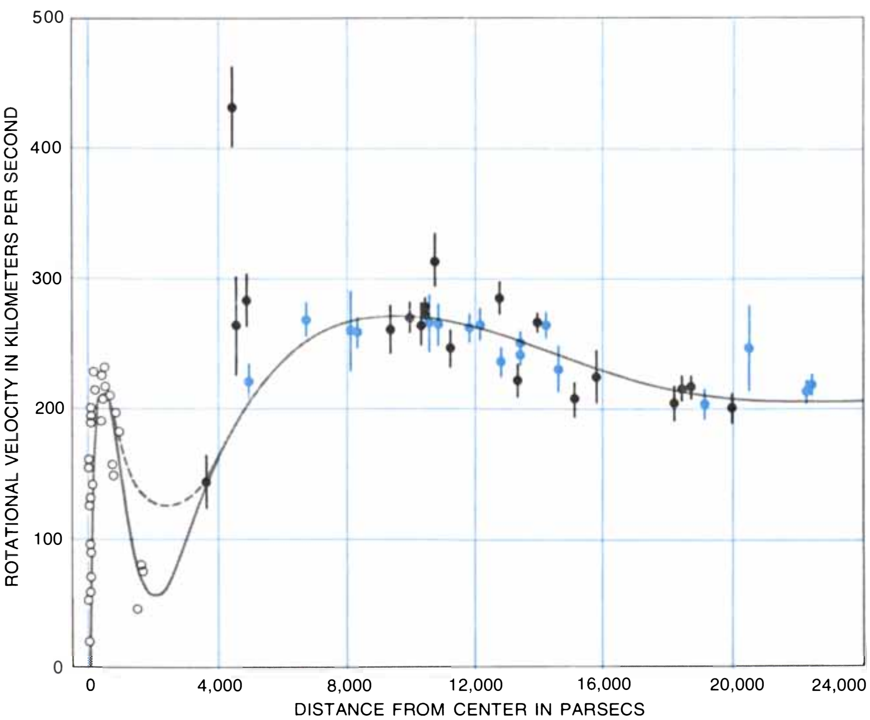
From spectra taken at 16 different angles across the nucleus of the galaxy we have deduced how the diffuse nuclear gas is distributed. Within some 400 parsecs of the nucleus the gas is concentrated into a very flat disk, perhaps 25 parsecs thick. The disk is rotating with velocities that reach 200 kilometers per second about 200 parsecs from the nucleus. In addition to the rotation, gas is streaming outward from the nucleus, principally in two directions about 180 degrees apart near the short axis. The velocities of the streaming gas reach 135 kilometers per second near the far side of the short axis. Moreover, in locations where the spectrograph slit crosses dust patches in M31, the observed velocities suggest that clouds of gas and dust are falling into the central plane of the galaxy. This gas could have been shed by evolving stars, and it could contribute to the mass of the disk. Although this model simplifies some of the complex motions we observe, it does account for their major features.

In 1963 Bernard F. Burke, Kenneth C. Turner and Merle A. Tuve of the Department of Terrestrial Magnetism began studying the motions of the un-ionized hydrogen gas in M31 at the radio wavelength of 21 centimeters. More recent radio investigations have been conducted by Morton S. Roberts of the National Radio Astronomy Observatory and S. T. Gottesman, V. C. Reddish and R. D. Davies of the Nuffield Radio Astronomy Laboratories at Jodrell Bank in England. Since the angular resolution of a radio telescope is lower than the resolution of an optical telescope, the actual detail that can be resolved in M31 is much less. In studies covering all of M31 to date, the diameter of the radio telescope beam on the sky has been 10 minutes of arc or greater. That corresponds to an ellipse some 2,000 parsecs wide by 9,000 parsecs long on M31 because of the foreshortening produced by the fact that we see the galaxy tipped only 13 degrees from edge-on. Our spectrograph slit covers about five parsecs by 400 parsecs on the galaxy. In spite of the difference in resolution, however, the agreement between the rotation curves resulting from the radio studies and those resulting from the visual studies is impressive for distances greater than 3,000 parsecs from the center of M31. (Closer than 3,000 parsecs from the nucleus there is too little un-ionized hydrogen to be detected easily by radio telescopes.)

The radio-wavelength observations have extended our knowledge to the outer limits of M31. I have mentioned that at visible wavelengths we have been able to determine velocities within the galaxy only out to the outermost known H II region 24,000 parsecs from the nucleus. Just last year Roberts and Robert Whitehurst of the University of Alabama extended the rotation curve out to 34,000 parsecs from the center with 21-centimeter observations. They found that the rotational velocity of un-ionized hydrogen remains constant at 200 kilometers per second between 24,000 and 34,000 parsecs from the nucleus. The mass contained between those limits is equal to 10^{11} times the mass of the sun, yielding a total mass for M31 out to a radius of 34,000 parsecs of 3×10^{11} times the mass of the sun. Since we still know very little about the boundaries of galaxies in general, this extension of the diameter of M31 is a matter of some importance.

So far I have discussed only the motions of the gas within M31 and by inference the motions of the hot young stars that ionize the H II regions. The study of the motions of stars themselves is more difficult. Part of the reason is, as I have mentioned, that the individual stars are faint and that their spectra are continuous. We do see broad, diffuse stellar absorption lines in the continuous spectra from regions near the nucleus. Indeed, it is possible to measure the velocities of stars from these lines, although the measurements are less precise than those that can be made with the sharp emission lines from the interstellar gas. Moreover, the analysis is complicated by the fact that stars near the nucleus of the galaxy are distributed not in a flat disk as the gas is but in a nearly spherical bulge. Hence we must obtain the spectra across a long projected path. At large distances from the galactic nucleus the velocities of individual stars are still too difficult to obtain.

We have measured the velocities of the stars on both the long and the short axes near the nucleus of M31 with a single absorption line. We anticipated that the velocities would indicate that the stars are following simple circular orbits around the center of the galaxy, and that they would exhibit none of the complexities of the gas velocities. This, we have found, is not the case. The general features of the stellar motions resemble those of the gas motions. There is a steep

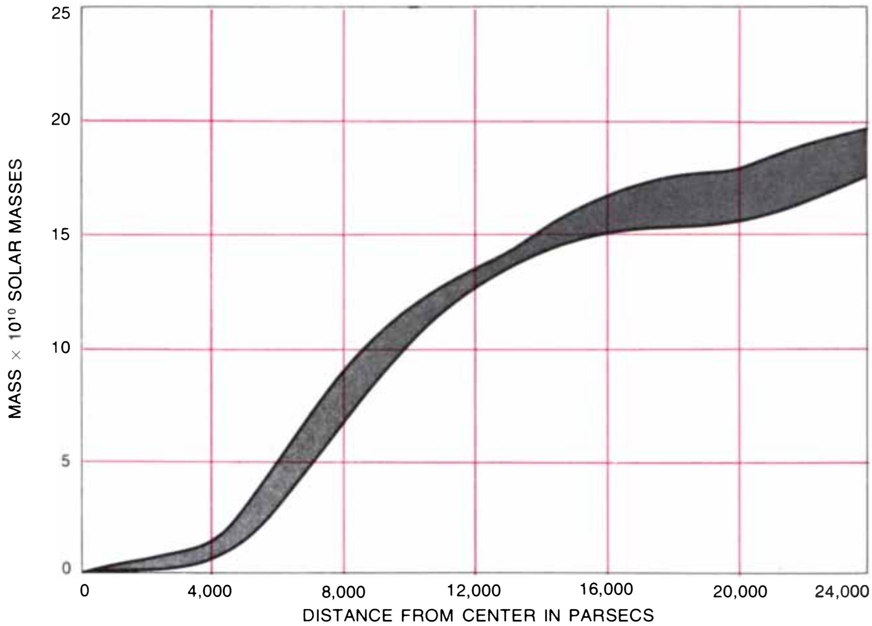


ROTATION CURVE of M31 shows the velocities of the gas within the galaxy as a function of its distance from the galactic center. The open circles represent the velocities determined from the diffuse gas within the nuclear disk. The black dots represent the velocities of the H II regions receding in the line of sight in the northern portion of M31; the color dots represent the velocities of the H II regions approaching in the line of sight in the southern portion. Vertical lines indicate the amount of observational uncertainty for each determination. Although there are two possible versions that have been calculated for the rotation curve near the nucleus (*broken line and solid line*), both alternatives show a peculiar deep minimum in the velocities of the gas at a radius of about 2,000 parsecs from the galactic center.

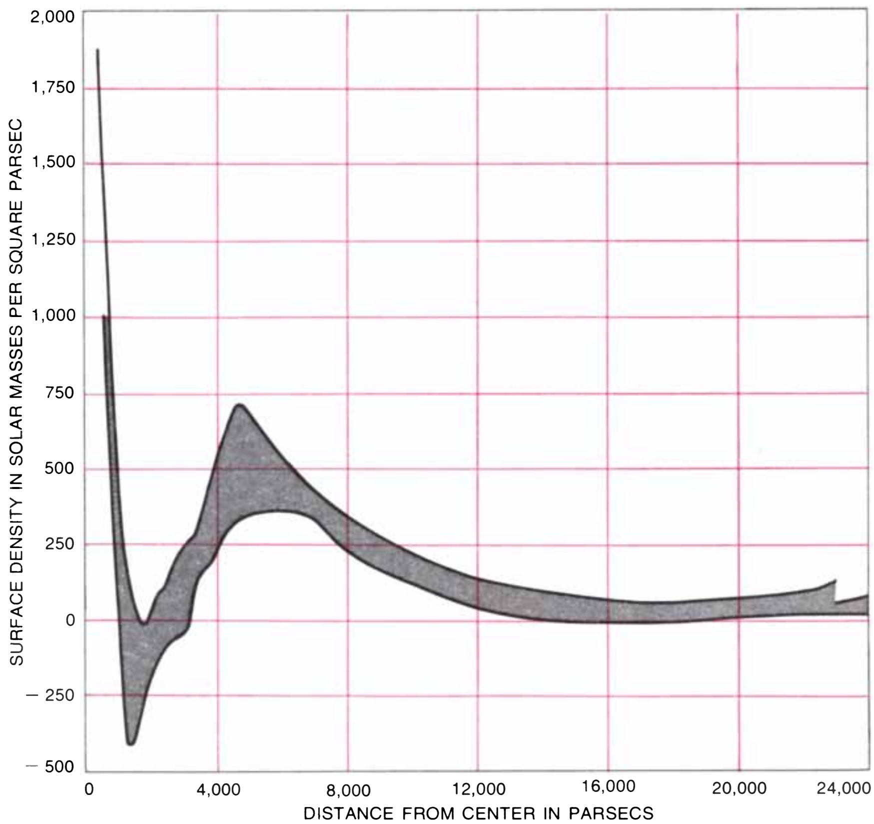
gradient in the velocities of the stars across the galactic nucleus reaching a minimum at 2,000 parsecs, and the stars seem to be moving outward along the short axis. In 1939 Horace Babcock had observed a minimum in the stellar velocities near 2,000 parsecs, but his observations had never been confirmed or widely accepted.

For several reasons these results are both unexpected and not understood. On the basis of our present understanding of stellar evolution it is believed the stars whose spectra we are observing near the nucleus of M31 are some four billion years old. In contrast, the irregularities in the motions of the gas should smooth out in less than 10 million years. Thus it is not clear how the old stars, which during their long life should have moved several times around the galaxy, can have the same motions as the young gas. Moreover, although it is relatively easy to work out mechanisms that impart irregularities to the gas motions, it is difficult to do the same for stars. Our present suggestion is that we are observing stars traveling in noncircular orbits that are dynamically stable. One possibility is that some of the mass of the nucleus of M31 is distributed asymmetrically. The inner spiral arms of the galaxy are decidedly asymmetrical, but it is unlikely that their mass is great enough to distort the orbits of the stars. Perhaps there are resonances, or other cooperative effects, that stabilize the orbits (as gravitational theories of galactic structure now predict). Stars are known to shed mass as they evolve; it is possible that the gas we observe has come from old stars. Detailed understanding must await more observations.

Beyond the region in M31 where the velocities of stars and gas fall to a deep minimum at 2,000 parsecs from the galactic nucleus, the H II regions have circular velocities that rise to 250 kilometers per second at 10,000 parsecs from the nucleus and remain at that velocity out to 24,000 parsecs. We can deduce a remarkably similar pattern of velocities for our own galaxy. A flattened disk of un-ionized hydrogen extends several hundred parsecs from the center, rotating with velocities of up to 200 kilometers per second and expanding at velocities as high as 135 kilometers per second. At a distance of 800 parsecs from the galactic nucleus both the velocity and the density of the gas are very low. Near the position of the sun, 10,000 parsecs from the nucleus, the stars and the gas are rotating around the nucleus at a velocity of some 250 kilometers per



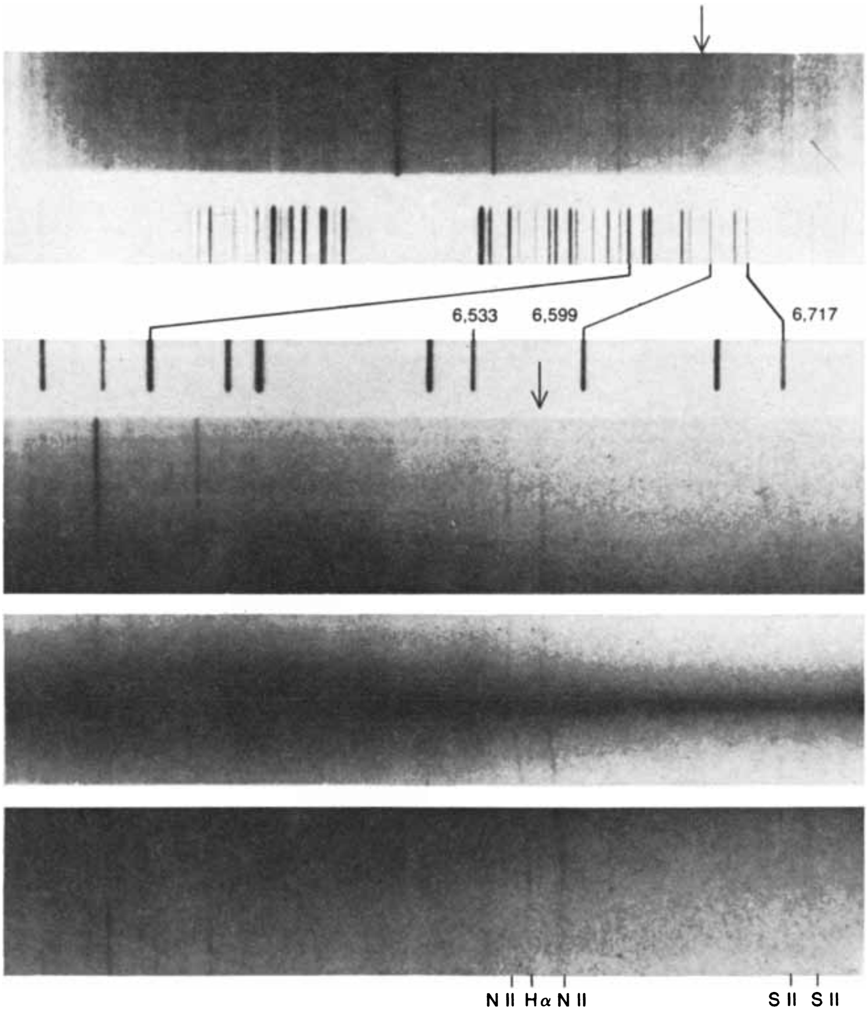
DISTRIBUTION OF MASS within M31 can be determined from the rotation curve of the velocities (*see illustration on opposite page*). The region in gray shows the range of values that the actual curve of the distribution of mass might have, and reflects the observational uncertainty in the rotation curve. The mass distribution displays the same deep minimum at a radius of 2,000 parsecs from the center as the rotation curve does, implying that there is a region of very low mass at that distance. The total mass of M31, out to the last observed H II region some 24,000 parsecs from the center, is about 1.8×10^{11} times the mass of the sun.



DISTRIBUTION OF DENSITY within M31 can also be determined from the rotation curve of the gas velocities. It too displays the same deep minimum, indicating that there is a region of low density some 2,000 parsecs from the center. Negative densities, which arise from the very steep fall of the rotation curve near 2,000 parsecs, are unrealistic and raise the question of whether the observed velocities of the gas near 2,000 parsecs arise from explosive or other nonequilibrium phenomena. For purposes of calculation the flat galaxy was approximated by a disk with no thickness; therefore the density (*vertical axis*) is given in terms of the number of stars per square parsec (area) instead of per cubic parsec (volume).

second. That velocity decreases to approximately 175 kilometers per second at a distance of 24,000 parsecs. The overall resemblance between these figures and those for M31 implies that the distribution of mass in our galaxy is similar to that in M31.

Spectrographic studies of M31 that we have carried out with C. Krishna Kumor of the Department of Terrestrial Magnetism also provide information about the relative abundances of the chemical elements in the galaxy. The abundances can be determined from the strength of the spectral lines emitted by atoms of the elements with respect to the strength of the hydrogen-alpha line. The abundances, like the velocities, vary with distance from the nucleus. In the diffuse gas of the nucleus the spectral lines of oxygen, nitrogen and sulfur are all stronger than the hydrogen-alpha line. Some of the H II regions 4,000 parsecs away from the nucleus also show these anomalously strong lines, although in the normal H II regions at that distance the lines of all three elements are weaker than the hydrogen-alpha line.



SPECTRUM VERY NEAR THE NUCLEUS OF M31 taken at the original low dispersion of 135 angstroms per millimeter (*top band*) is contrasted with three spectra taken at a higher dispersion of 28 angstroms per millimeter (*bottom three bands*). Middle two spectra are laboratory comparison lines of neon; the numbers are wavelengths of certain emission lines in angstroms. In the low-dispersion spectrum the forbidden emission line of singly ionized nitrogen is indicated by an arrow; all other emission lines are due to radiation from the earth's atmosphere. Gray background continuum is from the light of the stars in the nucleus. The exposure time was four hours. Increasing the dispersion of a spectrum spreads out the continuous background radiation and reduces its brightness, while leaving the sharp emission lines from the diffuse gas unaltered. Thus it increases the signal-to-noise ratio for a sharp feature, although longer exposures are necessary. In the three bands of the high-dispersion spectra the emission line of hydrogen alpha (*Hα*), the two weak lines of singly ionized sulfur (*S* II) and one additional forbidden line of singly ionized nitrogen (*N* II) is visible as well as the original one (*arrow*) in the low-dispersion spectrum. The lines are curved because of the rotation of M31. The exposure time was six hours for each spectrum.

Between 5,000 and 15,000 parsecs from the nucleus the nitrogen lines decrease in strength with increasing radius and the oxygen lines increase in strength. We infer from this observation that the abundance of nitrogen is decreasing with respect to the abundance of hydrogen by a factor of about one-half with increasing distance from the nucleus. The abundance of oxygen is also decreasing, but by a small factor: perhaps 1/1.2. The decrease in oxygen has a curious result, because under the conditions of temperature and density in an H II region, radiation by oxygen atoms acts to cool the gas in which they are embedded. Therefore a decrease in oxygen leads to an increase in temperature and to a strengthening of the line of doubly ionized oxygen with respect to the line of hydrogen alpha.

In our galaxy we can observe H II regions only near the sun, that is, at a distance of 10,000 parsecs from the galactic nucleus. In these regions the strongest spectral line is always hydrogen alpha. Leonard Searle of the Hale Observatories has studied giant H II regions in external galaxies other than the Andromeda nebula, and he finds an even larger variation in the strength of the lines of oxygen and nitrogen than Ford and I see in M31. A decrease in the abundance of oxygen, nitrogen and sulfur with increasing distance from the galactic nucleus is probably a general feature of normal spiral galaxies. Perhaps the interstellar gas near the nucleus has been enriched over several generations in the formation of stars. During its lifetime a star transforms by thermonuclear reactions some of the hydrogen in its interior into heavier elements. These elements are returned to the interstellar gas if the star explodes at the end of its life cycle, resulting in an interstellar gas that is richer in heavy elements than it was initially.

Our study of M31 has been a satisfying one. There have been surprises, such as the existence of the disk of gas in the galactic nucleus and the complexity of its motions. There have been puzzles, such as the motions of the stars close to the nucleus. There has been controversy, such as the interpretation of the minimum in the velocities observed near the radius of 2,000 parsecs. Overall, however, we have been able to determine the details of motions in a galaxy much like our own. These results suggest new problems to examine in order that we may better understand the dynamics of spiral galaxies in general and of our own galaxy in particular.

Some people worry about dampness in the cellar

Others let nothing worry them, saying, "Life is too short. There are still wonderful places to go and questions to be asked. Just look beneath the surface."

Soon, just to look is not quite enough. It is more fun to take a camera down beneath the surface.

How much Kodak product is used beneath the surface of lakes, oceans, estuaries, and right offshore from familiar summer beaches, we do not know. Quite a bit, we suspect. The popular underwater press may rival the under*ground* press in circulation, to say nothing of professional publications in underwater science and technology. Professionally, though, the fun part poses a problem. Particularly in the U.S.A., where taxpayers who do not enjoy their work are free to resent those who do. In other lands and times, the line between serious amateur and self-confident professional has not been drawn so hard.

How serious do you want to be underwater? How much confidence have you in your familiarity with underwater photographic principles and practice? Bluegill or coelacanth? A pretty lady lurking in the coral or an assessment of cropping pressure on a stand of phaeophytes? Do you know more of archaeology than of the severe limitation on infrared underwater?

Anyone who writes "Underwater Photography" on a self-addressed envelope at least 4" x 9" and sends it to Kodak, Dept. 55Y, Rochester, N.Y. 14650 admits ignorance of the subject and a desire to receive a small introductory pamphlet. When professional need is claimed, the same address can supply a lengthy bibliography on underwater photography and photogrammetry and a list of manufacturers reported to us as active in underwater photographic products, including underwater housings for Kodak still and movie cameras. Kodak XL movie cameras, designed as they are to use light very efficiently, have done some fine work underwater. We detect no great demand for special films for underwater photography. That may come.

The specimen below of medium-grade advertising art of the mid-'60s was carefully arranged to impress upon the mind that there are many EASTMAN Organic Chemicals.

Some 40 years earlier there had been a proud day when we could offer more than 100 compounds for the "chemical laboratory." Needs change, fashions in scientific emphasis change. Today the preparation of 5 grams of real good quality 2,3,5-"X" can be proposed as a fine, challenging project for the hyperambitious sophomore in Chem 172 whose instructor happens to play squash every Thursday with the first chemist of the entire profession in seven years to require a gram of that compound.

It is in the following categories that we now particularly solicit opportunities to meet needs through our talents and facilities for chemical synthesis and allied manufacturing arts:

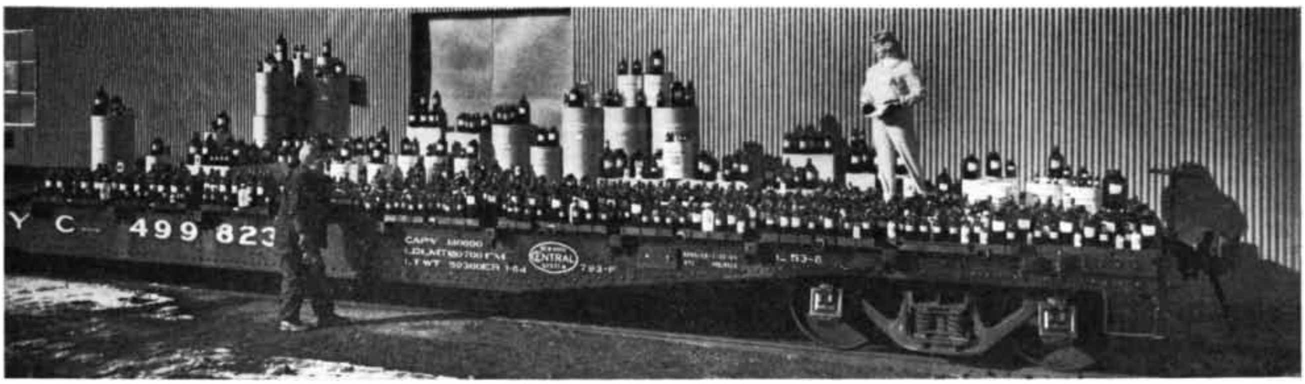
In biology and medicine: Compounds and formulations of recognized or prospective broad applicability in protein and polypeptide synthesis and analysis, in TLC and GLC, in fluorescent labeling, in scintillation counting, and in the various electrophoretic techniques. Also quantity production of ingredients and intermediates for manufacturers of pharmaceuticals and clinical reagents.

In electro-optical science and engineering: Liquid crystal preparations, compounds for the laser art, and spectral sensitizers beyond the scope of photography.

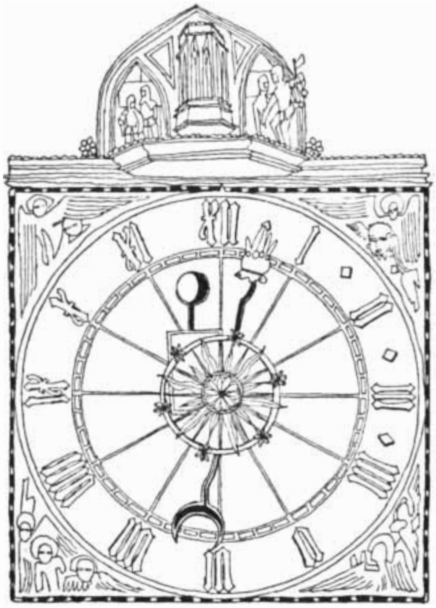
Correspondence and inquiries should be directed to Marshall Guntrum, Eastman Kodak Company, Rochester, N.Y. 14650.

Kodak

Adaptable to the environment



SCIENCE AND THE CITIZEN



Looking Backward

Astronomers at the Steward Observatory of the University of Arizona have reported what may be the most distant, and hence the most ancient, object yet identified on photographic plates. Its light may have taken 11 billion years to reach us. The object was originally detected as a radio source by John D. Kraus and his colleagues at the Ohio State University Radio Observatory, who listed it as OH 471 in the Ohio sky survey. Subsequently the Ohio group found that the source coincided in position with what otherwise seemed to be an ordinary faint star (18th magnitude) in a densely crowded field of stars. The identification was confirmed by precision radio measurements made at the Royal Radar Establishment at Malvern in England.

Suspecting that OH 471 might be a quasar, or quasi-stellar object (QSO), R. F. Carswell and P. A. Strittmatter recorded the object's spectrum with the 90-inch reflector at the Steward Observatory. The spectrum confirmed that OH 471 was indeed a QSO: a starlike object whose radiation is shifted toward the red end of the spectrum. Carswell and Strittmatter report in *Nature* that OH 471 exhibits a red shift of 3.4. Thus the prominent Lyman-alpha emission line of hydrogen, normally found at a wavelength of 1,216 angstroms in the ultraviolet part of the spectrum, was found to be shifted to 5,341 angstroms, near the middle of the visible spectrum. The highest red shift previously observed is 2.88 for the QSO 4C 5.34. In comparison the highest red shift known for an ordinary galaxy is .46.

According to the standard interpretation, all objects outside our own galaxy except for a few of our closest neighbors exhibit a red shift because they are receding as part of the general expansion of the universe; the greater the red shift, the higher the speed of recession and the more distant the object. OH 471's red shift of 3.4 implies that it is receding at 90 percent of the velocity of light, which corresponds to a distance of about 11 billion light-years.

Several hundred QSO's have been identified since the first one was discovered almost exactly 10 years ago. Because all but a few of the known QSO's have red shifts below 2.5 it was conjectured that objects with greater red shifts are genuinely rare. A quasar with a red shift of 2.5 is already so distant that its light has been traveling through space for more than 85 percent of the age of the universe, which is currently estimated to be about 12 billion years. Thus the light from OH 471, with a red shift of 3.4, may have started on its way only about a billion years after the "big bang" that hypothetically created the universe as we know it.

Maarten Schmidt of the California Institute of Technology has estimated that the number of QSO's may have reached a maximum when the universe was about 1.5 billion years old and has been declining ever since. Because the lifetime of a quasar is probably well under a billion years, the overwhelming majority must have evolved by now into less luminous objects, perhaps ordinary galaxies. On the other hand, the energy radiated by most QSO's is only now reaching our instruments. Schmidt estimates conservatively that more than a million QSO's should be observable with the largest optical telescopes.

The problem is to locate them. Many QSO's are "quiet" at radio wavelengths, and on photographic plates they are almost indistinguishable from ordinary stars. One distinguishing technique that has been successful in uncovering QSO's with moderate red shifts is to compare the strength of stellar images on separate blue-sensitive and ultraviolet-sensitive plates. QSO's with red shifts of up to about 2.2 usually appear brighter on the ultraviolet exposures, but at higher red shifts this distinguishing characteristic tends to disappear. OH 471, for example, is neutral in color. No one has yet come up with a good optical method for discovering high-red-shift QSO's that are also radio-quiet.

Flying Circus

Ever since the announcement in Moscow last spring that the first round of the strategic-arms-limitation talks (SALT I) between the U.S. and the U.S.S.R. had succeeded in some measure in curbing the nuclear-arms race in the area of strategic offensive missiles and defenses against them, it has been argued in certain circles that the U.S. should now be redoubling its efforts to modernize those strategic-weapons systems not expressly prohibited by the SALT I accords. For example, in recent months the Air Force has pressed plans in Congress and elsewhere for an entirely new long-range bombing aircraft, to be designated the B-1 [see "Science and the Citizen," May].

It now appears that another aspect of the "strategic modernization program" envisioned by the Department of Defense is an attempt to revive the long dormant idea of an active defense against Russian strategic bombers. Included in the U.S. military budget for fiscal year 1974 are funds for two projects regarded as key components of a projected continental antibomber system. The two proposed projects are AWACS (for airborne warning and control system) and OTH-B (for over-the-horizon backscatter radar).

According to a statement by Secretary of Defense Elliot L. Richardson, released recently to a subcommittee of the House Committee on Appropriations, AWACS aircraft are "designed to detect, identify and track approaching aircraft, and if they are determined to be hostile, to direct our interceptors against them. A small force of AWACS aircraft could replace the bulk of the existing ground-based aircraft warning and control system, which is quite vulnerable to nuclear attack." As currently conceived, the AWACS scheme would consist of an air-surveillance radar and the associated data-processing and communications equipment, all installed in a modified Boeing 707. Unlike the ground-based systems, this airborne system would be capable of detecting and tracking aircraft flying at low altitude over land or water.

During 1972, the Richardson statement reports, "two prototype radars for the AWACS system were flight-tested in Boeing 707 aircraft.... Analysis of the test results has been completed, and the radar built by Westinghouse was selected on the basis of superior performance." Continued development, testing and procurement for the AWACS project will cost, according to Department of Defense estimates, a total of $210 million during fiscal year 1974.

The OTH-B program, said Richardson, "would provide two fixed-base radar systems—one facing east and one facing west—for the long-range detection of aircraft approaching the North American continent.... The FY 1974 Budget includes $5.5 million to continue the OTH-B development program."

Although these funding levels for the next fiscal year may seem comparatively low, critics of the new expenditures point out that the long-term commitment involved is likely to be much greater. For example, according to a recent non-Government study of the U.S. air-defense modernization program by two independent strategic analysts at Washington University, the AWACS project alone will eventually cost "a minimum of $2 billion." Even this amount, the authors of the study, Richard D. English and Dan I. Bolef, contend, "represents merely the camel's nose of the modernization program planned for the Aerospace Defense Command (ADC)."

A third major component of the program, English and Bolef point out, "is the improved manned interceptor (IMI). No fewer than four different aircraft are under consideration for the 1970–1980 interceptor role, while for the 1980's and beyond six other projects are under serious study. The entire program deserves critical scrutiny.... Unless a public debate is opened on this subject U.S. taxpayers may find themselves once again committed, without their knowledge, to an expensive and perhaps unnecessary military program."

In their own review of the strengths and weaknesses of the proposed ADC systems, English and Bolef find much to be critical of. Besides questioning the main rationale given for the current revival of the continental air-defense idea ("the supposed 'continuing threat' posed by the existing U.S.S.R. bomber force and by the claimed deployment by the Russians of a new, improved strategic bomber") the two analysts argue that the specific system outlined by the Department of Defense "seems glaringly deficient. Granted that the radar and control elements would be far more effective than ground-based systems and more likely to survive, what of the weapons themselves? Both the SAM-D [a projected new surface-to-air missile] and the IMI are themselves ground-based units, vulnerable to attack by offensive missiles. It is for precisely this reason that the present ADC system is considered useless."

Moreover, write English and Bolef, the deployment of AWACS would have "much wider and more serious implications than its role in the proposed air-defense system would indicate. ADC is not the only Air Force command interested in AWACS capabilities. The Tactical Air Command (TAC) also wants AWACS—as a flying battlefield command post.... Thus AWACS is an essential element of a future U.S. capacity to launch and pursue an automated Vietnam-type war more quickly and efficiently."

Perhaps the best reason for opposing air-defense modernization, English and Bolef conclude, is that the proposed system, "possibly unnecessary, probably ineffective, could have as its only casualty the present and future SALT agreements it is supposed to encourage.... Should the Russians conclude that a new U.S. air-defense system coupled with a new U.S. bomber (the B-1) would really threaten their security, the least they could do would be to respond in kind. The worst they could do, should the costs of bombers and bomber defenses seem too great, would be to abrogate the SALT agreements."

The Growth of Nations

Every nation for which statistics are available showed economic growth during the decade of the 1960's, but the rates of growth ranged from an average of 25 percent per year to less than 1 percent. The data are assembled in the *United Nations Yearbook of National Accounts Statistics, 1971,* which reviews economic development during the decade that the UN designated in 1961 as the Decade of Development. Classifying nations as developing, developed market economies (meaning nations that do not have centrally planned economies) and centrally planned economies (the U.S.S.R. and the other countries of eastern Europe), the yearbook reports that the developing countries grew at an average annual rate of 5.2 percent, the developed market economies at 5.1 percent and the centrally planned economies at 6.7 percent.

Among developing countries the rate of growth ranged from an average of 25 percent for Libya to about 1 percent for Haiti, Guyana and Uruguay. Developing countries on the high side (7 percent or more) were Zambia, Panama, Mexico, Iran, Saudi Arabia, Thailand and the Republic of Korea. In the second half of the decade growth quickened in the developing countries (to 5.8 percent from 5 percent) but slowed in the developed market economies (to 4.6 percent from 5.3 percent) and in the centrally planned economies (to 5.9 percent from 7.1 percent).

The yearbook also provides estimates of per capita income in current U.S. dollars. Among the market economies the average of $920 in 1970 conceals a wide disparity between $2,660 in the developed nations ($4,274 in the U.S., which was at the head of the list) and $210 in the developing countries. At the beginning of the decade the disparity was smaller: $1,360 in the developed countries and $130 in the developing countries. Summarizing the report, the UN's Centre for Economic and Social Information noted that the figures on per capita income reveal "a continually widening gap in the... economic well-being" of the two groups.

Man among the Man-Apes

When did man arise? In recent years serious estimates for the first appearance of the genus *Homo* have ranged from 750,000 years ago to nearly two million years ago. Now it seems that the genus may be a million years older. If that is the case, a species of *Homo* lived contemporaneously with "man-apes" of the genus *Australopithecus* three million years ago.

The evidence for the greater antiquity of *Homo* is some 150 fragments of a single skull found near Lake Rudolf in Kenya last year by workers under the direction of Richard E. F. Leakey of the National Museum of Kenya. The fragments are from a stratum of sandy soil that lies 35 meters below a layer of volcanic rock known, on the basis of potassium-argon dating, to be 2.6 million years old. Leakey believes the age of the deeper stratum is 2.9 million years.

When the fragments were assembled, they made up much of the skull of a hominid with a cranial capacity of some 810 cubic centimeters. There are two species of *Australopithecus;* the smaller one has a cranial capacity of about 420 cc. and the larger a capacity of about

530 cc. The cranial capacity of *Homo sapiens* averages 1,400 cc.; the capacity of his fossil relative *Homo erectus* ranges from 900 to 1,100 cc. Leakey notes that the cranial capacity of the Lake Rudolf skull is too large for *Australopithecus,* and that the skull is quite unlike the skull of *Homo erectus.* He therefore proposes classifying the fossil as *Homo* sp. indet. (species indeterminate).

The oldest *Australopithecus* fossils date back 5.5 million years; the youngest, less than a million years. Leakey's father, the late L. S. B. Leakey, believed that a hominid contemporary of *Australopithecus* he had found at Olduvai Gorge in Tanzania was a member of the genus *Homo,* and he classified it *Homo habilis.* Moreover, at Swartkrans in South Africa, J. T. Robinson found among plentiful *Australopithecus* fossils fragmentary remains that were later classified as belonging to the genus *Homo.* The big-brained hominid of Lake Rudolf, however, is at least a million years older than any known so far.

Survival Centers

There are 291 species and subspecies of mammals and 340 bird forms listed as being rare or endangered in the Red Data Book published by the International Union for the Conservation of Nature and Natural Resources (IUCN). Approximately half of the endangered mammals and a fifth of the endangered birds are in zoo collections. As more and more mammals and birds in the wild are threatened with extinction, the use of zoos as survival centers for endangered species becomes urgent.

Przewalski's horse and the wisent (the European bison) are often cited as examples of the survival of species by breeding in zoos. Two papers in the journal *Zoologica* by John Perry and Dana L. Horsemen of the National Zoological Park and Donald D. Bridgwater of the Minnesota State Zoological Gardens report that six other species and subspecies of mammals and 10 bird forms have the potential to survive in zoos without further acquisition from the wild. The mammals are the Siberian tiger, the onager (a wild ass), the Formosan sika (a deer), the Père David deer, the scimitar-horned oryx and the addax (both antelopes). The birds are Swinhoe's pheasant, the mikado pheasant, Hume's pheasant, Elliott's pheasant, the nene and the cereopsis (both geese), the Laysan duck, the Hawaiian duck, the Cuban tree duck and the turquoise parakeet.

Mammals showing promising population increases are the mongoose lemur, the Golden Lion marmoset, the orangutan, the Sumatran tiger, the snow leopard, the Hartmann mountain zebra, the black rhinoceros, the pygmy hippopotamus, the vicuña and the Arabian oryx. Other mammals that appear to be breeding well in captivity but are not self-sustaining include the yellow-footed rock wallaby, the North China leopard, the Nubian wild ass, the Thailand brown-antlered deer and the Arabian gazelle.

Although some captive breeding has been successful, zoos on the whole remain consumers of wildlife rather than producers. Zoos with the largest numbers of a rare or endangered species tend to produce a disproportionately large share of the births. Perry, Horsemen and Bridgwater state that many of the species would undoubtedly multiply to satisfactory numbers if adequate breeding groups were brought together under the proper conditions. Most zoos, however, still emphasize diversity in collections. Room for growing numbers of endangered species must be found either by displacing more common species or by establishing rural survival centers. Unless zoos forgo an animals-for-exhibit-only philosophy and undertake survival-breeding programs, Perry, Horsemen and Bridgwater conclude, they "may be in danger of failing in their preservation and propagation objectives."

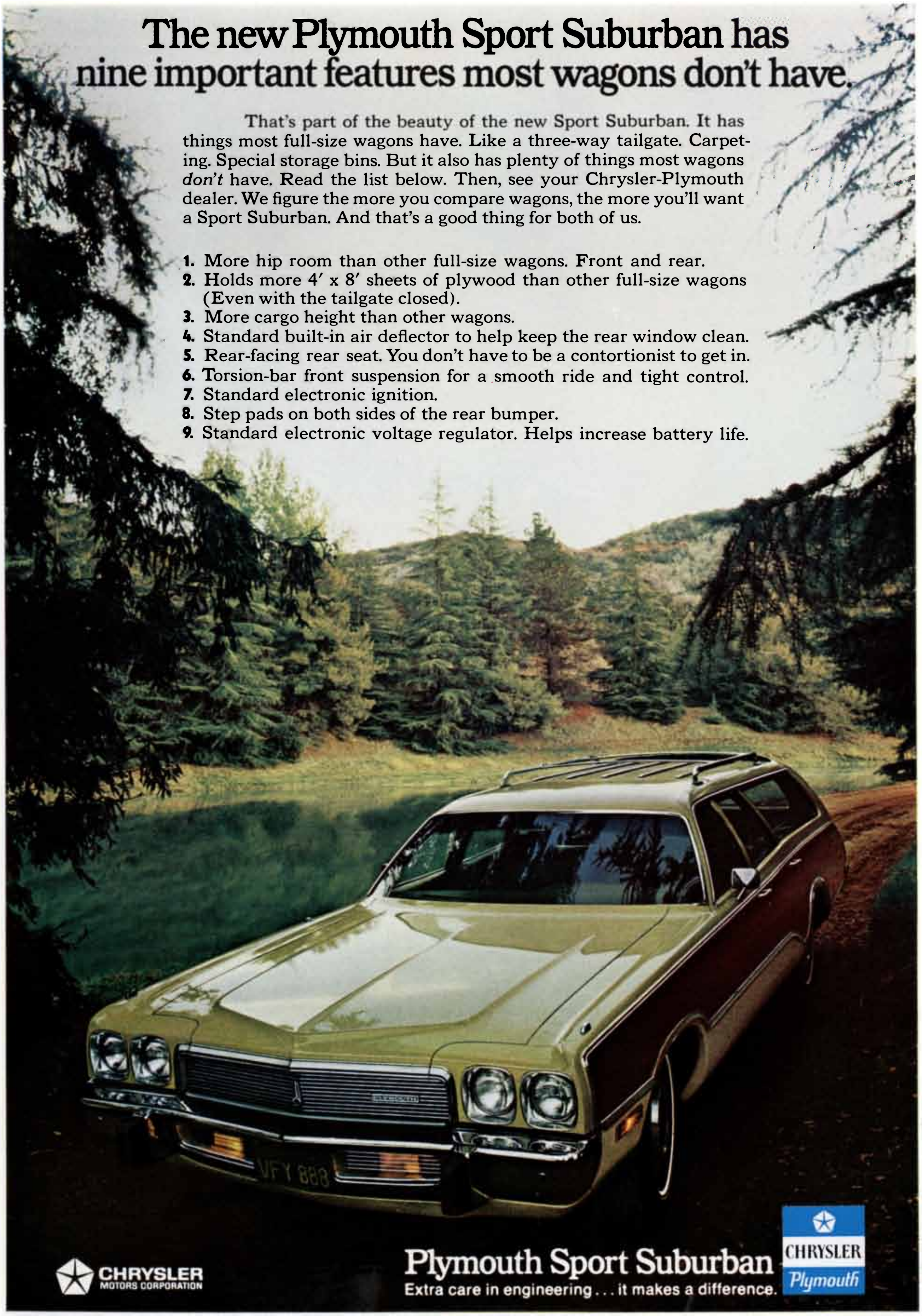
John's Other Life

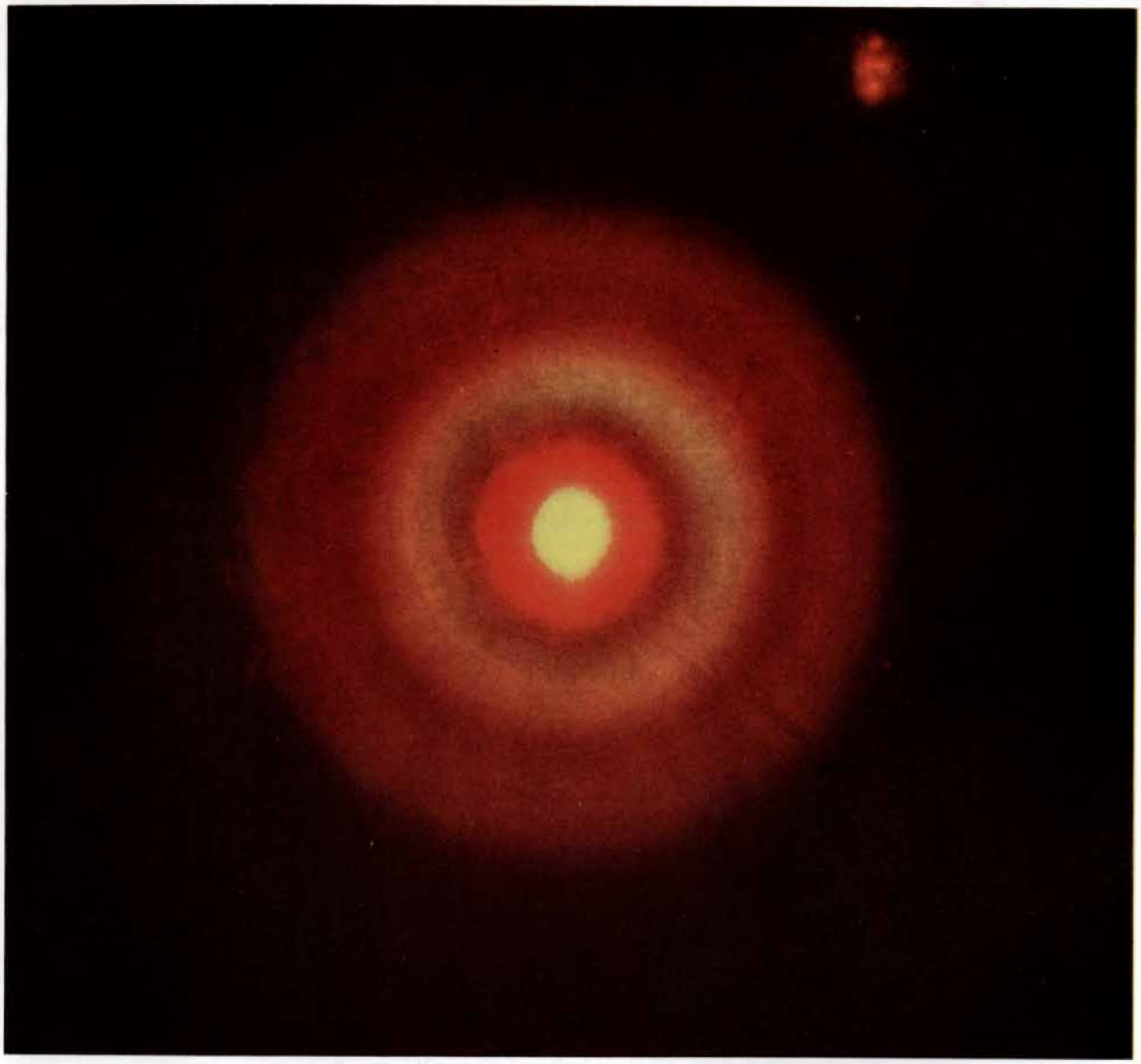
Most people who know the English poets know that John Keats died tragically early (in 1821, at the age of 25, of tuberculosis) but few are aware that he crowded two careers into his brief life: he studied medicine and was trained in surgery for as many years as he wrote poetry. Even people who know of Keats's early training have assumed that he was pushed into medicine against his will, resented it and left it as soon as possible. Not so, according to Robert Gittings, an English biographer of Keats. The poet picked a medical career on his own, worked hard at a sound training program and did well, and had actually qualified as a general practitioner when he gave up medicine to devote the rest of his life (four years, as it turned out) to the joys and pain of writing. Gittings tells the story in an article in *Journal of the American Medical Association.*

They did medical education differently in those days. Orphaned at 14 by the death of his mother, Keats was apprenticed by the age of 15 to a general practitioner, Thomas Hammond. Three years later he began to write poetry, but what he did was "largely devoid of both talent and promise." He completed the full five-year apprenticeship and by his 20th birthday had moved to Guy's Hospital in London and registered at the medical school of Guy's and St. Thomas's. He attended lectures on the theory and practice of medicine, on chemistry and on anatomy and physiology and—remarkably quickly—was given the post of "dresser," or assistant to a surgeon. At the same time, however, he was becoming aware of Wordsworth, from whom he assimilated not only the older poet's exact imagery of natural things but also a concept of the poet as "a healer of the ills of mankind."

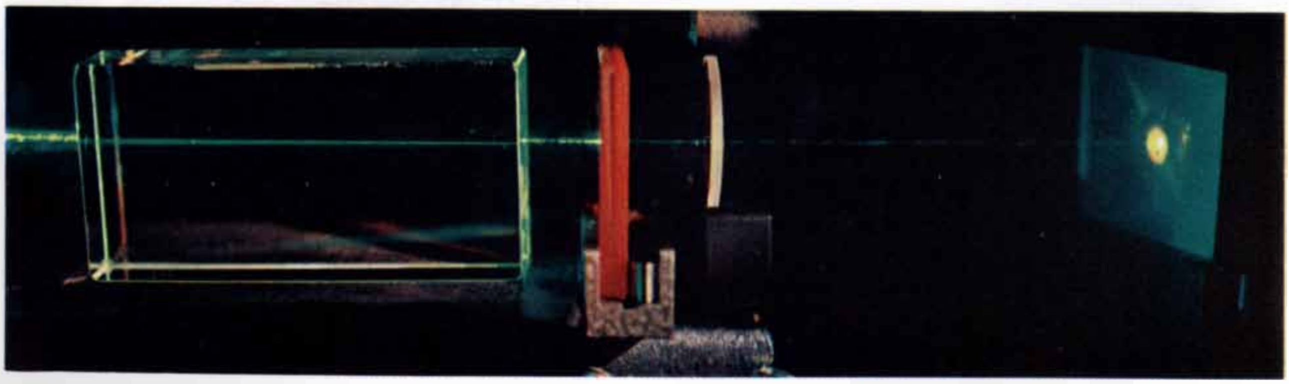
Now the tension between medicine and poetry increased. As his roommate later recalled, he talked only of being a poet and "began to affect a poetic style of dress 'a la Byron.... He also let his moustachios grow occasionally.' " His disaffection from medicine may have been heightened by assignment as an assistant to one William Lucas, Jr., "the butcher of the hospital," who (according to a more distinguished surgeon) made his colleagues " 'shudder from apprehension of his opening arteries or committing some other error.' " Yet Keats signed up for a second year, contemplating taking the examination for Member of the Royal College of Surgeons in early 1817. In May his first published poem appeared. Yet in July he took an examination for the License of the Society of Apothecaries and passed it, thereby qualifying for general practice (outside London; a London license exacted a larger fee).

In the fall of 1816, not quite 21, Keats was hard at work as a dresser and was attending advanced lectures on surgery. Then he wrote "On First Looking into Chapman's Homer" and showed it to Leigh Hunt, who published it, introduced him to Hunt's circle of poets and artists and wrote of Keats and Shelley as constituting "a new school of poetry." That praise, Keats's roommate wrote, "sealed his fate." The decision was reinforced, Gittings thinks, by Keats's conviction that he was "temperamentally unfit to perform a surgical operation." Whatever the decisive reasons, Keats did not sit for the Royal College examination in February, 1817. On March 3 he failed to renew his hospital appointment. "On that precise date was published a small grey-covered book, priced modestly at six shillings—*Poems* by John Keats."


The new Plymouth Sport Suburban has nine important features most wagons don't have.
That's part of the beauty of the new Sport Suburban. It has things most full-size wagons have. Like a three-way tailgate. Carpeting. Special storage bins. But it also has plenty of things most wagons *don't* have. Read the list below. Then, see your Chrysler-Plymouth dealer. We figure the more you compare wagons, the more you'll want a Sport Suburban. And that's a good thing for both of us.
1. More hip room than other full-size wagons. Front and rear.
2. Holds more 4′ x 8′ sheets of plywood than other full-size wagons (Even with the tailgate closed).
3. More cargo height than other wagons.
4. Standard built-in air deflector to help keep the rear window clean.
5. Rear-facing rear seat. You don't have to be a contortionist to get in.
6. Torsion-bar front suspension for a smooth ride and tight control.
7. Standard electronic ignition.
8. Step pads on both sides of the rear bumper.
9. Standard electronic voltage regulator. Helps increase battery life.
CHRYSLER
MOTORS CORPORATION
Plymouth Sport Suburban
Extra care in engineering . . . it makes a difference.
CHRYSLER
Plymouth



EXTREME INTENSITY of the ultrashort laser pulses used in the authors' experiments to explore the atomic and molecular properties of materials is demonstrated by the colorful effect captured in this photograph. A pulse of green light derived from a "mode-locked" solid-state laser was directed through a transparent piece of glass, distorting the refractive index of the glass in the path of the beam. The altered refractive index in turn reacted back on the light pulse to produce this circular pattern of spectral colors.



EXPERIMENTAL SETUP of the type employed to make the photograph at the top of the page was photographed in the laboratory of one of the authors (Alfano) at the City College of New York. Outside the glass block the laser beam was made visible with the aid of smoke. The reddish objects are filters. For this picture the image at right was formed on paper instead of photographic film.

Ultrafast Phenomena in Liquids and Solids

Intense laser pulses a trillionth of a second in duration are harnessed to probe the extremely rapid energy-transfer times characteristic of events in the submicroscopic world

by R. R. Alfano and S. L. Shapiro

At the submicroscopic level even the quietest, most peaceful scene in nature is in fact a riot of activity. Within every substance, no matter how passive its outward appearance, atoms and molecules are in continuous motion: rotating, vibrating, rocking back and forth, colliding with their neighbors—all in times measured in picoseconds (trillionths of a second). Until recently it has not been possible to observe such atomic and molecular motions on a time scale of less than about a nanosecond (a billionth of a second) because of the inherent limitations of the available experimental techniques. As a consequence nearly all our knowledge about these extremely rapid events has been derived somewhat indirectly from an analysis of the characteristic frequencies of the radiation they cause to be absorbed or emitted.

The recent development of the "mode-locked" laser, a solid-state device capable of generating extremely intense light pulses with a duration on the order of a picosecond, has in effect opened up this hitherto inaccessible world of ultrafast phenomena to direct exploration. The light pulses from such a laser can be so intense that they tear an atom or a molecule apart into its basic building blocks (protons, neutrons and electrons). The laser pulses can also alter the way an atom or a molecule behaves by disturbing its structure.

A dramatic example of the power of this new research tool is provided by the photograph at the top of the opposite page. The picture shows a novel effect observed when a sufficiently intense pulse of green light from a mode-locked laser is directed through a piece of transparent glass. The laser pulse significantly changes the refractive index of the glass, causing the altered refractive index in turn to react back on the light pulse in such a way as to generate a brilliant array of colors spanning the entire visible spectrum.

In this article we shall discuss techniques for using such ultrashort laser pulses to probe rapid energy-transfer processes in various materials. Historically this effort has evolved from two different directions. In one approach the experimenter studies the characteristic frequency spectrum of a material in order to derive information about its structure and its internal processes; in the other approach he subjects the material to some outside shock and then measures the material's characteristic "relaxation time," that is, the time it takes for the atoms or molecules of the material to return to their normal states.

Measuring Frequency Spectra

The frequency spectra of the light emitted and absorbed by materials have been the subject of investigation ever since man first noticed that objects have color. The first systematic scientific study of the interaction of light with matter was carried out by Isaac Newton, who in 1666 succeeded in splitting the white light of the sun into different colors with a prism. Newton reported these results in his first scientific paper, which appeared in 1672. His theory that "ordinary white light is a mixture of rays of every variety and colour" was vigorously attacked by such eminent contemporaries as Robert Hooke and Christiaan Huygens, causing Newton to consider giving up natural philosophy altogether. Newton also believed that light consisted of tiny corpuscles. When in 1803 Thomas Young demonstrated that light consisted of waves, he too was vehemently assailed, because it was felt that he was undermining the corpuscular theory of Newton.

In 1873 James Clerk Maxwell published his general wave theory of electromagnetism, explaining light propagation in terms of waves that could be of any frequency. Thus the colors observed by Newton correspond to frequencies that can be detected by molecules sensitive to visible light and interpreted in the eye-brain system. In 1905 Albert Einstein, trying to explain the ejection of electrons from metallic surfaces by light, introduced (with the aid of Max Planck's quantum principle) the concept that the energy in a light beam travels through space concentrated in the bundles called photons. By the end of the 19th century many materials had been shown to emit and absorb well-defined spectral frequencies. The origin of these mysterious discrete spectral frequencies was firmly established in the 1920's with the development of the theory of quantum mechanics, which united the corpuscular and wave theories of light.

According to quantum theory, atoms and molecules have distinct energy states corresponding to different configurations of the electrons around nuclei. Discrete spectral lines appear when a transition from one energy state to another occurs in the atomic or molecular system and the excess energy is emitted or absorbed in the form of a light wave. Spectral lines yield information about atomic and molecular structure by telling us where the allowed energy states lie and which transitions between these energy states are possible. Different frequency regions yield information on the various types of molecular motion that occur in matter.

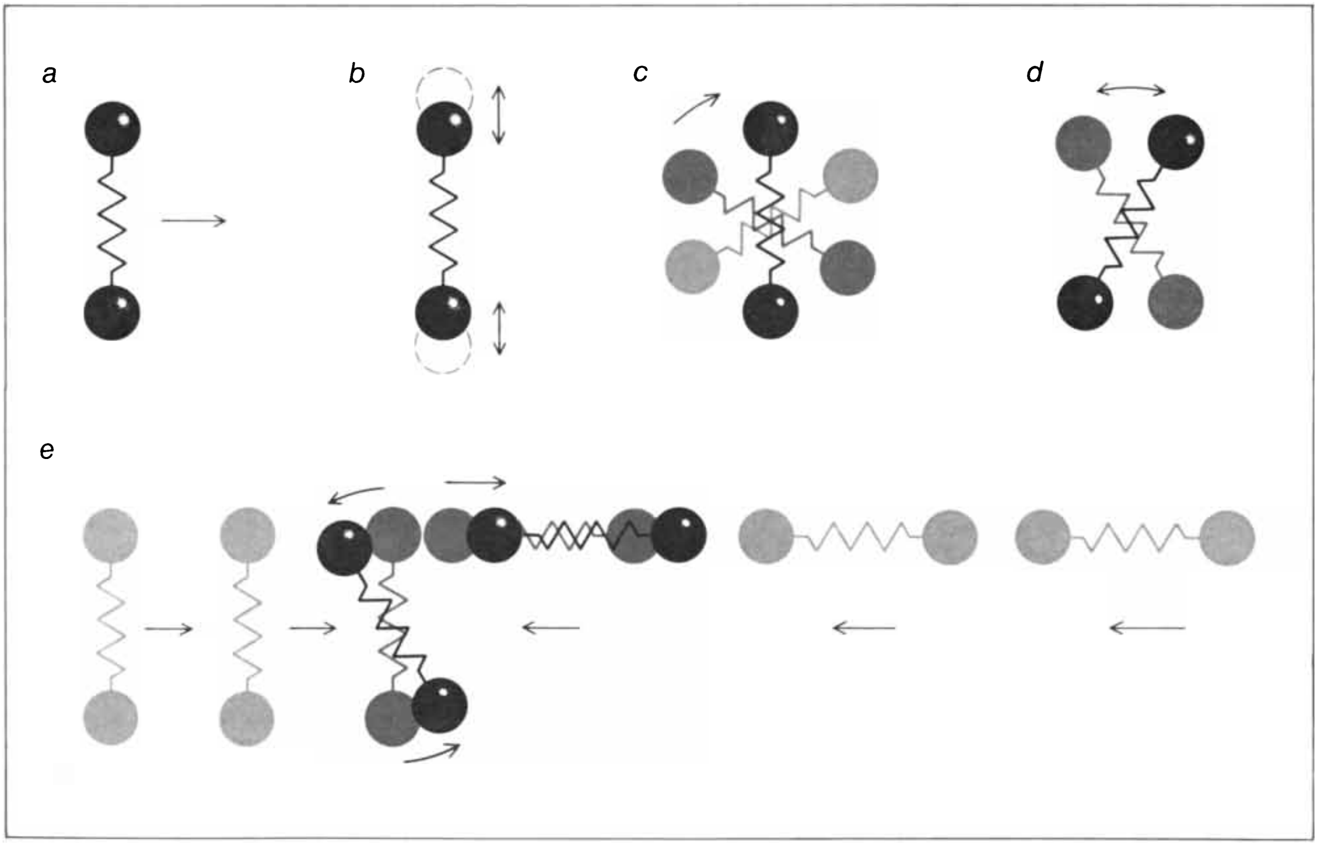
Additional information about the structure of materials from their spectra became accessible after an important effect was discovered experimentally by C. V. Raman in 1928. This phenomenon, now commonly called the Raman effect, occurs whenever monochromatic light passes through any transparent material. The emerging light is somewhat less monochromatic than the incident light; new frequencies are added to the original beam as "side bands" but are so weak that a sensitive apparatus is required for their detection. Some of the new frequencies are discrete and well separated from the original frequency and are attributable to the vibrational motion of atoms within molecules; these vibrational frequencies were the ones first observed by Raman. Vibrational frequencies are low compared with visible-light frequencies because they arise from the motion of the massive nuclei, whereas the visible frequencies arise from the motion of the much lighter electrons.

In the 1930's new, lower Raman frequencies were discovered. These frequencies are associated with rotational motions of molecules and with acoustic waves in liquids and solids. A subsequent discovery revealed a frequency band that clusters around the original light frequency and is slightly broader than the spectrum of the monochromatic light; this spectrum is caused by the disordered thermal motion of the individual atoms or molecules. Within the past decade light-scattering research has had a rebirth because of the invention of the laser and of new tools for examining the frequency spectrum. As a result the frequency-analysis approach has furnished a wealth of new information on molecular and atomic behavior. It is possible to obtain an estimate of the lifetime of a particular energy transition using the frequency-analysis approach (by computing the reciprocal of the frequency bandwidth of the transition); this indirect method of investigation is rarely employed, however, because it often leads to large errors.

Measuring Rapid Relaxation

Rapid molecular-energy-transfer relaxation times were first measured with mechanical devices, because advanced electronic techniques did not yet exist. In 1859 Edmond Becquerel devised a spinning-disk technique capable of resolving times as short as a ten-thousandth of a second [*see top illustration on opposite page*]. The property most important for measuring a short time interval with any mechanical device is either a rapid rotational velocity or a rapid translational velocity. In Becquerel's scheme two slotted disks are mounted on a shaft with the luminescent sample placed between the two disks. As the disks spin, light from a light source passes through the first disk, exciting the sample and causing it to luminesce. The luminescence is observed by the experimenter behind the second slot at a time determined in part by the angle formed by the slots in the two disks and in part by the angular velocity. By varying the velocity of rotation or the angular displacement of the second disk one can measure the duration of the luminescence.

In 1905 Robert W. Wood devised a technique similar to Becquerel's that demonstrated a lifetime as short as five hundred-thousandths of a second for the luminescence of zinc sulfide bombarded with radiation emitted by radium [*see middle illustration on opposite page*]. Wood coated the rim of a wooden disk with zinc sulfide and placed a speck



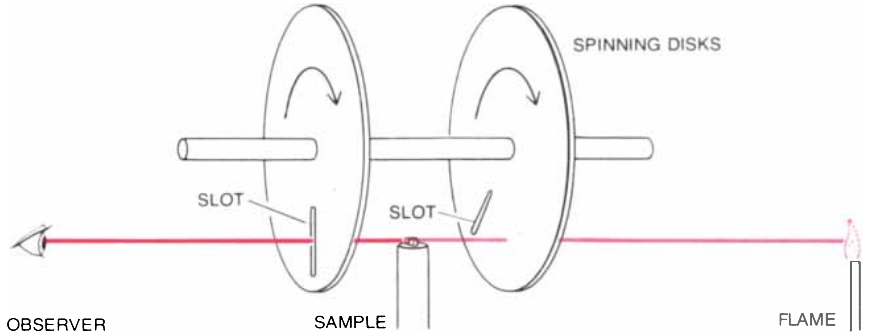
VARIETIES OF MOLECULAR MOTION are represented schematically by the spring-coupled two-atom systems in this drawing. Sample molecular motions shown include a translation (*a*), a vibration (*b*), a rotation (*c*), a libration (*d*) and a collision (*e*).

of radium above the disk on the point of a needle. The disk was mounted on the shaft of an electric motor. He viewed the apparatus in a dark room and found that the luminescence came from a sharp point when the disk was stationary or rotating slowly but was drawn out into a streak when the disk was rotating rapidly. The length of the streak divided by the velocity of the rim of the wheel then yielded the lifetime of the luminescence. In 1921 Wood, using a spark to excite the sample on a disk, was able to measure times as short as two millionths of a second.

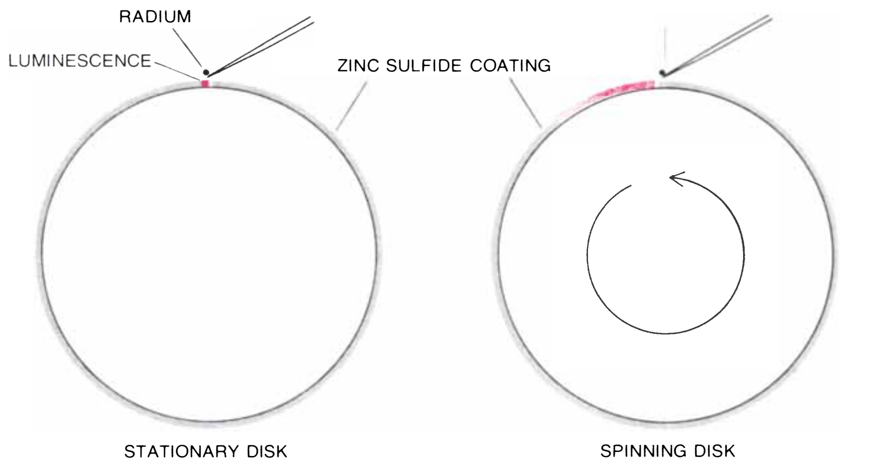
Wood and his colleague Charles E. Mendenhall later found that liquids could be investigated by another mechanical technique. They used a high-pressure pump to force a liquid containing a fluorescent dye through a nozzle at high velocity [*see bottom illustration at right*]. A tiny hole was made at the end of the nozzle for sunlight to enter and excite the dye. The fluorescent tracks they observed were less than a tenth of a millimeter long; from this observation they concluded that the fluorescent lifetime for the dye in water was less than four ten-millionths of a second.

Another ingenious mechanical device for measuring rapid relaxation times was the streak camera invented by S. I. Vavilov and V. L. Levshin in 1926 [*see top illustration on next page*]. In this scheme a spark of short duration excites a sample; luminescence from the sample then passes through a slit, is collected by a lens, bounces off a mirror and is imaged onto a photographic plate. If the mirror is rotated at high velocity, the luminescence is recorded as a streak on the photographic plate, indicating by its length the luminescent lifetime of the sample.

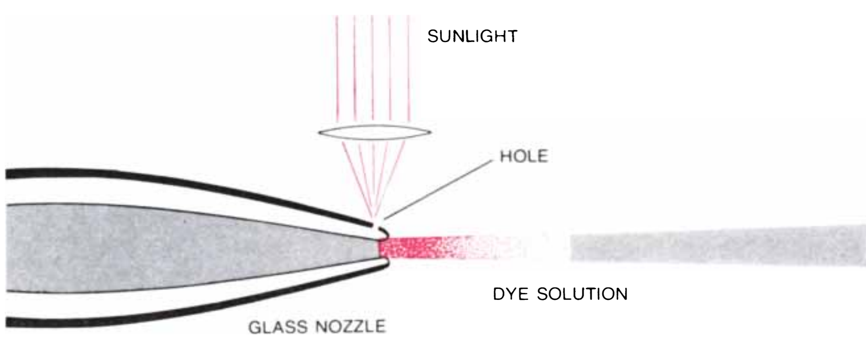
About 1925 electronic techniques began to replace mechanical devices for measuring rapid relaxation times. The first of these techniques made use of a shutter that was based on electro-optic effects discovered by John Kerr in 1875. The thought that led to Kerr's investigation was in his own words "that if a transparent and optically isotropic insulator were subjected properly to intense electrostatic force, it should act no longer as an isotropic body upon light sent through it." (An isotropic body is a body whose properties are the same in all directions.) Kerr found that when a direct-current electric field is applied across a piece of glass, the index of refraction of the glass is different in the direction of the field. In technical terms, a birefringence is said to be induced.



PHOSPHOROSCOPE, an ingenious mechanical device invented by Edmond Becquerel in 1859, was capable of resolving luminescent lifetimes as short as a ten-thousandth of a second. The sample was placed between two spinning disks. Light from a flame source passed through a slot in the first disk, exciting the sample and causing it to luminesce. The luminescence observed through the slot in the second spinning disk enabled the experimenter to determine the luminescent lifetime of the sample. In Becquerel's original apparatus a complicated set of gears was attached to the disks and the device was cranked by hand.



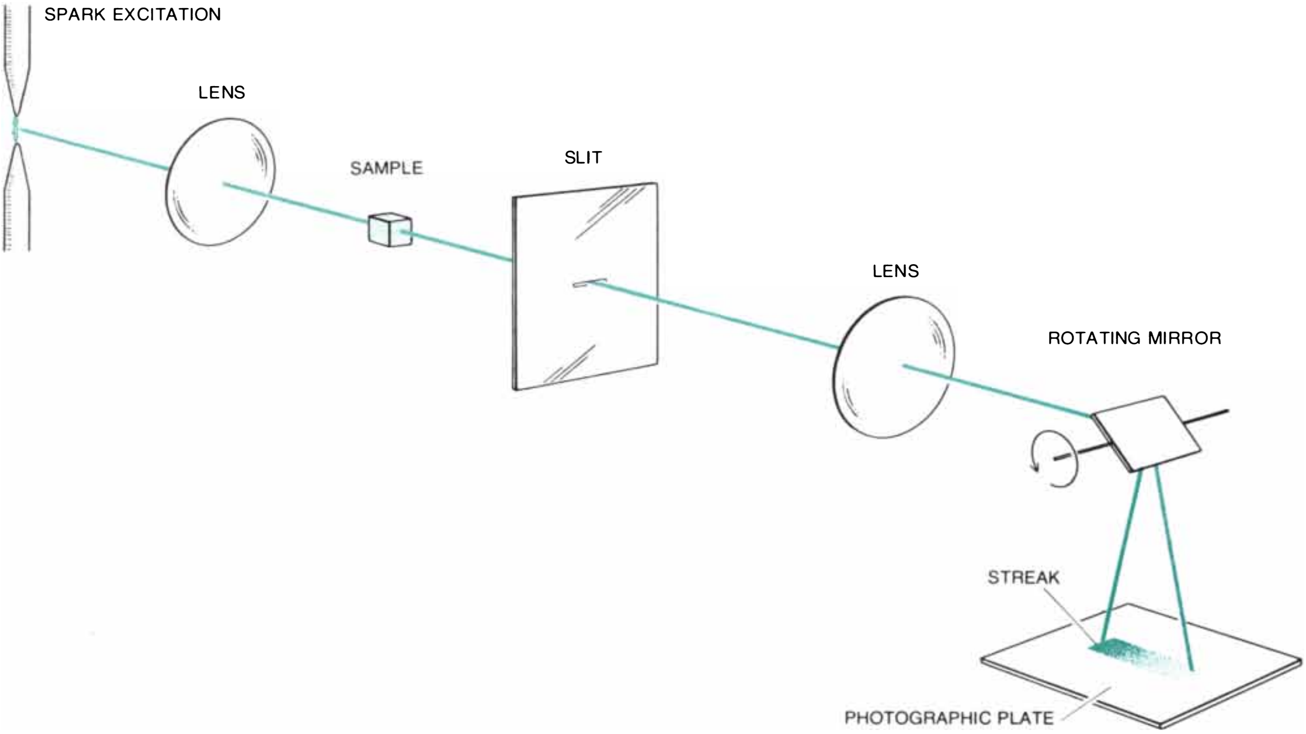
SIMILAR TECHNIQUE, devised by Robert W. Wood in 1905, was employed to demonstrate that the luminescence stimulated from zinc sulfide by radiation from radium has a lifetime of only five hundred-thousandths of a second. Wood coated the rim of a wooden disk with zinc sulfide and caused the coating to luminesce by placing a speck of radium above the disk on the point of a needle. When the disk was stationary or was rotated slowly (*left*), the luminescence came from a small point; when the disk was rotated rapidly (*right*), the luminescence was drawn out into a streak. The lifetime of the luminescence was obtained by dividing the length of the streak by the velocity of the rim of the disk.



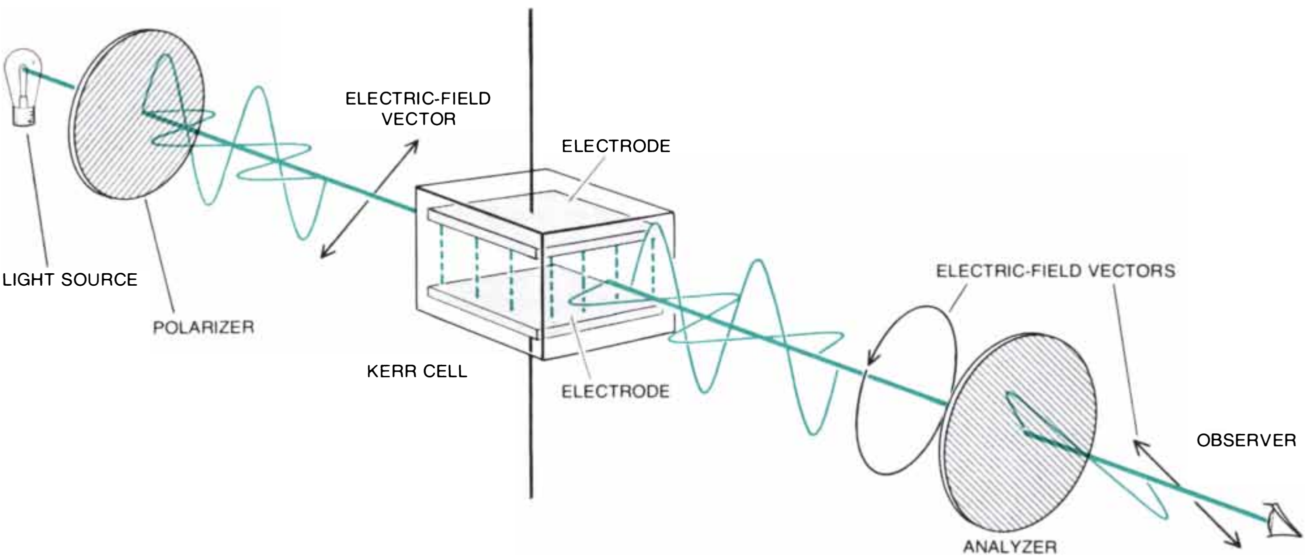
FLUORESCEIN DYE was found by Wood and his colleague Charles E. Mendenhall to have a fluorescent lifetime of less than four ten-millionths of a second with the aid of the mechanical technique depicted here. The solution of fluorescein in water was forced through a glass nozzle at high velocity by means of a high-pressure pump. Fluorescence was excited in the dye solution by sunlight entering through a small hole near the end of the nozzle. The fluorescent lifetime was measured by dividing the length of the fluorescent streak (typically less than a tenth of a centimeter long) by the velocity of the liquid.

According to Kerr, the birefringence is caused by dielectric particles that become polarized and tend to align themselves along the direction of the electric field. The particles subsequently attract one another, which leads to a compression and to an increase in the refractive index along the lines of electric force. Based on principles discovered by Kerr, a direct-current optical shutter, or Kerr gate, was built by Henri Abraham and J. Lemoine in 1899. Although these two investigators succeeded in demonstrating that Kerr's device was capable of measuring rapid time intervals, a quarter-century was to pass before it was used systematically for that purpose.

The direct-current Kerr gate consists of two metal electrodes immersed in a cell of liquid situated between crossed polarizer and analyzer filters [*see bot-*



MECHANICAL STREAK CAMERA, invented by S. I. Vavilov and V. L. Levshin in 1926, was capable of measuring even shorter luminescent lifetimes. Light from a spark discharge was focused on a sample, causing it to luminesce. The luminescence from the sample then passed through a slit, was collected by a lens, bounced off a rotating mirror and was imaged on a photographic plate. The rapid rotation of the mirror "smeared" the luminescence across the plate, forming a streak whose length indicated the luminescent lifetime of the sample. Electronic techniques subsequently replaced such mechanical devices for measuring rapid molecular-relaxation times.

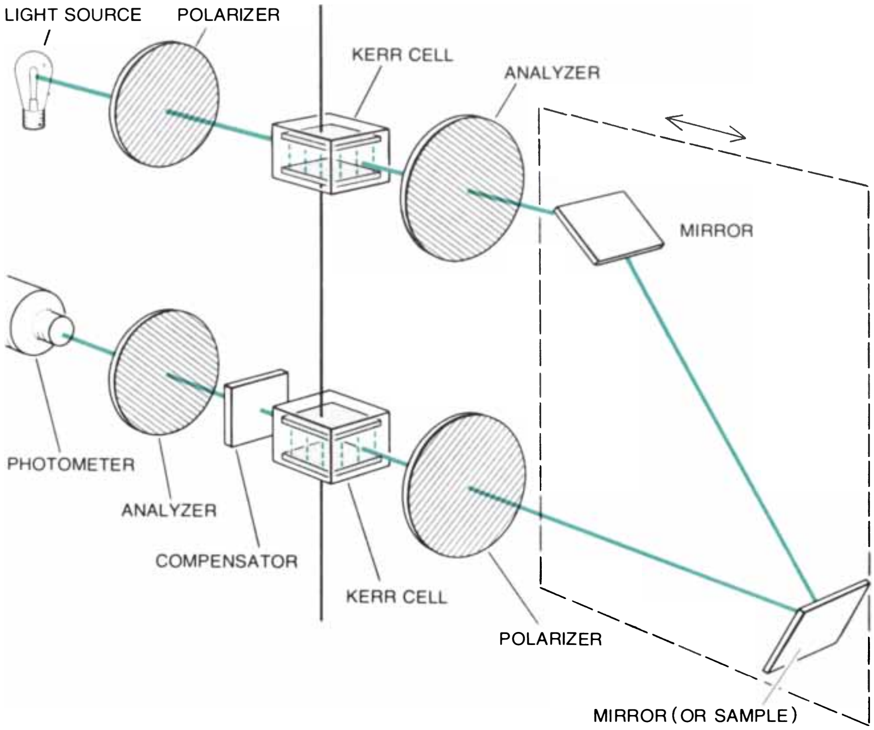


DIRECT-CURRENT KERR GATE, an early electronic device for measuring rapid molecular-relaxation times, was based on electro-optic effects discovered by John Kerr in 1875. The device consisted essentially of two electrodes immersed in a liquid cell situated between crossed polarizer and analyzer filters oriented at +45 degrees and —45 degrees with respect to the axis of the applied electric field. When no voltage is applied to the Kerr cell, none of the incident light, which is linearly polarized by the polarizer, passes through the analyzer. When a high voltage is applied, however, the light becomes elliptically polarized on passing through the Kerr cell, with the result that the polarization component of the light parallel to the axis of the analyzer is transmitted.

tom illustration on opposite page]. The polarizer is oriented at an angle of +45 degrees with respect to an axis perpendicular to the metal plates and the analyzer at −45 degrees. A light beam passing through the polarizer is therefore polarized at an angle of 45 degrees. The polarization of the light remains unchanged when the beam passes through the cell because there is no preferable new direction in the isotropic liquid. The gate remains closed (that is, no light is transmitted through the system), because the light has no component in the analyzer direction if the analyzer is oriented at 90 degrees with respect to the direction of polarization. When a strong direct-current electric field is applied across the liquid, however, the gate transmits light, because the light beam becomes elliptically polarized on passing through the liquid cell; as a result the component of light polarized along the analyzer axis is transmitted. The elliptical polarization arises from the fact that the light wave can be split into two components; the one that is polarized along the electric-field axis slows down or speeds up with respect to the one that is polarized at right angles to the electric-field axis, because of the change in refractive index due to the field.

If an electric-field pulse of short duration is applied to the electrodes of the liquid cell, only the portion of light coincident with the electric-field pulse passes through the analyzer of the gate. The resolution time, or fastest shutter time, of this type of gate is determined by the shortest duration provided by an electrical voltage pulser: about a nanosecond. Even though the time resolution of the shutter is about a nanosecond, however, experimenters through the early 1940's had to resort to an additional bit of craftiness to measure short lifetimes with it. Electronic devices such as the high-speed oscilloscope and the fast photomultiplier were unavailable at that time, and rapid relaxation times had to be measured with complicated shutter schemes.

One such scheme, an improved version of a design devised originally by Enrique Gaviola in 1926, was used by W. Szymanowski in 1935 to measure lifetimes as short as a nanosecond with an accuracy of a few percent [*see illustration on this page*]. The apparatus works as follows. Two Kerr cells are electrically connected to some voltage source, which is varied synchronously and sinusoidally at a characteristic frequency. Light from a light source passes through the first Kerr cell and is modulated at this frequency. The light is reflected from a system of mirrors and arrives either in phase or out of phase at the second Kerr cell, depending on the length of the light path between the cells. The phase difference can be adjusted by adding a compensator in the second Kerr cell.

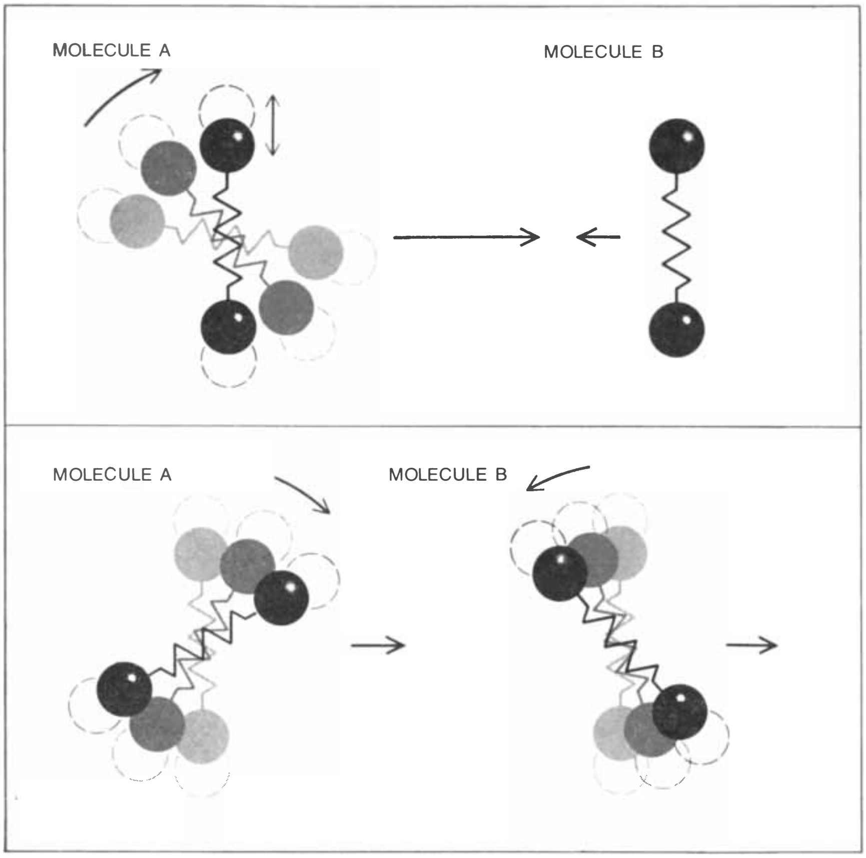


COMPLICATED SHUTTER SCHEME based on the principle of the direct-current Kerr gate was designed by Enrique Gaviola in 1926; an improved version of this scheme was later used to measure fluorescent lifetimes as short as a nanosecond (a billionth of a second) with an accuracy of a few percent. The apparatus consists of two Kerr cells hooked up in series with the voltage across them varying synchronously and sinusoidally at a characteristic frequency. Light from a light source is modulated electronically in the first Kerr cell. After being reflected by a system of mirrors the light passes through the second Kerr cell only if it arrives in phase with the modulation. If a fluorescent sample is placed at the position of one of the mirrors, the induced fluorescence from the sample will be delayed with respect to the frequency of the second Kerr cell by an amount that depends on the relaxation lifetime of the fluorescence. Phase differences can be adjusted by moving the part of the apparatus within the broken line, by varying the ellipticity of the polarization with a compensator or by changing the frequency. A plot of the intensity of the light received as a function of frequency enables one to derive fluorescent lifetime of the sample.

If one of the mirrors is replaced by a fluorescent substance, the fluorescent light will no longer arrive in phase at the second Kerr cell because of the time delay attributable to the relaxation lifetime of the fluorescence. The light that passes through the second Kerr cell is detected by a photometer and depends on that lifetime. By varying the frequency at rates comparable to the lifetime one can obtain a curve of the intensity at the photometer as a function of frequency, and from this curve one can derive the fluorescent lifetime. This "phase shift" technique was developed because the photometers could measure only the average intensity of the light and were not yet fast enough to respond to rapid variations in light intensity.

The first use of high-speed photomultipliers capable of responding to rapid lifetimes was made as recently as 1948. At that time nuclear physicists were very interested in determining the lifetime of certain phosphors used in scintillation counters. The method by which a scintillation counter monitors the decay of nuclei is based on the fact that particles released by a nuclear event can cause ionization in liquids or solids with the consequent emission of light. The light waves can then be detected by high-speed coincidence-counting circuits, and the decay time of the nuclear event can be inferred from a measurement of the decay of the light intensity. Since some nuclear decays are of exceedingly short duration, the characteristic lifetime of the scintillator must be rapid indeed.

In 1948 George B. Collins measured fluorescent lifetimes by allowing the emission from a sample of radium to impinge on a phosphorescent substance and then detecting the light emission from the phosphor as a function of time



ENERGY IS EXCHANGED between different types of motion in a collision involving molecules. Before the collision depicted in this drawing, for example, molecule *A* is translating, rotating and vibrating, whereas molecule *B* is just translating (*top*). After the collision the distribution of translational, rotational and vibrational modes is altered (*bottom*).

with a high-speed photomultiplier connected to an oscilloscope. He found that the passage of a single million-electron-volt electron produced enough light in the phosphor for observing both the rapid rise of the fluorescence determined by the minimum response time of the photomultiplier circuit and the decay of the fluorescence, which yielded the fluorescent lifetime. Later, bursts of X rays were used to excite scintillating samples for the measurement of lifetimes in conjunction with photomultipliers and oscilloscopes.

Using an ionized particle is probably not the most accurate way of measuring fluorescent lifetimes, however, because the lifetime depends on the means of excitation. Excitation with a gamma ray, for example, may yield a different decay time than excitation with a neutron or a beta ray. This difference in decay time enables the experimenter to discriminate between emission of different particles but would not necessarily determine the true fluorescent lifetime of the material.

A more conventional scheme for measuring fluorescent lifetimes, not subject to such a drawback, is to excite a sample with a pulse of ultraviolet radiation obtained from a hydrogen flashtube and to detect the fluorescence with a photomultiplier connected to an oscilloscope. Such a technique has been used by I. Berlman and his co-workers at the Argonne National Laboratory to measure fluorescent lifetimes in a variety of materials, with the times measured being as short as a nanosecond.

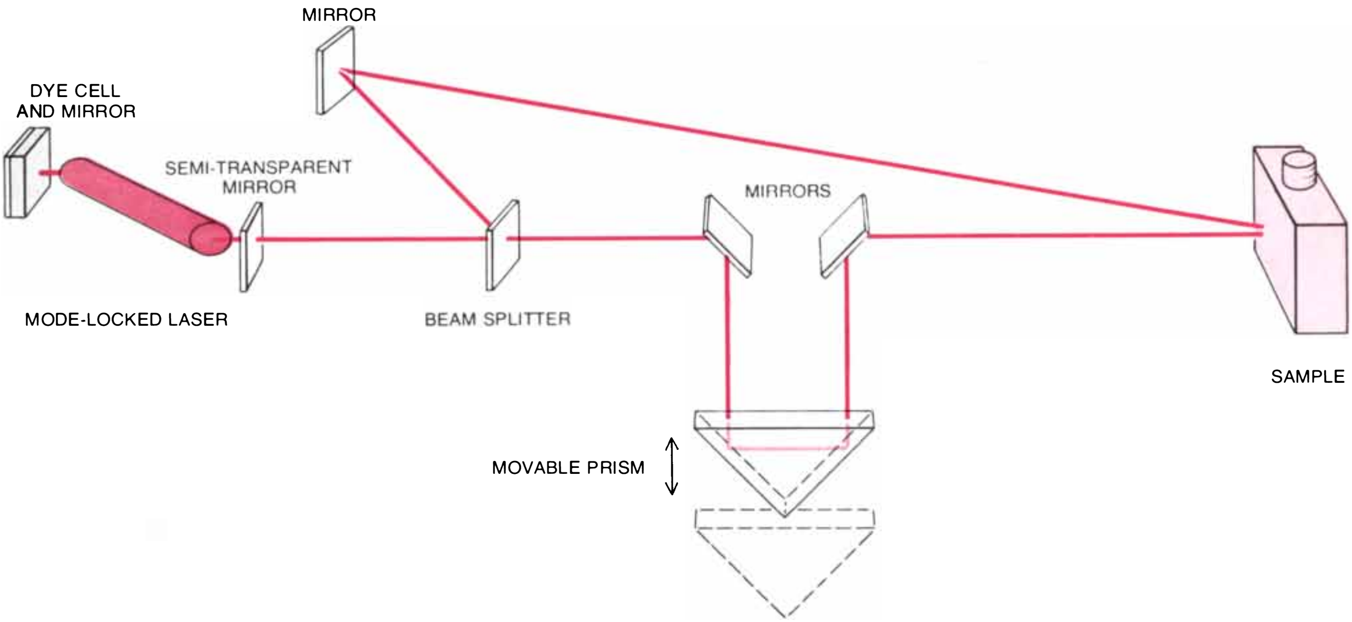
Harnessing the Picosecond Laser

With the advent of ultrashort laser pulses one can now study, in addition to rapid fluorescent processes, the dynamics of energy transformation between the different types of motion that occur in solids, liquids and gases. For example, a molecule moving in a liquid with a particular velocity and vibrating and rotating at a well-defined rate can collide with another molecule at rest and transfer some or all of its kinetic, vibrational and rotational energy. In a real material one has a collection of colliding molecules that interact, exchange energy and come to equilibrium between the various types of motion. If one particular vibration or rotation is selectively excited in a molecule by a light beam, this motion will decay by transforming its energy into other vibrations, rotations and molecular motion by means of collisions in an interval of about a picosecond in a liquid or a solid.

In crystals, where atoms are interconnected in a regular periodic array, the atoms are in collective vibration. Depending on the complexity of the crystal and the constituent molecules, the lattice can have many vibrational modes. The frequencies of these vibrations extend from the audible to the optical, and their modes are called acoustical and optical phonons. The energy of lattice vibrations is quantized in analogy with the photon, which is the quantum of energy of a light wave. The complexity of the forces between the molecules results in a coupling between these different vibrations, or phonons, so that a single vibration excited by a laser pulse can readily decay into other vibrational motion or heat.

To measure fast decay times and to determine routes of energy transfer between molecules, one needs a clock with a time-scale index of a picosecond. Just as a kitchen clock cannot accurately measure events that take place in less than a second (such as the explosion of a firecracker or the rapid blink of an eye), so a fast-decay-time clock cannot perform its function unless it has a time-measuring capability faster than the event to be measured. Such a clock with the necessary time-scale index is available; it uses the light pulse emitted from a mode-locked laser. The duration of the light pulse (about a picosecond) denotes the smallest divisional marking on the clock's time scale, corresponding to the markings for seconds on an ordinary clock.

In the laser clock the constant speed of light traveling over a known distance provides the basis of the time-measuring mechanism. It takes a light beam one nanosecond to travel a foot in space. An interval of time on the clock is determined by the separation between two light pulses, which can be varied by arbitrarily spacing the two pulses in a manner analogous to the setting of the hour and minute hands on an ordinary clock. As an example, consider two light pulses traveling toward a target on the same line and separated by a distance of a foot. The second pulse will obviously arrive a nanosecond after the first. The clock is calibrated by knowing the coincident position of the two light pulses. Then the two pulses can be separated by any known distance. Coincidence is accomplished by a process that is accentuated when the two light pulses



TWO SYNCHRONOUS ULTRASHORT LIGHT PULSES suitable for calibrating a fast-decay-time clock to an accuracy on the order of a picosecond (a trillionth of a second) can be produced by taking the emission from a mode-locked laser and reflecting part of the beam with a semitransparent mirror while allowing the rest of the beam to pass through the mirror. The separation of the two pulses is then adjusted by means of a scheme that directs the pulses along different paths to a target. In this scheme a movable prism delays one pulse with respect to the other by changing the length of one optical path between the source and the target.

intersect; any "nonlinear" optical effect (such as "frequency-doubling" in a suitable crystal) will do. Various nonlinear processes used in picosecond-pulse clocks will be discussed below.

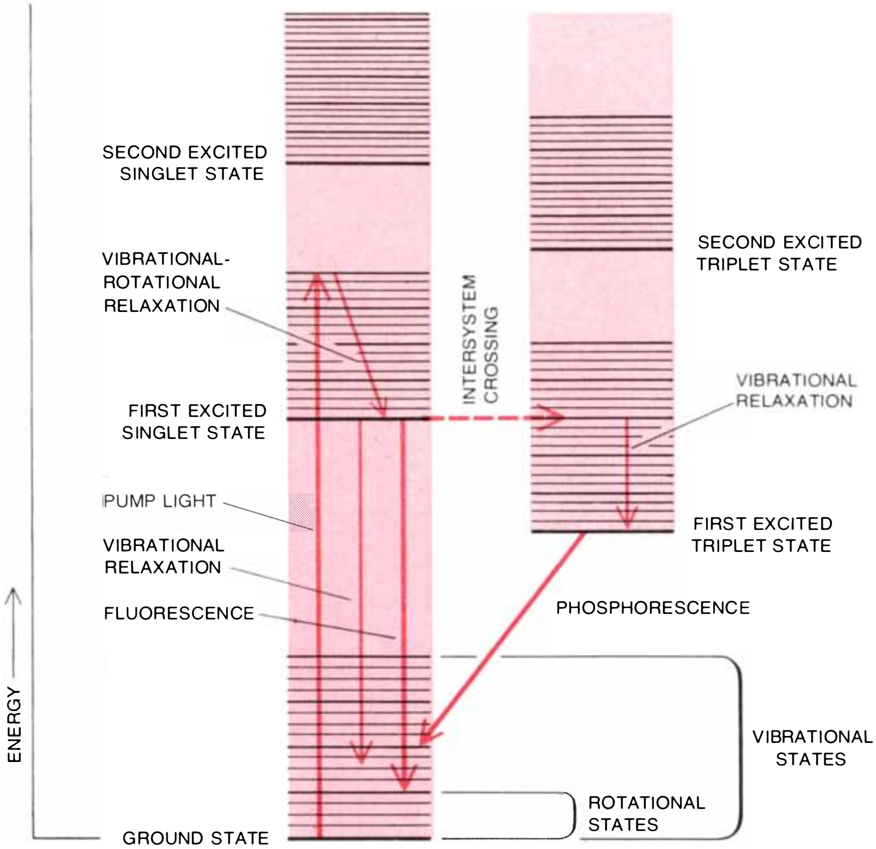
Two synchronous ultrashort light pulses can be produced by taking the emission from a mode-locked laser, reflecting part of the beam with a semitransparent mirror and allowing the rest of the beam to pass through the mirror and then adjusting the separation of the two pulses by means of a scheme that directs the pulses along separate paths to a target [*see illustration above*]. To probe the time evolution within the microscopic world (the creation and annihilation of vibrations, rotations and so on) the first picosecond light pulse initiates the physical process to be studied and the second picosecond light pulse probes the time evolution of the process at different delay times.

For example, a laser pulse excites the system of molecules into a vibrational state, a rotational state or an electronic state by inducing a change of state on absorption of the light pulse. If the molecules are excited with a pulse longer than the decay time one is interested in, the molecules would be excited and would decay within the pulse width of the laser pulse, so that the measurement of time evolution would be difficult or impossible. For such a measurement one needs an ultrashort pulse that will excite the particular type of molecular motion abruptly, leaving the molecules to decay freely under the influence of the forces within the microscopic world. The molecular decay can then be observed either by probing the number of remaining excited molecules with the second pulse or, if the excited molecule emits light, by observing the light decay with an ultrafast camera that is opened with the second pulse.

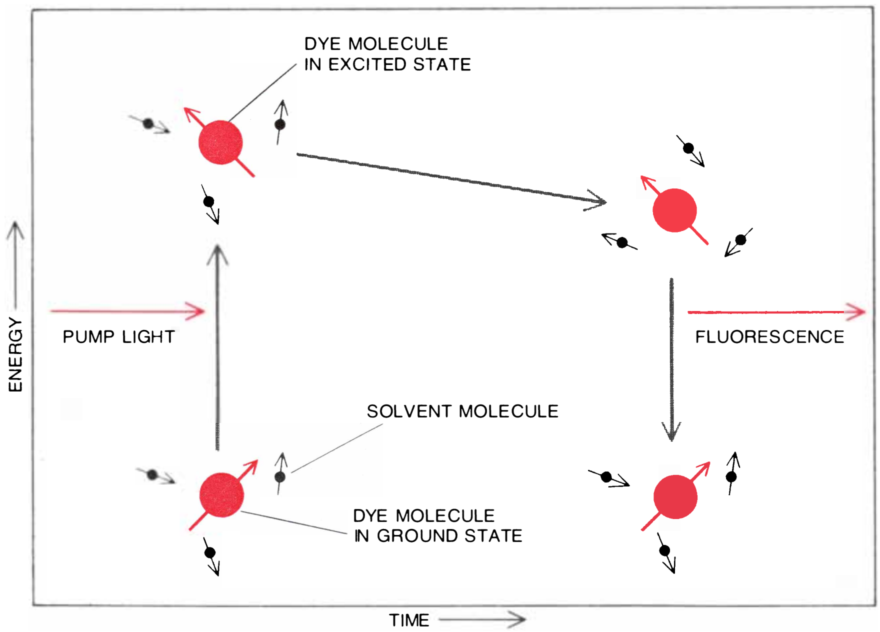
Let us now discuss how picosecond laser pulses are generated by a mode-locked laser, how the duration of these pulses is determined and how the pulses are used to measure ultrafast processes in liquids and solids.

The active medium of the mode-locked lasers used to generate high-power ultrashort light pulses must have a wide fluorescent bandwidth in order to

ULTRASHORT LIGHT PULSES are measured by means of the two-photon fluorescence technique developed in 1967 at Bell Laboratories. The technique is based on the special optical properties of an organic solution in which two photons of light must be absorbed to excite a molecule to a higher energy state from which it can decay spontaneously back to the ground, or lowest-energy, state, emitting fluorescent light in the process. Bright spots are produced in such a medium where pulses collide with each other, because of the intensity at the collision point. The fluorescence from the liquid cell can be photographed, as demonstrated in this picture made by one of the authors (Shapiro). In this case the collisions occur because the beam, which enters from the left, is reflected back on itself by a mirror at the right. The length of the bright spots divided by the velocity of light in the medium yields a measurement of the duration of a single pulse. Laser pulses as short as .3 picosecond have been measured with this technique.



ENERGY LEVELS of a typical organic-dye molecule are shown in this highly schematic diagram. A light pulse can "pump" such a molecule from its ground "singlet" state to a sublevel of the first excited singlet state, from which the molecule decays rapidly by nonradiative processes (vibrational or rotational relaxation) to the lowest sublevel of the first singlet state. The molecule can then fluoresce, returning directly to the ground state, or cross over to a system of "triplet" energy levels, from which it can be deactivated by emitting phosphorescence in the process of returning indirectly to the ground singlet state.



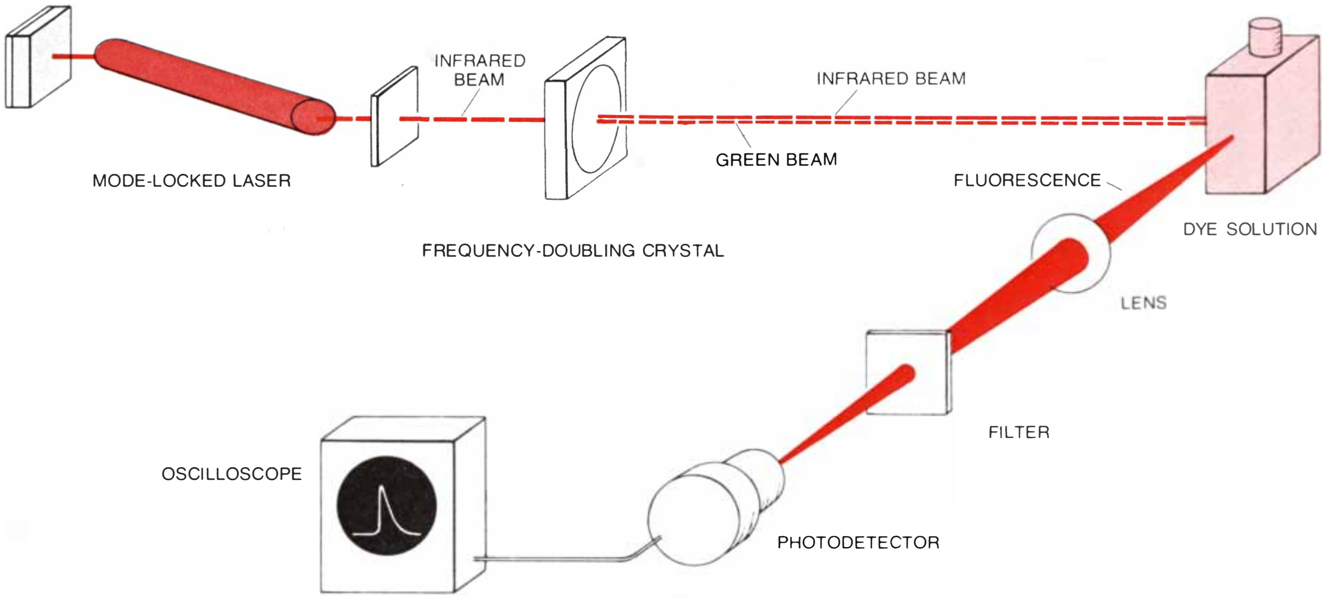
SOLVENT-SOLUTE INTERACTION involving dye molecules dissolved in a suitable solvent can be monitored by measuring changes in the intensity and wavelength of the fluorescence from the excited dye molecules. In this case the solvent molecules (*black*) rearrange themselves spatially in response to a new, excited configuration of a dye molecule (*color*). The fluorescent wavelength becomes time-dependent because the energy levels of the dye molecule shift in time in response to the dynamic molecular interaction.

produce picosecond pulses. The laser medium is contained between two mirrors, which form an optical cavity as in an ordinary laser. This cavity restricts the oscillating wavelengths within the broad bandwidth to just those wavelengths that satisfy nondestructive interference conditions; in other words, an integral number of wavelengths must equal the round-trip optical length of the cavity. The laser oscillates only at discrete frequencies, corresponding to those wavelengths called cavity modes. These closely spaced modes are amplified by stimulated emission, and a typical output spectrum can consist of hundreds or even thousands of them.

To obtain intense short-pulse emission one must establish a definite phase relation between the modes with a modulating element placed in the cavity. This element locks the modes in phase and is the key to the mode-locked laser. A favorite element for mode-locked ruby lasers and neodymium-doped glass lasers is a bleachable dye whose absorption at the laser frequency decreases with light intensity. In a neodymium-glass laser many modes probably start oscillating simultaneously with random phases. When laser action begins over a wide bandwidth, there are fluctuations that have a duration equal to the inverse of the initial bandwidth; the laser has a head start on the production of short pulses because short fluctuations are already present. The dye attenuates the most intense of these fluctuations or pulses less than it does any other pulse, so that after hundreds of passes back and forth through the laser cavity, nonlinear-absorption discrimination results in one prevailing intense pulse.

The dye must be able to return to its ground state much more rapidly than the round-trip time of the cavity so that it can perform its function again and again. Because of the nonlinear characteristics of the dye, the peak of a pulse is attenuated less than its wings, resulting in pulse-sharpening. This pulse-sharpening proceeds until the maximum number of modes that can possibly be coupled together are in phase; then the pulse duration is just equal to the inverse bandwidth of this maximum number of modes, and it can become no shorter.

Pulse-sharpening is equivalent to adding side bands in phase in the frequency domain; the more side bands that are added, the shorter the pulse duration is. Side bands originate because any wave circulating in the laser cavity is amplitude-modulated at the round-trip time by an absorption loss at a frequency that



ULTRASHORT FLUORESCENCE-DECAY TIMES can be measured with this system, developed by Michael E. Mack of the United Aircraft Corporation. Picosecond light pulses from a mode-locked laser excite the sample; the resulting fluorescence is then detected by a fast photodiode and displayed on a traveling-wave oscilloscope. In a typical experiment a neodymium-glass laser is used to generate short, powerful pulses at the infrared wavelength of 10,600 angstroms (*long dashes*); these pulses are then converted by the nonlinear optical process of second-harmonic generation (frequency-doubling) inside a potassium-dihydrogen-phosphate crystal to the green wavelength of 5,300 angstroms (*short dashes*), a more appropriate wavelength for pumping most materials.

is the inverse of the round-trip time. Therefore any frequency propagating in the cavity develops side bands in phase, which in turn develop side bands in phase, the process continuing until all the modes in the cavity have a well-defined phase relation with one another.

The secret of the high power obtained from the mode-locked laser is that when a certain number of modes oscillate together in phase, the peak power is that number times the average power when they oscillate randomly. Thus if out of a large number of modes oscillating with random phases 1,000 oscillate in phase, the pulse can be 1,000 times as intense. In this way an originally messy waveform can be turned into a well-behaved, intense picosecond pulse circulating in the laser cavity. The transmission through one of the cavity mirrors provides the output for experiments on rapid relaxation times.

One method of measuring the duration of ultrashort pulses utilizes the phenomenon of two-photon fluorescence. In this approach, which was first employed by one of us (Shapiro) at Bell Laboratories in 1967 at the suggestion of J. A. Giordmaine, one first chooses an organic-dye solution that cannot absorb a single photon of light (because there are no energy levels at the single-photon frequency) but can absorb two photons of light (because there are energy levels at twice the photon frequency). Since it takes two photons to excite a molecule in the solution, the amount of absorption depends on the intensity squared of the incident radiation. After being excited to a higher energy level the molecules can return to the ground state by emitting fluorescent light, the amount of fluorescence being proportional to the amount of absorption.

When two light pulses collide in such a medium, the intensity increases significantly at the crossing point, and the absorption and the fluorescence at that point are much higher. When the fluorescence produced by the colliding pulses is photographed, a bright spot appears at the point of collision, and the length of the bright spot divided by the velocity of light in the medium gives a measure of pulse duration. With this technique pulses as short as .3 picosecond have been photographed, and the resolution of the technique should in principle be as good as .01 picosecond, or a hundred-trillionth of a second [*see bottom illustration on page 49*].

The light pulses emitted from mode-locked lasers are an ideal tool for studying the factors governing the transitions between molecular energy levels. The high energy and power of the laser pulses are suitable for pumping large numbers of molecules to excited states, and the short duration of the pulses is suitable for studying molecular transitions by probing the time behavior of light emitted from excited states. Until recently light emitted from optically excited molecules for a duration of about a ten-millionth of a second was called fluorescence, whereas light emission continuing thereafter was usually called phosphorescence. Today the definition is more technical, fluorescence corresponding to "spin-allowed" electric dipole transitions and phosphorescence to "spin-disallowed" transitions. Fluorescence and phosphorescence can occur only after molecules are excited by another energy source and can be differentiated from the exciting source by shifts in wavelength, time behavior, polarization properties and angular distribution.

The new techniques enable one to examine the buildup and decay of emitted light. The initial stages of fluorescence are sensitive to the energy redistribution within a molecule and to its environment. Processes where excited molecules change energy level without emitting light, called nonradiative transitions, can also be studied. To understand the mechanisms leading to fluorescence, let us follow the routes of deactivation of a typical organic molecule through its energy levels after it has been excited by a light pulse [*see top illustration on opposite page*].

The electronic level corresponding to the ground, or lowest, "singlet" energy state of the molecule is characterized by an electronic configuration where the paired electrons in the outermost orbits spin in opposite directions. "Triplet" states arise when the electrons spin together in parallel in the same direction, the nomenclature originating with the fact that on close examination triplet states consist of three closely spaced energy levels. Extra sublevels in the diagram above singlet and triplet levels denote vibrational and rotational degrees of freedom of the molecule associated with that electronic state.

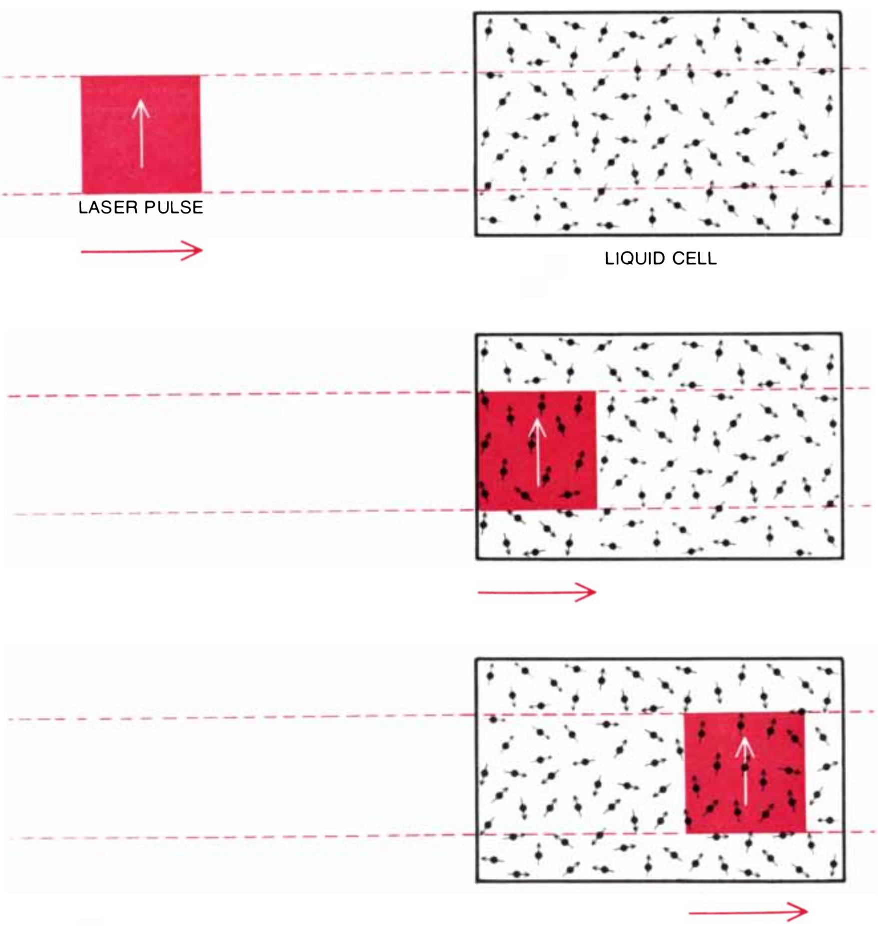
The level corresponding to the first excited electronic singlet state can also have vibrational and rotational energy levels associated with it. A light pulse can excite a molecule in the ground state to a sublevel of the first excited singlet state. Direct transition from a singlet state to the lowest-energy triplet state by light does not occur in the electric-dipole approximation, because light waves cannot supply the rotational momentum necessary to flip an electron spin from the antiparallel to the parallel orientation. From the sublevel of the first excited singlet state the molecule can decay to the lowest vibrational state in the excited singlet state by transferring its vibrational and rotational energy to the surrounding medium by means of rapid collisions with other molecules. This decay process within an electronic manifold is a nonradiative transition and is characterized by a nonradiative decay time.

Once the molecule reaches the lowest singlet level it can return to the ground state by emitting fluorescent light, transfer its energy into heat, cross over to the manifold of different spin levels or transfer its energy to another molecule. Weak fluorescence from a dye indicates either a strong coupling between the singlet and triplet levels (leading to a fast intersystem crossing rate), or nonradiative depopulation of molecules to the ground state by vibrational decay, or a very slow decay rate between the singlet states.

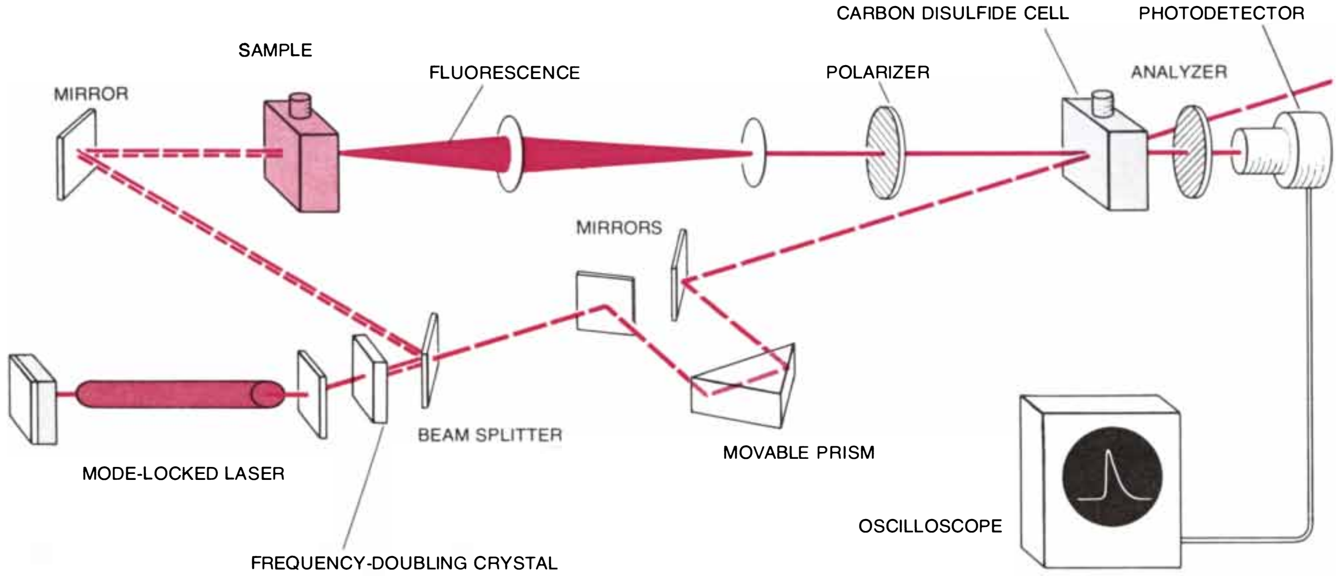
A key for understanding molecular interactions is provided by the time behavior of fluorescence, because if a dye molecule is dissolved in a solvent, it can be excited into a new molecular configuration, and the solvent molecules will rearrange themselves spatially in response to the new configuration [*see bottom illustration on page 50*]. This spatial rearrangement is very rapid and can be monitored by changes in fluorescent intensity and wavelength. The fluorescent wavelength becomes time-dependent when the energy levels shift in time in response to dynamic molecular interactions. Such a dynamic interaction occurs when solvent molecules instantaneously "see" a solute-dye molecule change its electronic shape and dipole moment on excitation. A large difference between dipole moments in the ground state and the excited states leads to strong dynamic interactions with the solvent dipole moment; the molecules rapidly readjust by strong electrical dipole-dipole interactions. These dipole interactions cause the fluorescent wavelength to become time-dependent, so that solvent-solute interactions can be studied by choosing solvents with different electrical characteristics.

Many techniques have been developed recently that harness picosecond laser pulses for the study of the fluorescence process in the ultrashort time domains. In the simplest type of fluorescence-decay measurement, developed by Michael E. Mack of the United Aircraft Corporation, the mode-locked laser beam excites the fluorescence, which is then detected by a fast photodiode and displayed on a traveling-wave oscilloscope [*see illustration on preceding page*]. The resolution of such a system is typically a third of a nanosecond, which is sufficient for measuring the fluorescence-decay time for a number of dyes. For the measurement of relaxation lifetimes the duration of the laser pulse must be shorter than the relaxation lifetime and the laser-pump light must be absorbed in the material. The ruby laser and the neodymium-glass laser fulfill these essential criteria by emitting short, powerful pulses, which can be converted by nonlinear optical processes to any wavelength from the ultraviolet to the infrared.

The fundamental output wavelengths of these lasers are the red wavelength of 6,943 angstroms for the ruby laser and the infrared wavelength of 10,600 angstroms for the neodymium-glass laser. By second-harmonic generation (frequency-doubling) in a potassium-dihydrogen-phosphate (KDP) crystal powerful pulses can easily be generated at the ultraviolet wavelength of 3,472 angstroms and at the green wavelength of 5,300 angstroms. Most substances can be pumped with just the fundamental



SHORT-LIVED BIREFRINGENCE can be caused in a liquid by the passage of an intense pulse of polarized laser light (*color*). As the light pulse enters the liquid cell it causes the randomly oriented molecules (*small black dots with arrows*) in the liquid to become reoriented in the direction of the pulse's electric-field vector (*large white arrow*). As a result the refractive index of the liquid in the region of the pulse is increased in the direction of the electric-field vector, producing a local, time-dependent birefringent condition.



OPTICAL GATE with a shutter speed of eight picoseconds is used by the authors to measure extremely rapid events such as the rise and decay of the fluorescence of dye molecules. In this system, developed by Michel A. Duguay and John W. Hansen of Bell Laboratories, a mode-locked laser generates an intense, picosecond pulse at the infrared wavelength of 10,600 angstroms (*long dashes*); the infrared beam then passes through a frequency-doubling crystal, which converts part of the beam into a green beam at a wavelength of 5,300 angstroms (*short dashes*). A dielectric mirror transmits the infrared beam and reflects the green beam. The infrared pulse travels along an optical path (whose length can be varied by a movable prism) before entering a cell containing liquid carbon disulfide, where it induces a short-lived birefringence in the liquid. Meanwhile the green beam is absorbed by the sample, exciting that substance to emit fluorescence. The fluorescence is collected by a lens and is passed through the gate, overlapping the infrared beam in the cell. Birefringence in the liquid induced by the presence of the infrared pulse causes the portion of the fluorescence beam that is coincident in time with the infrared pulse to change its polarization from linear (due to the first polarized filter) to elliptical. In this manner an eight-picosecond section of the fluorescence beam is transmitted through the analyzer for examination.

and second-harmonic pulses emitted by these two lasers.

The oscilloscope-photodiode system is not fast enough to give information on nonradiative relaxation times, fluorescent rise times or ultrashort fluorescent lifetimes. The use of picosecond pulses as the basis of a clock mechanism eliminates these electronic time-resolution difficulties. In two early independent experiments J. Shelton and John A. Armstrong of the International Business Machines Corporation and Richard I. Scarlet, Joseph F. Figueira and Herbert Mahr of Cornell University pumped dye molecules in solution to a higher electronic state with intense picosecond pulses; both groups then used a weak second pulse to measure the rate at which the molecules returned to the ground state. Nonradiative relaxation times as short as six picoseconds were measured in this way.

The next major advance was the development of a technique for actuating an optical gate with picosecond pulses. In this technique, which is applicable for any light-emitting process over a wide range of wavelengths, the mode-locked laser beam both excites the fluorescence and triggers the gate, which operates like a superfast camera shutter. The device is called a gate because when it is open it allows light to pass, whereas when it is closed it prevents its transmission. One idea behind the optical gate is to try to use the picosecond pulses themselves in a device, because then the time resolution will be in picoseconds and electronic problems will not arise.

The operation of an optical picosecond shutter is based on the principle of the Kerr shutter. The time resolution of a conventional Kerr shutter is not high enough for the measurement of really rapid decay rates, but these difficulties have been overcome with a picosecond Kerr optical gate developed by Michel A. Duguay and John W. Hansen of Bell Laboratories. In this device the temporally ultrashort electric-field pulse applied across the gate is provided not by an electronics system but by the electric field of the light pulse itself.

Since light is an electromagnetic wave made up of electric and magnetic fields it is an ideal source for an electric field. The strength of the electric field associated with the light emitted by an ordinary incandescent lamp, however, is too low to operate the shutter. In a mode-locked laser pulse the electric field can be as high as a million volts per centimeter, which is ideal for operating the ultrashort shutter.

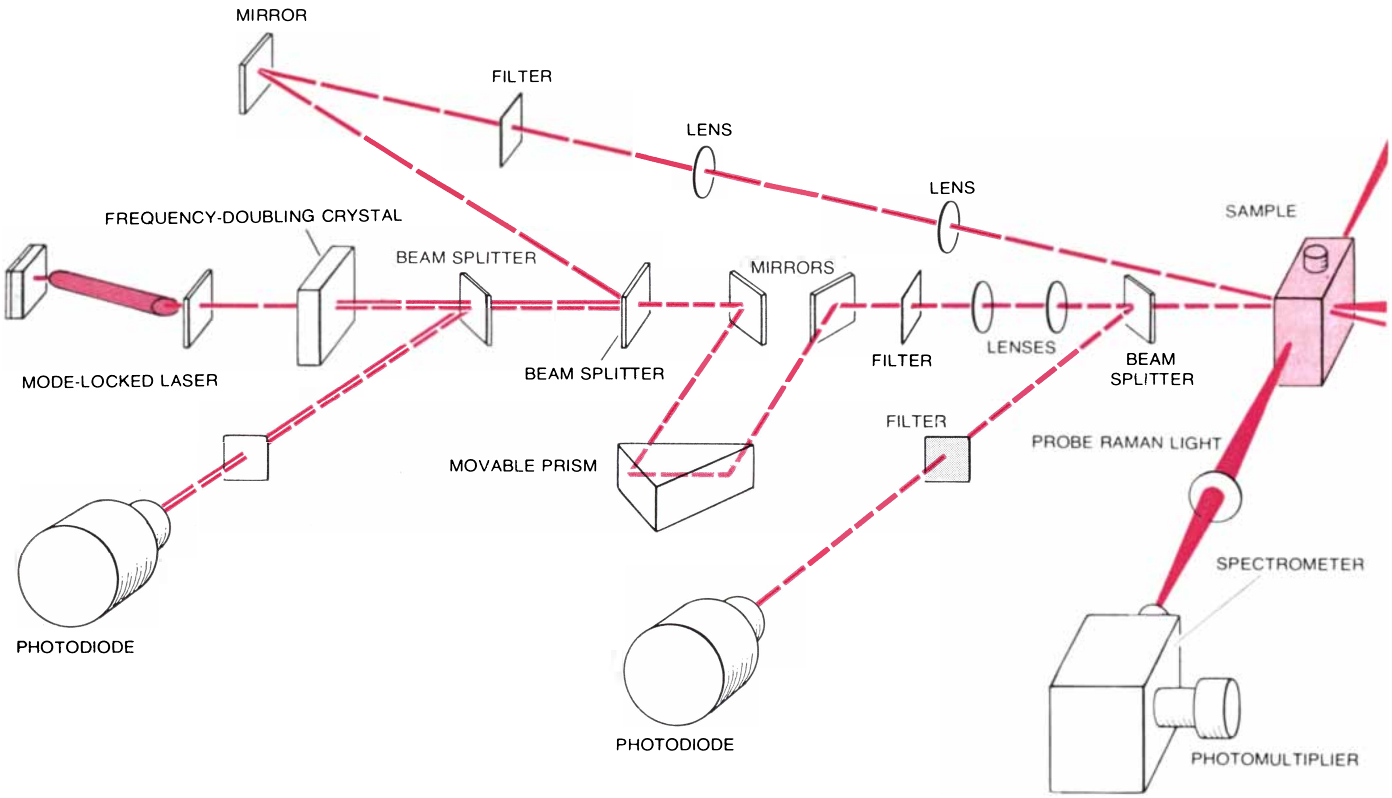
As the laser pulse travels through the liquid, its electric field polarizes the molecules contained within the laser light-pulse duration, thereby inducing a short-lived birefringence as a function of time [*see illustration on opposite page*]. Note that the direct-current Kerr gate and the picosecond gate differ; the latter requires no metal-plate electrodes in the liquid, because the intense electric field is provided by the light pulse, not by a voltage source. The time response of the induced birefringence in a given region is the result of both the pulse duration of the laser (typically six picoseconds) and the orientational response time of the molecules, which can be as short as a few picoseconds.

Picosecond optical gates with shutter times of about eight picoseconds have been constructed, and they have been used successfully to measure very rapid events [*see illustration above*]. In such a system a mode-locked laser generates an infrared pulse of high intensity and short duration at a wavelength of 10,600 angstroms. This beam passes through a frequency doubler (a KDP crystal), where 10 percent of the beam is converted into a new green beam at a wavelength of 5,300 angstroms. A dielectric mirror transmits the infrared beam and reflects the green one. The infrared pulse travels along an optical path whose length can be varied with a movable prism. The infrared pulse enters a centimeter-long cell containing liquid carbon disulfide (CS_2) and induces a short-lived birefringence in the liquid.

Meanwhile the green pulse is absorbed, thereby exciting the sample to fluoresce. The fluorescence is collected by a lens and is passed through the gate at an angle of about five degrees with respect to the direction of the infrared beam. The fluorescent beam is positioned by a lens at the cell so that it overlaps the infrared beam in the cell. The polarizer is oriented so that the fluorescence becomes linearly polarized at an angle of 45 degrees with respect to the plane of incidence on passing through the cell. As long as the intense infrared beam does not pass through the carbon-disulfide cell, the fluorescence remains linearly polarized on passing through the cell and the analyzer extinguishes the fluorescence beam.

Birefringence induced by the presence of the infrared pulse causes the portion of the fluorescence beam that is coincident in time with the infrared pulse to change its polarization from linear to elliptical. Then light is transmitted through the analyzer. Essentially an eight-picosecond section is carved out of the fluorescence profile for examination. Different sections of the profile can be carved out by delaying the shutter pulse with respect to the pulse that excites the fluorescence in order to examine the time behavior of the fluorescence. The exact coincidence time between the infrared pulse and the green pulse can be obtained by measuring the intensity profile of the green pulse with respect to the time it takes for the pulse to pass through the gate. By monitoring the fluorescence signal as a function of delay time between the fluorescence excitation and shutter-opening time, the actual buildup and decay of the fluorescence can be determined.

This novel picosecond shutter was



RAMAN CLOCK exploits the phenomenon of spontaneous and stimulated Raman scattering produced by picosecond laser pulses to measure the ultrashort lifetimes of molecular vibrations in liquids. In this system, developed by the authors while they were at the General Telephone and Electronics Laboratories, the molecular vibrations are created in the sample through stimulated Raman scattering by means of a powerful infrared pulse from a mode-locked laser. This treatment increases the population of molecules in the excited vibrational state in direct proportion to the intensity of the stimulated-Raman-scattering light. A second, weak probe pulse at half the wavelength (green) is then applied some time after the first pulse and undergoes spontaneous Raman scattering off the excited vibrations. The intensity of the probe Raman light is proportional to the number of vibrations still present in the sample. By varying the delay time between the pump pulse and the probe pulse (with a movable prism) and by measuring the inten-

used by Duguay and Hansen to measure the rapid fluorescence decay time of the dye molecules 1,1′ diethyl-2,2′-dicarbocyanine iodide and cryptocyanine to be 22 and 14 picoseconds respectively. We have used the new shutter to measure the entire temporal behavior of the fluorescence intensity for erythrosin dye in water. The fluorescence of erythrosin rises quickly, reaches a sharp peak 60 picoseconds after excitation and rapidly decays to 37 percent of its intensity in a time of 90 picoseconds. These results yield a vibrational-rotational relaxation time in the excited electronic singlet manifold of 40 picoseconds.

An investigation of how different solvents affect these characteristic times reveals in some measure how vibrational and rotational energy is transferred within the same molecule or to surrounding molecules. The fluorescence-decay time of 90 picoseconds for erythrosin, for example, indicates a strong coupling between the singlet and the triplet levels, resulting in a very fast intersystem-crossing rate, since the lifetime can be this short only if molecules are depopulated from their excited state by mechanisms other than fluorescence.

The use of a picosecond Kerr gate has enabled Figueira and Mahr to measure exciton lifetimes in cadmium selenide of a few hundred picoseconds. (An exciton is a conceptual entity created in a crystal lattice by the exceedingly temporary pairing of an excited electron and an electron "hole.") Other applications of the gate have included measurements by us in collaboration with Michael Seibert of the General Telephone and Electronics Corporation of the fluorescence intensity profile of chlorophyll molecules, indicating a new step in the process of photosynthesis. Furthermore, we have shown that the picosecond broad-band emission discovered in our experiments [*see top illustration on page 42*] is ideal for detecting short-lived species through their characteristic absorptions. P. M. Rentzepis of Bell Laboratories has also reported measurements of nonradiative relaxation times in dye molecules.

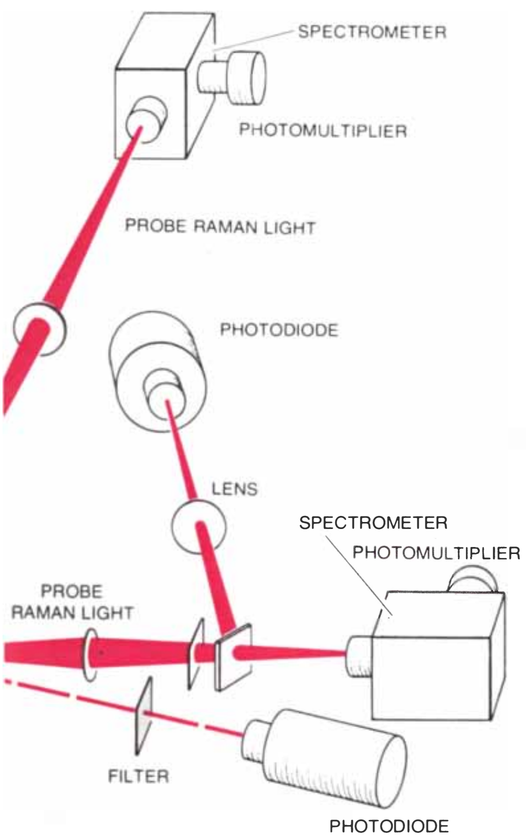
The transfer of the rotational motion of a molecule to its environment has been demonstrated by Kenneth B. Eisenthal, Karl H. Drexhage and T. Chuang of IBM with another picosecond-probe technique. They used a linearly polarized ultrashort light pulse to preferentially excite molecules oriented in a specific direction, thereby creating an anisotropic distribution of excited molecules. Then they monitored the rotation of these molecules back to an isotropic distribution by probing the absorption of the medium with light of a different polarization at later times. Their technique makes use of the phenomenon known as dichroism (the same phenomenon that enables Polaroid sunglasses to absorb one polarization of light while passing another, thereby eliminating the glare of reflections). It provides a unique way of investigating the interactions of a molecule with its environment. Solute molecules can be dissolved in different solvents and excited with picosecond pulses. Direct information on solvent-solute interactions is obtained by measuring the rate of return of the excited distribution to a randomly oriented distribution.

Experimental measurements of orientational relaxation times are surprising because they agree with a simple model that ignores detailed molecular interactions and instead treats a molecule rotating in a liquid much as one might treat a large rotating sphere (except that the gravitational forces in the molecular case are negligible). The surrounding medium is treated as a macroscopic entity. A specific molecular interaction that was expected to be important and to influence the orientational relaxation time is the hydrogen-bonding interaction. In this interaction alcohol molecules are bonded together with the positively charged hydrogen atom of one molecule bonded to the negatively charged oxygen atom of another molecule. This specific interaction, however, was found to play a negligible role in determining the rotational relaxation time of the large dye molecule rhodamine 6G in many liquids.

Further Applications

We shall now describe how picosecond pulses have been adapted to the measurement of vibrational lifetimes in liquids and to the determination of optical-phonon lifetimes in crystals. The adaptation of picosecond pulses for such measurements is perhaps their most important application so far for gaining an understanding of the transfer of energy within and between molecules.

A voluminous scientific literature accumulates each year on the normal vibrational modes of molecules in liquids and on optical phonons in crystals. The principal information for determining the structure of molecules and materials is derived from the frequency, line width, depolarization and intensity of spectral lines in Raman and infrared spectra. All these data tend to present a picture of a rather quiescent molecular world. Actually the molecular world is full of agitation and commotion as it continuously undergoes structural transformations. Because of the lack of an ultrafast clock with picosecond resolution, little attention had been paid to these interesting phenomena. The fact that everything happens so quickly (typically in a tenth of a nanosecond or less) had made it impossible to track the events. Now that these rapid transformations can be followed with picosecond pulses physicists are for the first time obtaining direct answers to such fundamental questions as "How long do molecular vibrations last?" and "What are the routes of decay of a molecular vibration or a phonon to other degrees of freedom in



sity of the probe Raman light created from the scattering off the excited vibrations, one can determine both the growth and the decay of the excess vibrations. By measuring the intensity spectrum of the probe Raman light with respect to delay time, the decay routes and growth and decay times of subsidiary vibrations can also be measured.

100 reasons to invest $395* in the time and money machine:

To obtain the mean and standard deviation of a series of numbers, for example, you simply key in the individual numbers, press the [Σ+] key after each entry, then the [$\bar{x}$] key to display the mean, and the [x⇄y] key to display the standard deviation. Because the HP-80 has the formula for calculating the standard deviation

$$S_x = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

built into its logic circuitry, it automatically performs all the averaging, subtraction, addition, squaring and square root extraction required by the formula.

The HP-80 Business Calculator.

The 9-ounce HP-80 has 40 specific capabilities built into its internal circuitry, which combine to perform more than 100 different financial calculations involving a relationship between time and money. All necessary programs, including a 200-year calendar, are built into the HP-80's solid-state memory.

It used to require up to 20 minutes to solve such time and money problems as bond calculations, figuring compound interest, loan repayments, depreciation amortization, investment analysis, sinking fund and statistics. To make matters worse, they required constant reference to cumbersome tables. You'll notice this is all in the past tense.

With the advent of the new HP-80, the financial counterpart of the pocket-sized HP-35, these problems can be solved literally in seconds, and without reference to any tables.

You just enter the data, push a few keys, and read the answer — accurate to within the last penny on a million-dollar transaction.

Since it can perform most financial calculations in about 10 seconds, a user can save up to 20 minutes on a single problem. The price is $395 (*U.S. only, plus tax).

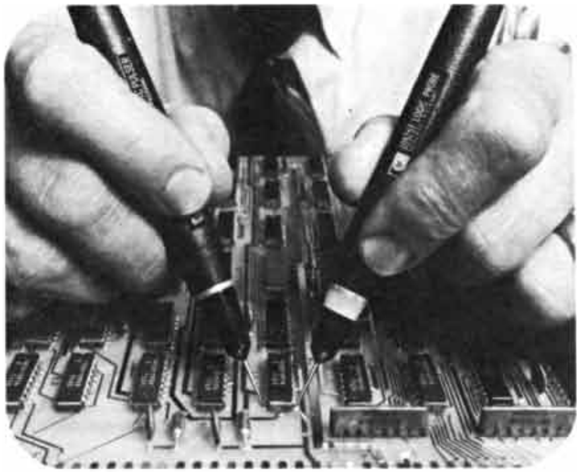
If you're in business, you can't afford to be without it. Ask for full information.

How to live happily in an increasingly digital world.

Most electronic products have changed drastically in the last few years. Everywhere you look, digital logic is displacing the familiar analog circuitry. And you'll be deceived if you believe that trouble-shooting these new products is the same as it was before the advent of the ubiquitous integrated circuit (IC). There are at least two important differences.

In the first place, you can easily check if any of the discrete components in an analog circuit are functioning properly by testing its input and output leads, usually two or, at worst, three. In a digital circuit, on the other hand, each IC has 14 or 16 leads and you may need to know what's going on at all the leads simultaneously.

Furthermore digital circuits are intentionally designed with low output impedances to make them insensitive to noise . . . perhaps their greatest single virtue. Unfortunately for circuit troubleshooters, this means that each IC input is clamped in either



a high or a low logic state; hence the only way to perform a stimulus/response test on an IC has been to unsolder its input lead before injecting a test pulse and observing the output.

We're happy to report that things have changed for the better in the world of digital logic testing, with HP's IC Troubleshooters.

First there's the Logic Clip which instantaneously indicates the logic state of all 14 or 16 nodes on an IC without cables or power connections. Or the Logic Comparator which compares an in-circuit test IC with a duplicate IC known to be good and identifies any dissimilarities.

For testing individual IC nodes, there is a series of Logic Probes covering the various IC families which immediately indicate logic states and pulse activity. New partner to the Logic Probes is the Logic Pulser which automatically injects a single pulse that drives a logic-low node high and a logic-high node low for 300 nsec. These are an effective stimulus/response test team — without unsoldering leads and despite low IC output impedances.

We've told the full story in an easy-to-read booklet that's yours for the asking.

For more information on the products described in these pages, fill out the coupon or write to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, Calif. 94304.

00312

Please send me information on the following:
( ) HP-80 Business Calculator
( ) IC Troubleshooters

Name____________________

Title____________________

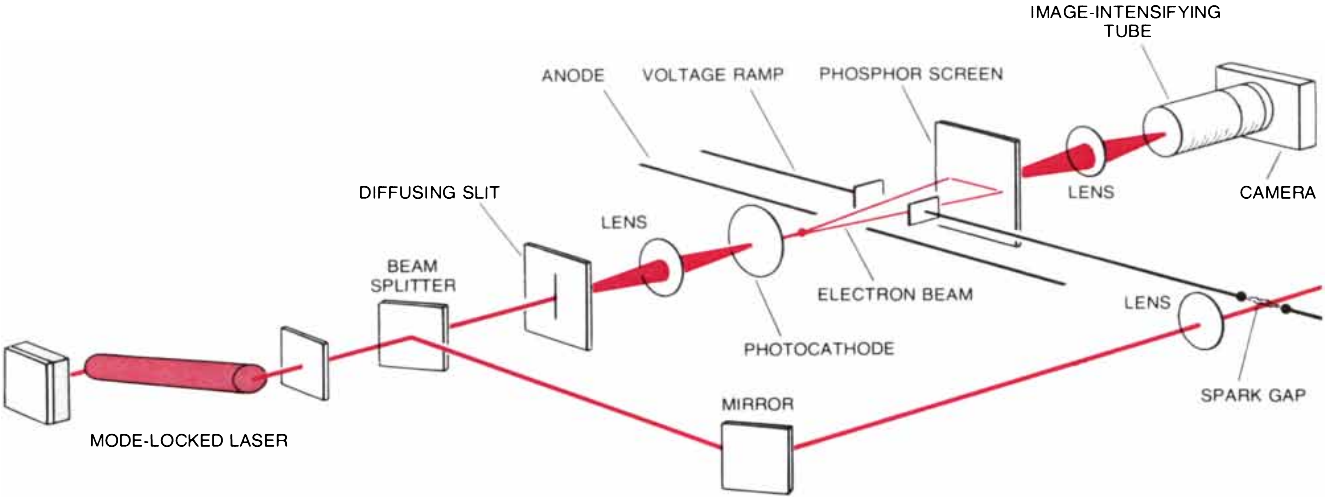
Company____________________

Address____________________

City__________ State______ Zip______

HEWLETT hp PACKARD

Sales, service and support in 172 centers in 65 countries.
Palo Alto, California 94304. Offices in principal cities throughout the U S



PICOSECOND STREAK CAMERA, developed recently by workers in several countries, has an electronic circuit fast enough to resolve atomic and molecular events on a time scale of about a trillionth of a second. Light from such an event enters a slit and is focused onto a photocathode, where electrons are released. The electrons are accelerated through an anode and are then deflected by a voltage ramp, which is actuated by focusing a portion of the original picosecond pulse into a spark gap. The increasing voltage in the ramp streaks the electrons across a phosphorescent screen so that the electrons released at different times strike the screen at different positions. A photograph of the resulting phosphorescent streak then gives a measure of the lifetime of the event.

liquids and crystals?" It is now possible to explore how such vibrations are transformed into one another in the same molecule or are transmitted to neighboring molecules. Moreover, by mixing different liquids one can learn how vibrations are transferred between different types of molecules.

Molecular vibrations and rotations last only picoseconds in liquids because the molecules move very rapidly, colliding frequently with one another. The collision rate is high because liquids are dense (1,000 times denser than air), so that molecules find themselves in close proximity to one another, where the electrical interaction forces are strong. Phonons in crystals collide frequently with one another for much the same reason. At room temperature the thermal energy is so high that much of it is converted into lattice-vibrational energy, with the result that immense numbers of phonons roam through the crystal. Under these conditions the dominant decay mechanism determining the lifetime of an optical phonon is a collision with another phonon that is already present in the crystal. This collisional decay is inferred to be a three-phonon interaction, where one phonon collides with a second phonon to yield a third phonon. At low temperatures, however, the decay mechanism must be quite different, because now few thermally excited phonons are present in a crystal and phonon collisions are relatively rare. Thus the predominant decay mechanism must be a spontaneous breakup of the phonon into two new phonons. The two new frequencies generated must satisfy the conservation laws of energy and momentum. In addition the frequencies must be limited to lattice-vibrational modes of the crystal, which are finite in number and uniquely determined by the crystal structure. Therefore only certain possible routes of decay exist for phonons. If the crystal contains an impurity, phonons can also decay by interacting with the impurity.

In liquids vibrational energy in a molecule can be transferred either to other vibrational modes of the molecule or to adjacent molecules. Symmetry of the molecular vibration is a determining factor in selecting decay routes. If one mixes two liquids, the vibrational motion of one molecular species can be transferred through collisions to the vibrational modes of the other species.

The mechanism of a clock capable of measuring ultrafast vibrational lifetimes exploits the phenomenon of spontaneous and stimulated Raman scattering produced by picosecond laser pulses. In this type of experiment the vibrations are excited through stimulated Raman scattering by means of a powerful picosecond laser pulse at the infrared wavelength of 10,600 angstroms [*see illustration on pages 54 and 55*]. This treatment increases the population of the vibrational state in direct proportion to the intensity of the stimulated-Raman-scattering light. A second picosecond probe pulse of weak intensity and different wavelength (5,300 angstroms) is then applied as a function of time after the first pulse and is spontaneously scattered by the created phonons. The intensity of probe Raman light is proportional to the number of vibrations still present. By varying the delay time between the pump pulse and the probe pulse with a movable prism and by measuring the intensity of the probe Raman light created by the scattering of the excited vibrations, one can determine the rise and fall times of the excess number of vibrations. Moreover, by measuring the intensity of the probe Raman light with respect to delay time, the decay routes and the growth and decay times of subsidiary vibrations created by the excited vibration can be measured.

Two different vibrational-decay times have been measured in liquids: depopulation times and dephasing times. As an example of the dramatic influence that phase can have on the time behavior of a system we have mentioned picosecond-pulse emission from lasers resulting from the fact that the modes of the light-wave field are in phase. Molecules in liquids can also vibrate in phase with one another, producing an enormous vibrational amplitude as they oscillate in step, a far greater amplitude than when they vibrate out of step. Such coherent molecular vibrations are generated in the stimulated Raman process where coherent light waves excite the medium. Vibrations oscillating in phase can go out of step with one another through interaction forces in the medium, and the vibrational amplitude then drops precipitously to a low value. The exponential

decay of the vibrational amplitude due to molecules' going out of step is called the dephasing time. The depopulation time is a measure of the decay of molecules to an entirely new state and is unlike the dephasing time, where the molecules vibrate out of step but can still remain in the same vibrational state. The depopulation time is a measure of the actual disappearance of created vibrations.

The measurement of depopulation times and dephasing times can be conveniently separated by experimental techniques. To measure the decay of the coherent oscillation by dephasing, the probe beam that is scattered by these vibrations at different delay times must come in at a particular direction such that conservation of momentum is preserved between the momentum vectors of the probe beam, the scattered beam and the coherent molecular vibrations. There are actually two particular directions corresponding to Raman frequencies that are upshifted and downshifted by the molecular vibrations. The depopulation of molecular vibrations is measured by incoherent Raman scattering of the probe beam by vibrations. The probe beam is sent in at an arbitrary angle with respect to the vibrations, and the scattered light is detected at an arbitrary angle. In this procedure the amplitude of the scattered light depends only on the population, since the vibrations at arbitrary angles contain little phase information. Strictly speaking, the two lifetimes are somewhat related. For example, if molecules oscillate in phase and one of them depopulates to a new state, then one less molecule is in phase and so the phase amplitude drops.

Using probe techniques, in 1970 we measured the optical-phonon lifetime in a calcite crystal to be 22 picoseconds at 100 degrees K. and 8.5 picoseconds at room temperature; we later estimated vibrational lifetimes in liquid nitrogen and benzene. In 1971 Alfred Laubereau, D. von der Linde and Wolfgang Kaiser of the Technical University of Munich, working with similar probe techniques, studied the decay of polyatomic molecules in the liquid phase and also investigated the decay of lattice vibrations in diamond. Later they extended their work to measure both the dephasing times and the depopulation times in ethyl alcohol and trichloroethane liquids. For ethyl alcohol they found a population lifetime of 20 picoseconds and a dephasing time 80 times shorter. The probe technique has also been extended by us to measure decay routes in liquids, where new "daughter" vibrations created from the decay of another vibration were observed.

Three recent developments in picosecond diagnostic techniques lead us to expect further advances in the measurement of rapid relaxation times. These developments are (1) the invention of a picosecond streak camera, (2) the discovery of a new technique for generating still shorter light pulses and (3) the development of a picosecond dye laser that is "tunable" over a range of wavelengths.

In the streak camera, which has an electronic circuit fast enough to measure picosecond events, light from a slit is focused onto a cathode where electrons are released and accelerated toward a phosphorus substance, which emits light [*see illustration on page 58*]. A voltage increasing with time is provided by focusing a picosecond pulse into a spark gap; the voltage streaks the electrons across the phosphor so that electrons released at earlier times appear at a different position on the phosphor than electrons released later. A photograph of the phosphorescence then gives a measure of the time behavior of the event. Such cameras have been built by A. John Alcock and Martin C. Richardson at the National Research Council in Canada, by D. J. Bradley, B. Liddy and W. E. Sleat at Queens University of Belfast in Northern Ireland, by L. Coleman and S. Thomas of the Lawrence Livermore Laboratory of the University of California, and by several groups in the U.S.S.R.

The second recent development is the discovery of mode-locking procedures capable of generating still shorter light pulses by M. J. Colles at Harvard University and by Kaiser and his co-workers at the Technical University of Munich. Pulses as short as .3 picosecond can now be generated reliably, and it seems probable that still shorter pulses can be obtained by this method.

The third recent advance, the development of a tunable picosecond dye laser, was made by E. Ippen, S. Shank and A. Dienes of Bell Laboratories. This laser combines ultrashort time resolution with convenient wavelength flexibility; moreover, it emits repetitive pulses and thus is adaptable to the averaging techniques required for low signal levels.

We therefore look forward to measurements of still shorter relaxation times, and to further applications of these powerful new techniques that illuminate the mechanisms of energy transfer in the submicroscopic world.


GM
MARK OF EXCELLENCE

Once again, we've designed the most versatile camera in the world.

Once we'd designed our first multi-purpose camera, we probably could have kept on selling it forever. Because it was the most versatile camera in the world.

But we're obsessive perfectionists, so we couldn't keep ourselves from improving our original. And we improved it in so many ways (16 to be precise) that it became a whole new camera: The Polaroid MP-4 Multipurpose Land Camera.

The MP-4 may *look* familiar. (And some of the things it does may *be* familiar. Like delivering color pictures in a minute, and black-and-white in seconds.) But it has a new range of capabilities.

For example, its modular design, coupled with a wider choice of accessories, lets you build (and pay for) only the exact photographic system you need.

If you want additional capabilities later on, you can add them later on.

It's easy to get optimal lighting for your subjects because the lighting arms are now completely adjustable. And the lamp heads, themselves, are removable (with heat shields) so they can be shifted anywhere. Even the switches and accessory outlets are more conveniently located.

Macrophotographs (formerly up to 10 times actual size) can now be made up to 14.8 times actual size. And critical focusing for high macro- and photomicroshots is facilitated by the addition of an aerial image focusing screen.

With the MP-4, subjects can be subjected to photographic scrutiny from almost any angle without being repositioned. Furthermore, when the unit is on a table, it can shoot charts on the wall or large objects on the floor. These feats are possible because the MP-4 column can rotate 360° around its vertical axis, while the camera can rotate 360° around its horizontal axis.

But the MP-4 camera is more than just movable. It is *removable* from the column. So when the subject can't come to the camera, the camera can go to the subject. At which point the camera with its tripod mount becomes a studio camera.

We could go on about the new MP-4 enlarger head, the self-cocking Copal shutter that accepts 5 Tominon lenses or the universal mount that lets you use other cameras *including* Polaroid's SX-70* (a camera that revolutionizes instant photography the way the first Land camera revolutionized conventional photography).

But you get the point. We took an incredibly versatile camera and made it even more versatile in many ways. And more sensible in other ways. And more economical in some ways.

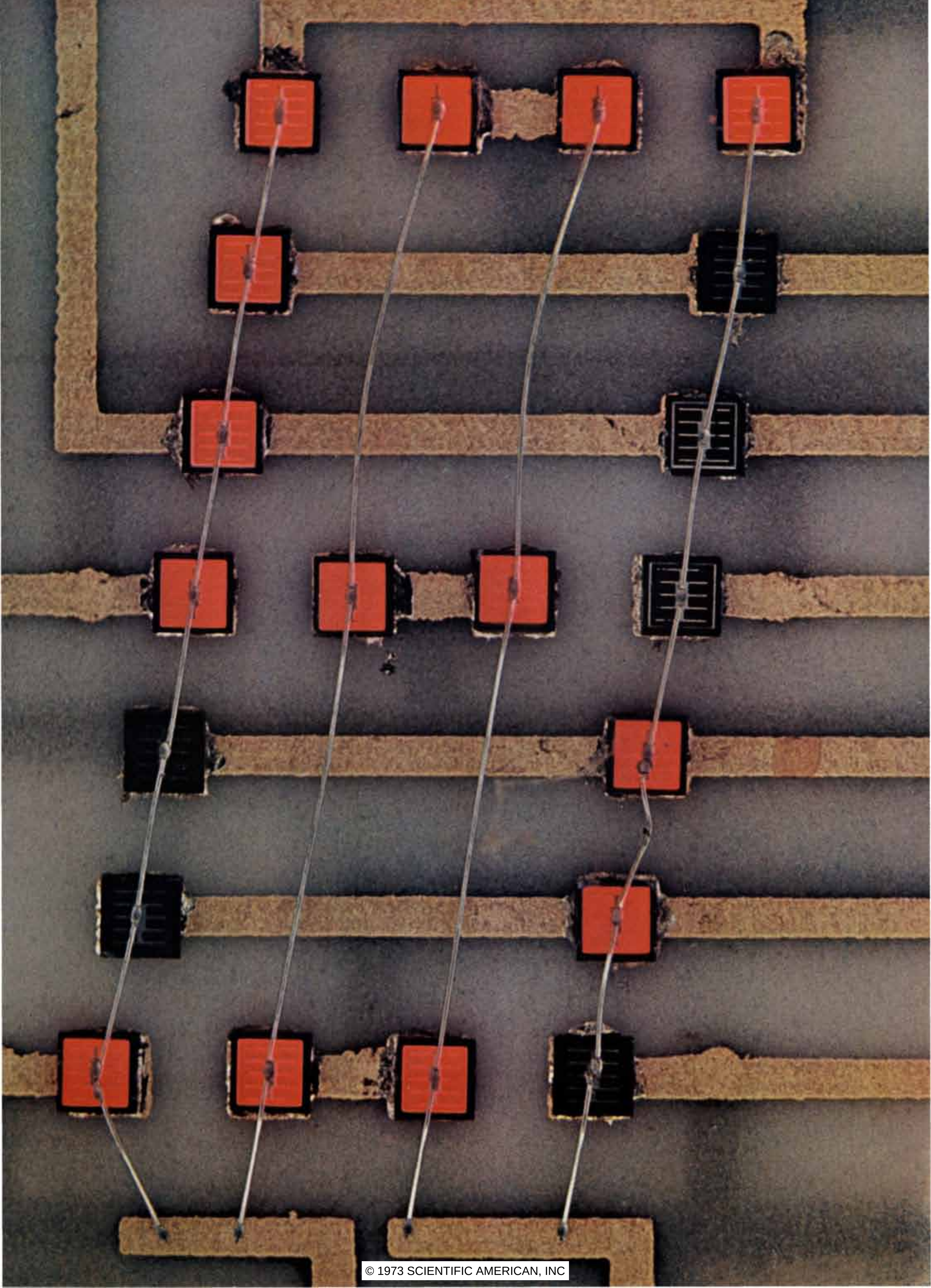
If you want a demonstration or more information, write to Polaroid Corporation, Dept. 57-257, Cambridge, Mass. 02139. (In Canada: 350 Carlingview Dr., Rexdale, Ont. In Puerto Rico: P.O. Box 2032, Ceramica Annex, Carolina 00630.)

WHAT IT CAN DO ON THE SPOT.

Copy wall charts
Enlarge negatives
Make 4 x 5 prints
Copy graphs
Record test results
Make 3¼ x 4¼ prints
Copy EKG's
Copy 35mm slides
Make 3¼ x 4 slides
Record fingerprints
Copy X-rays
Make product pictures
Make metallographs
Record defective parts
Photograph models
Make halftone prints
Record gross specimens
Illustrate research reports
Record legal evidence
Make close-ups of samples
Copy original artwork
Size type for mechanicals
Record electrophoresis results
Document research projects
Make positive/negative photographs
Make photomicrographs
Create graphic designs
Produce catalog artwork
Make macrophotographs
Take "before" and "after" records
Make copyprints of original photos
Illustrate training procedures
Record destructive tests
Copy engineering drawings
Record assembly sequences
Convert continuous tone material to line art
Copy results of nondestructive tests
Photograph transient phenomena
Photograph fluorescent phenomena
Make record of prototypes
Record insured valuables
Produce advertising pictures
Record bacterial cultures
Copy medical illustrations
"Memorize" laboratory setups
Copy scintillation photos
Photograph chromosomes for karyotyping
Et cetera.

The new Polaroid MP-4 Multipurpose Land Camera.

*Presently available in selected markets in the U.S. Polaroid®



ELECTRONIC NUMBERS

Numeric displays are now seen not only on laboratory and industrial devices but also on pocket calculators and wristwatches. Deciding what kind of display to use involves both technology and economics

by Alan Sobel

Electronic numbers are becoming a familiar sight not only on laboratory and industrial instruments but also on consumer items such as pocket calculators, digital clocks and electronic watches. A variety of electronic devices for displaying numbers have emerged in recent years. The rapid development of numeric displays has taken place both because of a growing demand and because of advances in electronics technology, particularly miniaturization and large-scale integrated circuitry, which have made it possible to produce electronic devices that are small in size and low in cost. Some of these devices require numeric readout displays. The kind of numeric display used in a device depends on the technical requirements, the cost of the display (and the electronic circuitry associated with it) and the aesthetic requirements (how the designer thinks the display will appeal to the user).

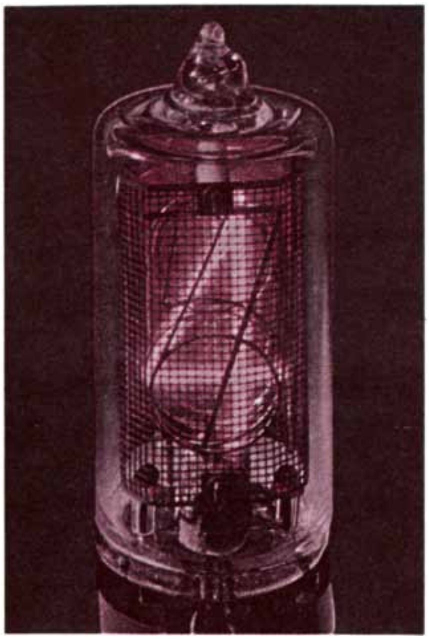
Although the technology of some of the new numeric devices is fascinating in itself, the evolution of electronic numbers can perhaps be better understood by looking at it in terms of an interaction between high technology, which determines what can be built, and the marketplace, which determines what can be successfully sold. To indicate a measurement most electrical and electronic devices have in the past incorporated some version of the d'Arsonval galvanometer, which consists of a wire coil suspended in the field of a permanent magnet. When an electric current passes through the coil, the coil turns and causes a pointer attached to it to move across a scale. Such meters can be made to be accurate to somewhat better than 1 percent of the full scale, but accuracies of .1 percent of the full scale are difficult to attain. When higher accuracy is required, more complicated techniques come into play; they generally involve balancing a calibrated electric current against the unknown current. Such techniques allow much more accurate measurements, but they are complex, slow and expensive. Furthermore, if the quantity to be measured varies rapidly, they do not work well.

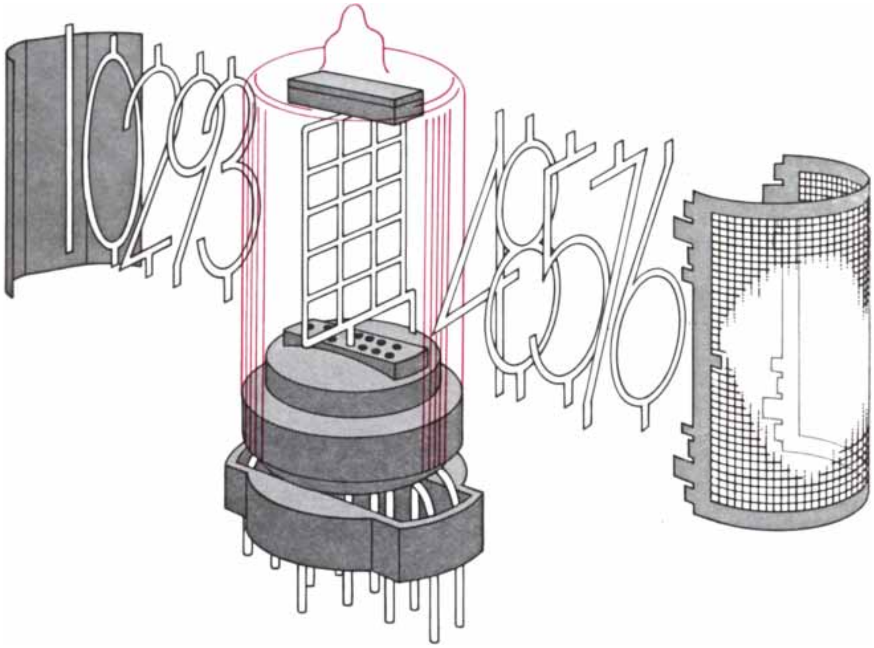
The development of transistors and later of integrated circuits has led to electronic measuring instruments of much greater precision. For example, multimeters capable of measuring voltage, current and resistance with an accuracy of one part in 100,000 are now available at prices under $2,000. These instruments do not require the operator to balance two currents; the input is read automatically and rapidly, and a new reading can be made in a fraction of a second. Accuracy of this order was formerly available only in standards-testing laboratories. The new electronic instruments not only are portable but also are more reliable than their predecessors.

Instruments that make rapid, high-precision measurements demand numeric readouts, since analogue presentations cannot display the results in any simple way. Moreover, many instruments are now designed to feed data directly to a digital computer as well as present the measurements to a human operator. Even low-precision two- or three-digit meters have some advantages over analogue displays, and they are rapidly replacing the meter movements that have been standard equipment in display panels. The digital meters can be read more easily and more accurately by untrained people who might have trouble interpreting a pointer on a scale. (On the other hand, it is harder to spot trends on a digital meter or to separate a trend from "noisy" fluctuations of the measured quantity.)

Another kind of measurement for which the d'Arsonval meter is not well suited is counting, particularly the counting of rapid and irregular events such as the disintegration of radioactive atoms. The Manhattan project of World War II called for many such measurements and spurred the development of electronic counters based on vacuum-tube circuitry. The amplified signal from the vacuum tubes went to small neon glow lamps that registered the count. The lamps were arranged in a binary code, and it proved to be a nuisance to mentally convert the binary numbers to decimal ones. Various techniques were devised to make a decimal readout possible. At the outset the display most frequently used was the "thermometer," or ladder, display. Each digit of the decimal readout required a column of 10 neon lamps. Only one lamp in any column would light at a time. In order to read a number it was necessary to look across the array of columns and note the digits that were lighted. Although these decimal displays were an improvement

ARRAY OF LIGHT-EMITTING DIODES for a single digit of a numeric display is shown greatly enlarged in the photograph on the opposite page. The red number 5 has been generated by selectively turning on appropriate light-emitting diodes in the array. The unenergized diodes are black. The diodes are made of gallium arsenide phosphide, a semiconductor that emits red light with a wavelength of about 6,550 angstroms. The filter that normally covers the numeric display has been removed to make the connective wiring visible.





NIXIE TUBE, a gas-discharge device, was the first commercially successful numeric indicator for electronic instruments. It consists of 10 shaped-metal numbers that are stacked one behind the other. The metal numbers are insulated from one another, and each of the numbers can serve as the cathode. The anode is a fine metal mesh. The entire assembly (*right*) is placed inside a glass bulb that contains neon with a small amount of mercury. When a voltage is applied between the anode and any cathode, the gas surrounding the cathode breaks down and emits light. The result is a luminous number (*photograph at left*). The Nixie tube was developed by the Burroughs Corporation in the mid-1950's. It is made in a variety of sizes and continues to be used in substantial quantities.

over the binary form, they were generally disliked because they were hard to read. What was needed was a means of displaying a number "in line" rather than scattered up and down several columns.

Of the devices that were designed in the mid-1950's to meet this requirement the most successful was a gas-discharge device called the Nixie tube. At the time the Nixie tube was introduced it was not at all certain that it would become the dominant digital device for electronic instruments. There were two major competitors: incandescent lamps and electroluminescent numbers. There were several ways in which incandescent lamps could be driven from the outputs of vacuum-tube counters, and the lamps could be used to illuminate masks or to edge-light plastic panels to produce a number display. The circuitry required to power these displays was more complicated and more costly than what was needed for the Nixie tube. Moreover, the incandescent indicators themselves were relatively expensive. The electroluminescent numbers were made from powdered phosphors that emit light when they are subjected to an electric field. Unfortunately the early electroluminescent lamps had short and unpredictable lifetimes and they gradually faded as serious competitors to the Nixie tube.

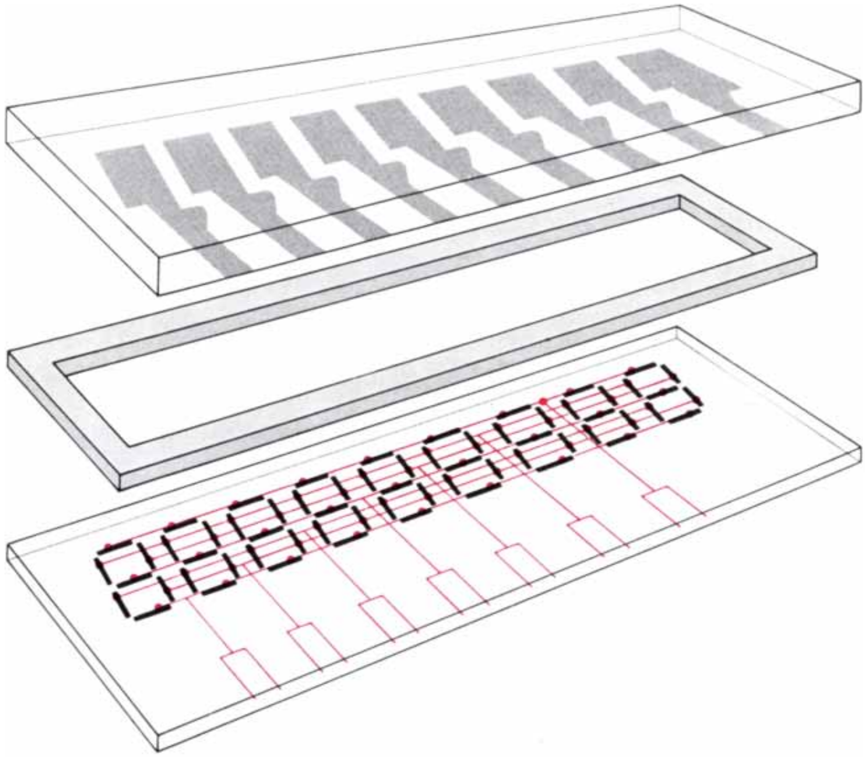
The name Nixie came about accidentally. A draftsman making drawings of the device labeled it NIX I, for numeric indicator experimental No. 1. His colleagues began referring to it as "Nixie," and the name stuck. The tube contains 10 metal cathodes, each shaped to form a different number. The cathodes are insulated from one another and are stacked one behind the other. The anode is a metal mesh. The entire assembly is in a glass bulb that contains neon gas with a small amount of mercury. When an electric potential of about 180 volts is applied between the anode and any cathode, the gas near the cathode breaks down and emits light. With a proper choice of gas pressure and cathode dimensions almost all the light comes from the immediate vicinity of the energized cathode, and the result is a luminous orange-red number.

The Nixie tube was first marketed commercially in 1956. It is still sold by its originator, the Burroughs Corporation, and by Burroughs' licensees in many countries. It is available in a variety of sizes and is widely used in measuring instruments of all kinds and in office equipment such as calculators and copying machines. The tube has been successful because it is reliable and has a

NIXIE NUMBERS are shown in sequence from 0 through 9 in this photograph of a section of a testing panel that is being used to determine the longevity of Nixie tubes. The tubes in the photograph have been in constant operation for more than 100,000 hours. Visibility of all the luminous numbers is good in spite of the fact that the number-shaped cathodes are stacked one behind the other.

long lifetime. Because it is a familiar device to design engineers the Nixie tube continues to be sold in large quantities.

The voltages to operate Nixie tubes are provided by circuits called drivers. Originally Nixie tubes were designed to be driven by vacuum tubes, which themselves operate at high voltages. Modern integrated circuits, however, operate at very low voltages, and interface circuits are required to drive Nixie-tube displays. These driving circuits are readily available from a number of sources, but the need for interface circuits, which provide a high voltage, is one reason why the Nixie tube is being challenged.



PLANAR GAS-DISCHARGE DISPLAY utilizes the same principle that the Nixie tube does to produce glowing numbers. The essential features of the planar device Panaplex II, made by the Burroughs Corporation, are shown in the drawing. The cathodes are metallic bars deposited on a flat base. Corresponding cathode segments for all the digits are bussed, or interconnected, by a buried conductor. Each digit has a separate anode, which is formed by depositing a transparent layer of conducting tin oxide on the inner surface of the front plate. A spacer separates the base and the front plate. Neon with a small amount of mercury is introduced into this space, and when a suitable voltage is applied between a cathode bar and its anode, the gas around the bar breaks down and begins to glow. Numbers are generated by sequentially energizing appropriate anodes and bussed cathodes, a technique that is known as strobing. This is done so rapidly that the viewer sees a steady light. The resulting numbers, as is apparent in the photograph, have a pleasing appearance.

The pocket electronic calculator is perhaps the most dramatic example of a new electronic instrument requiring a digital display. These small machines can add, subtract, multiply and divide, and many can perform even more complex calculations. Instruments with such capabilities would be impossible without large-scale integrated circuitry (known as LSI). To perform its calculations an instrument may need more than 1,000 diodes and transistors, and all are contained on one or two integrated-circuit "chips." The results of the calculations are displayed by numeric devices. Because pocket calculators are usually operated from small batteries, the power demand of the numeric display is more critical than it is for larger instruments that operate from a power line. Moreover, because of the size of the pocket calculator the numerals are usually smaller than those needed for other kinds of instrument. The numeric devices must be mounted close together and occupy as little space as possible. Since these calculators are sold to consumers rather than to business and industry, low cost is essential if they are to find a market.

The requirements of pocket calculators have had a strong impact on numeric-indicator design. The size of the market has stimulated manufacturers to develop new systems and new devices. There will probably be more than three million pocket calculators sold throughout the world in 1973, and since each calculator has from six to 16 digits the total demand for individual digits will be roughly 25 million. By way of comparison some five million Nixie tubes were sold throughout the world in 1972.

The Nixie tube has a shaped form for each digit; other devices use bars or dots to form the numbers. Seven-stroke devices form numbers by lighting the appropriate segments of a rectangle with a horizontal bar in the middle. The dot arrays are usually four by seven or five by seven [*see illustration on next page*].

Each type of device needs a different means of translating the signals received from the computing or measuring element of the instrument. The translation device is called a decoder. Although it is a relatively simple device to build with integrated circuits, it is not inexpensive. It is therefore often preferable to have all the digits to be displayed share one decoder.

One way to organize the display is to have a separate driver for each element of each digit. That is expensive. For an eight-digit display using seven-stroke numbers, for example, 56 drivers would be needed. If the eight digits are connected in a matrix arrangement in which the driving circuits are shared among all the digits, however, it is possible to reduce the number of drivers. This approach is termed multiplexing. For an eight-digit, seven-stroke display connected in a matrix, only 15 (8 + 7) drivers are required.

If in such an arrangement we want to activate only one segment of one digit, we send a current down a column and across a row. The segment at the intersection of that row and that column will receive the sum of the voltages from the two drivers. Other segments connected to the energized row and column will also be subjected to roughly half of the total voltage. If this half-voltage causes the display segment to produce some light, it could be difficult for an observer to distinguish between segments that should be on and those that should be off. This problem is called crosstalk, and unless it is solved some cross talk will be heard from the people who want to read the numbers.

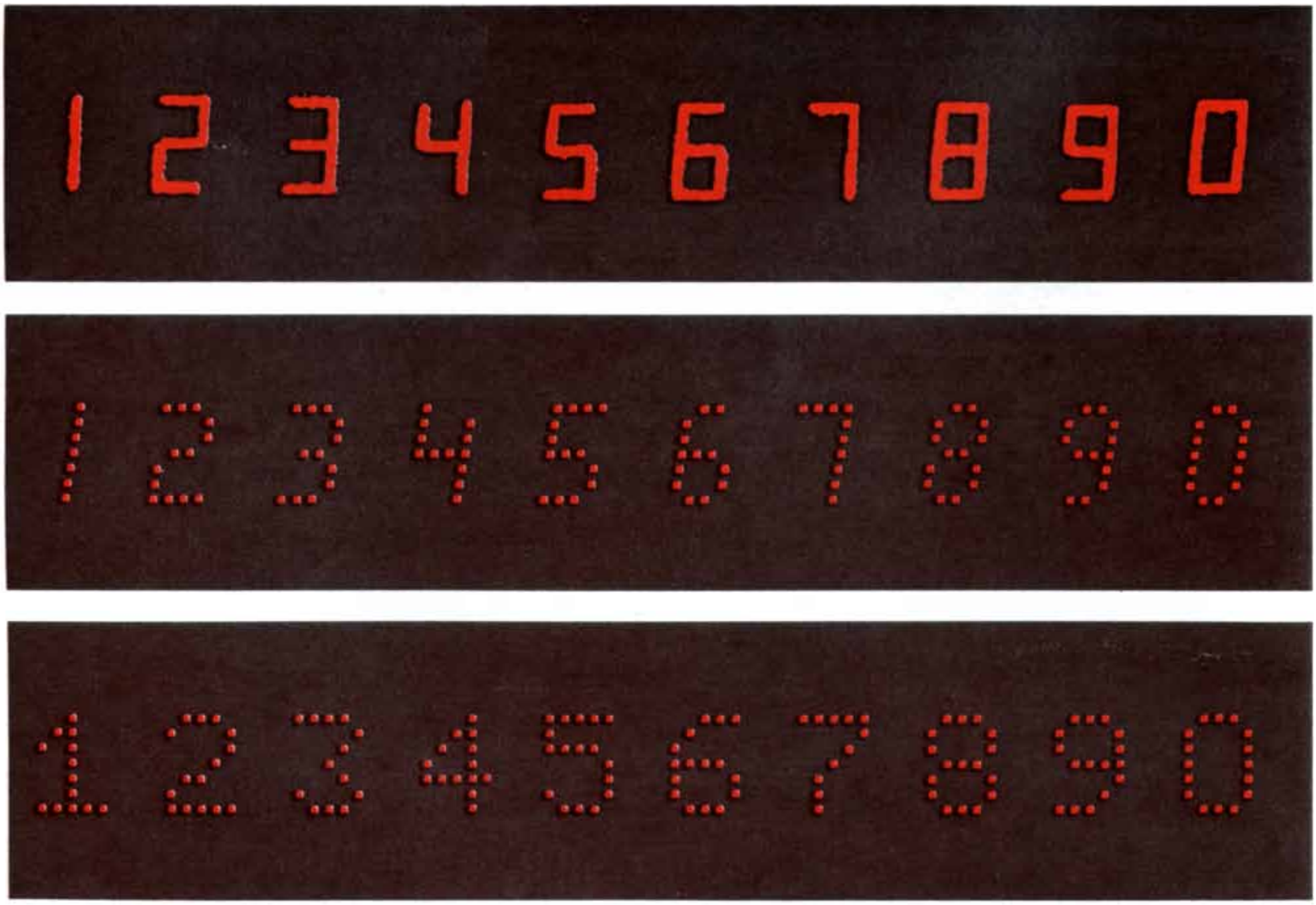
What is required is that each display element have a sharp on-off threshold so that partial voltages give rise to no light output. Many of the display devices satisfy this requirement, but there are some that do not. Some devices operate on direct current and so need a sharp threshold in only one direction; others operate on alternating current and should have sharp thresholds in both directions.

Another problem is that in multiplex operation each digit is "strobed" (is on) for only part of the time. Hence it must radiate all the light that the viewer will see during only part of the time. For example, in an eight-digit display each digit is on for an eighth of the total time and must emit eight times its average output during the on interval. Furthermore, the time it takes to pulse all the digits in the display must be short enough so that the observer does not see an objectionable flicker. These requirements set a limit on the number of digits of a particular kind that can be used in a multiplex display.

Multiplex operation requires a more complicated organization of the electronics than the simpler scheme of a driver for each segment. For a large number of digits this extra complexity is clearly better than the many drivers needed for nonmultiplex operation, but for a display of only a few digits the saving in drivers may not compensate for the increase in programming electronics. As complicated integrated circuits become cheaper, however, the cost of the programming electronics is going down, making multiplex operation favorable for smaller displays.

When the market for numeric displays began showing signs of great expansion, the Nixie tube was already a mature device. That is, it had been manufactured for a long time and costs had been reduced about as far as possible: to about $1.50 per tube in large quantities. Some further price reductions might be achieved, but substantial savings seem unlikely. Furthermore, it is difficult to make Nixie tubes in the very small sizes (less than three-eighths of an inch high) that are needed for pocket calculators.

A new approach utilizing the gas-discharge principle was devised. The cathode electrodes around which the gas glow concentrates were formed into seven-stroke rectangles in a single plane.



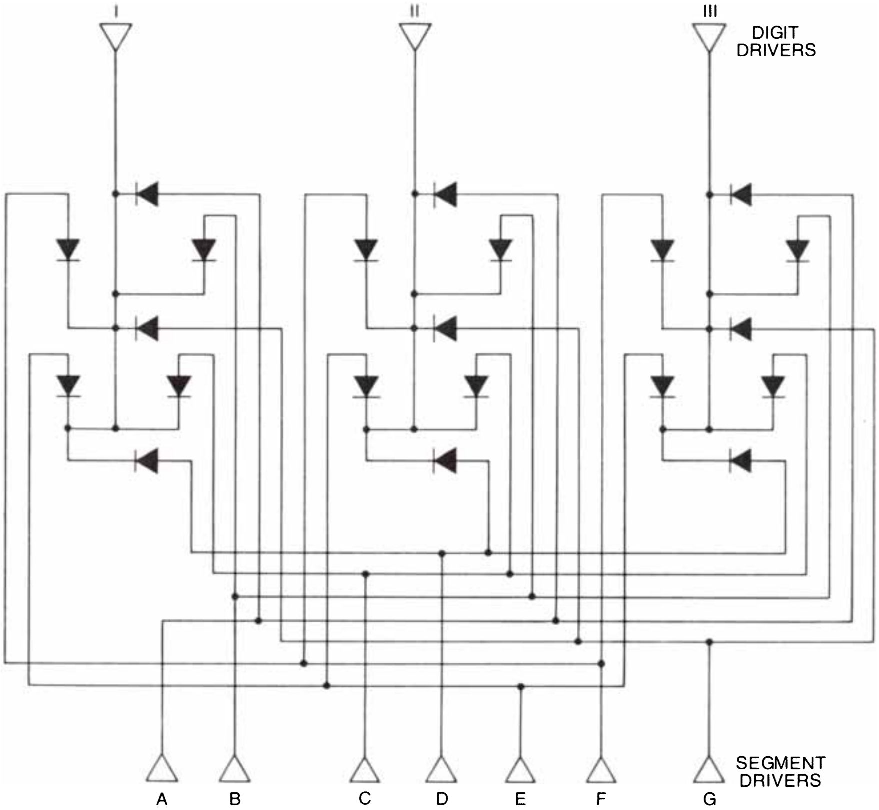
THREE TYPES OF ARRAY for forming electronic numbers have been developed. The simplest is the seven-stroke array, basically a rectangle with a horizontal bar. With this arrangement each number from 0 through 9 can be generated, as shown in the top photograph. All seven segments of the array can be seen in the number 8. The four-by-seven dot array (*middle photograph*) also is a rectangle with a horizontal bar. The five-by-seven array (*bottom*) often is filled in so that it has 35 dots. Such an array can generate not only numbers but also all the letters of the alphabet. The top photograph shows a light-emitting-diode display on the Model 80 pocket calculator made by Hewlett-Packard. The other two photographs show light-emitting-diode displays that are made by same company.

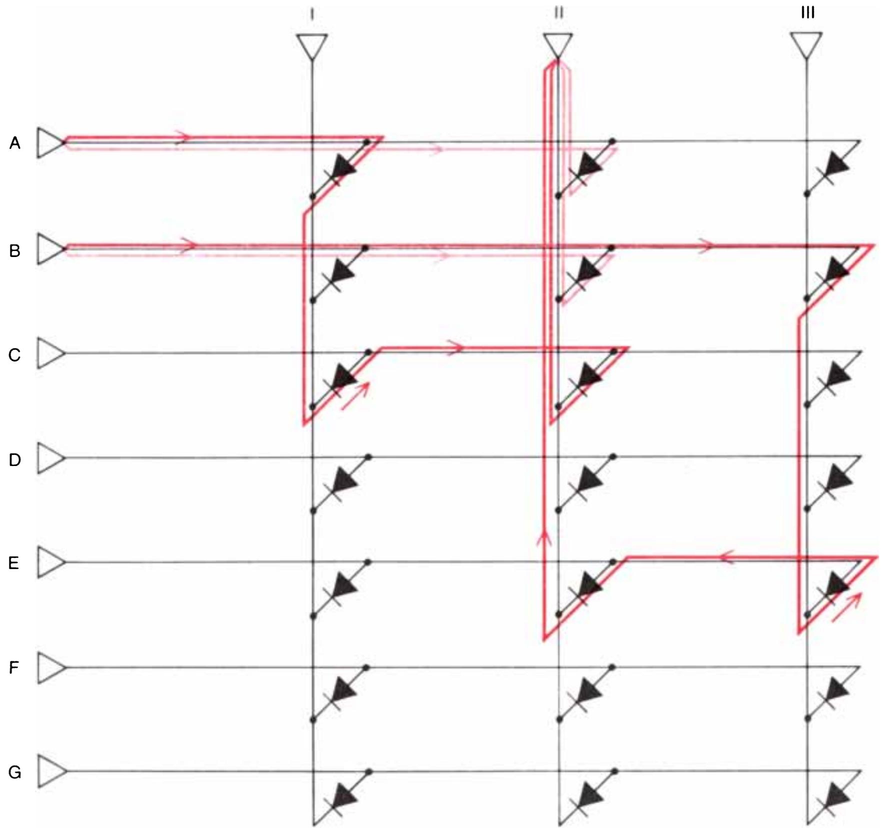
This stratagem resulted in thinner devices than the ones that could be made with Nixie tubes and opened the way to a more economical fabrication of the electrodes. The cathode electrodes are usually made by depositing a metallic paste on a suitable base and baking the paste, a technique called thick-film fabrication. Several manufacturers have produced variations of such planar gas-discharge devices. Since the numbers are all in one plane, they do not jump back and forth as the Nixie numbers do.

The segments in the seven-stroke matrix have the sharp thresholds needed for multiplexing. Although the total voltage across the gas discharge is high (about 200 volts), the difference between on and off voltages can be only 30 volts. Thus low-cost driving circuits can be used. The overall power economy is good even when it is necessary to include a converter to step up the voltage from a nine-volt battery. Multiplex displays of 10 or 12 digits can be made with planar gas-discharge devices. Costs range upward from about $1 per digit but will doubtless come down soon.

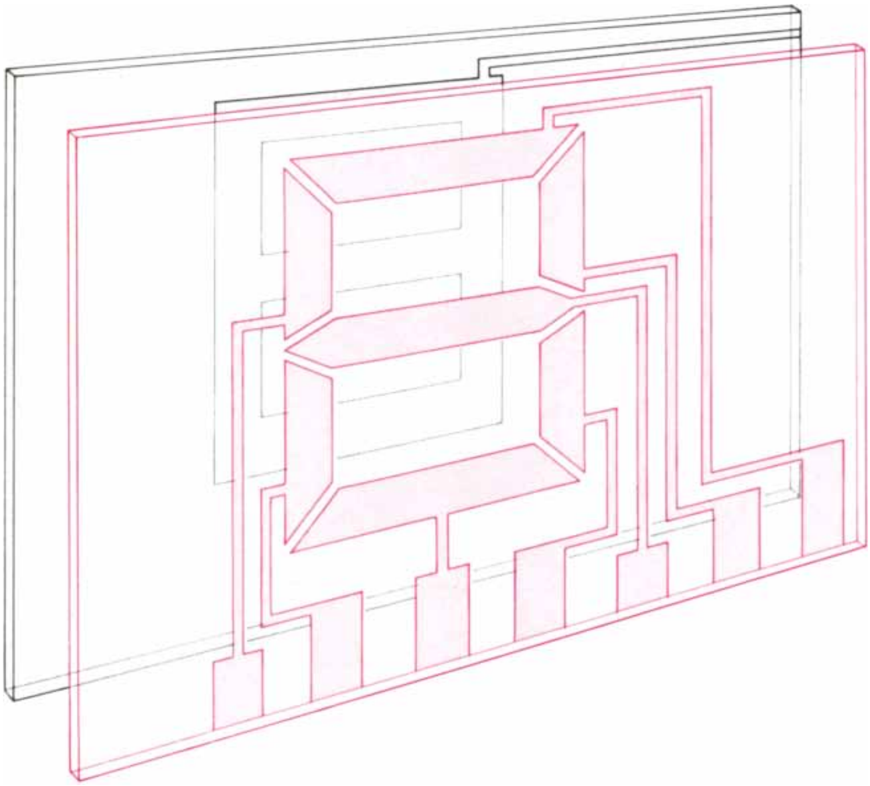
Two new technologies have made their appearance in numeric displays: liquid crystals and light-emitting diodes. Liquid crystals are fluids that are partially ordered, so that they have some of the optical properties of crystals (hence their name). Although they have been known for nearly a century, their recent application to numeric displays dates back to the discovery in 1968 that they have readily usable electro-optical properties [see "Liquid-Crystal Display Devices," by George H. Heilmeier; Scientific American, April, 1970]. Light-emitting diodes (LED's) are semiconductors that emit light when a suitable electric current is applied to them. The semiconductor materials most commonly used in them today are gallium phosphide and gallium arsenide phosphide.

At present there are two types of crystal used in displays: dynamic-scattering liquid crystals and field-effect liquid crystals. Dynamic-scattering liquid crystals are clear in the absence of an electric field. When an electric field is applied, they turn cloudy and scatter light. The effect is rather like a frosted piece of glass. The devices can be made transmissive for rear-lighting applications, reflective for ambient-lighting applications or semireflective for both kinds of operation. The usual liquid-crystal numeral is arranged as a seven-stroke digit [*see top illustration on next page*]. The strokes of the digit are created by transparent elec-

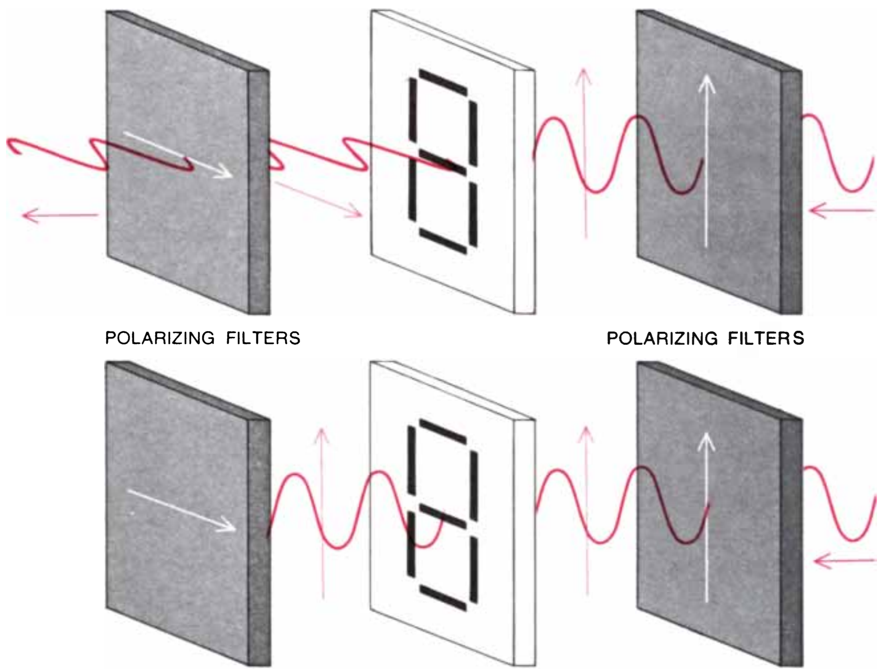




MULTIPLEX CIRCUIT indicates how three seven-segment digits can share segment and digit drivers (*top*). The diagram shows semiconductor diodes, but other kinds of element could be used if they have suitable thresholds. The same array is redrawn (*bottom*) to emphasize the three-column, seven-row matrix configuration. Desired paths for energizing a 1 in the middle digit are shown as light-colored lines. Some "sneak" paths that could turn on unselected segments are shown in the darker color. If sufficient light is generated by the sneak currents, it may be difficult to determine which of the segments are supposed to be on.



DYNAMIC-SCATTERING LIQUID-CRYSTAL DISPLAY typically consists of a nematic liquid crystal between two plates of glass on which electrodes have been deposited. The seven-segment electrode on the inner surface of the front plate is transparent. The continuous electrode on the rear plate can be either transparent or silvered. In the absence of an applied electric field the nematic liquid crystal is clear. When an electric field is applied to any segment, the liquid crystal in that region becomes turbid and scatters light. Numbers are formed by energizing appropriate segments. Dynamic-scattering displays can be either reflective or transmissive. In the latter case auxiliary rear lighting is usually required.



FIELD-EFFECT LIQUID-CRYSTAL DISPLAYS require polarizing filters on both sides of the basic cell, which contains a twisted nematic type of liquid crystal sandwiched between two plates with transparent electrodes. When polarized light passes through the liquid crystal, the plane of polarization of the light is rotated 90 degrees (*top*). When an electric field is applied to the liquid crystal, it untwists and no longer rotates the light (*bottom*).

trodes on the inner surfaces of two sheets of glass between which is a thin film of liquid crystal. If the device is to be reflective, the electrode on the rear surface is made of metal to serve as a mirror.

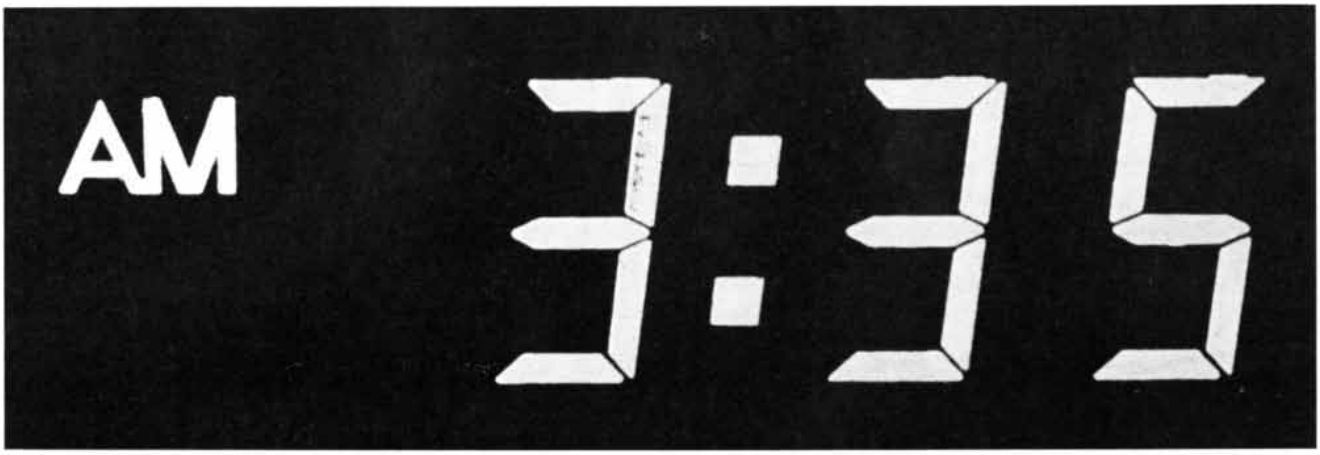
It requires very little current (on the order of microamperes per square centimeter) to produce the liquid-crystal scattering effect. Direct-current operation shortens the life of the device; hence alternating current is used. The voltage can be 25 volts or less, so that liquid-crystal displays are well suited to low-power applications such as the face of an electronic wristwatch. Transmissive displays call for auxiliary lighting. The contrast between the numbers and the background in the early displays was not very good, but improvements have been made rapidly. For reflector-backed displays there are some angles at which the numbers disappear under some lighting conditions. Another disadvantage of dynamic-scattering liquid crystals is that they do not have sharp enough thresholds for multiplexing, although driving techniques to overcome this drawback have been developed.

Field-effect liquid crystals, a newer development, operate by rotating the plane of polarization of incident polarized light [*see bottom illustration at left*]. Their thresholds are good enough for them to be used with multiplexing in displays of a few digits. They can operate in either the transmissive or the reflective mode. They can be operated on direct current, and their power requirement is even lower than that of dynamic-scattering liquid crystals. Typically they operate at about eight volts, but threshold voltages of less than one volt have been achieved.

Many companies are conducting intense research and development in liquid crystals. The temperature range over which these substances function is being extended; it is now from below zero degrees Celsius to more than 80 degrees C. (176 degrees Fahrenheit). The lifetimes of the devices are being lengthened and electro-optical performance is being improved; in short, manufacturers are learning how to make liquid-crystal displays of high reliability at low cost. They promise to become the least expensive as well as the least power-demanding of all numeric displays. Prices of 50 cents per digit should be attained in the near future. The fact remains that they are still new materials and their large-scale commercial use is only beginning.

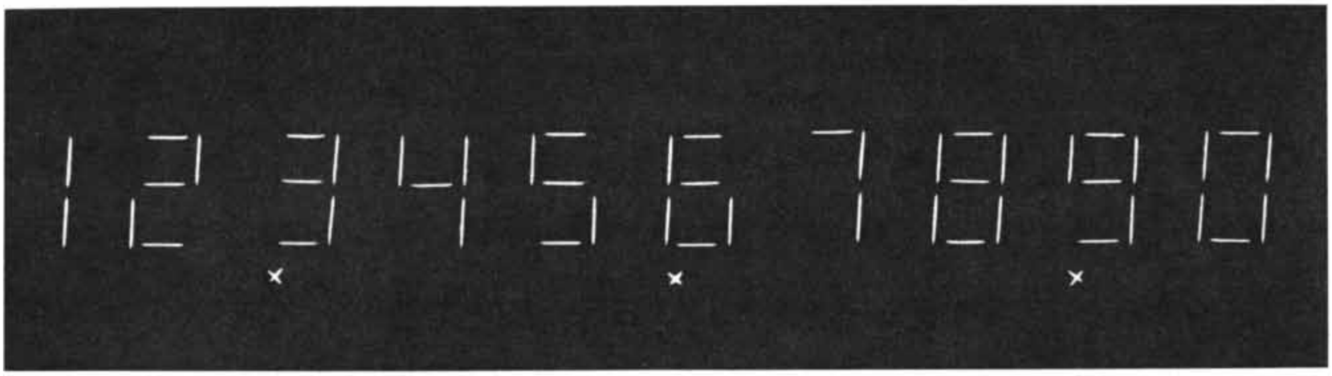
The light-emitting diode is an outgrowth of work on semiconductor materials. When an electric current is ap-

DIGITAL ELECTRONIC WATCHES are providing a new consumer outlet for numeric displays. The watch on the left has a dynamic-scattering liquid-crystal display made by the Optel Corporation. Watch on the right, made by Gruen Industries, Inc., has field-effect liquid-crystal display made by the International Liquid Xtal Company. Both systems operate by reflecting ambient light.



DIGITAL ELECTRONIC CLOCK has a liquid-crystal numeric display with a 12-hour format and with separate A.M. and P.M. indicators. The liquid-crystal display is made by American Microsystems, Inc., which also is a maker of electronic clock circuits.

ELECTRONIC CALCULATORS with large liquid-crystal numeric displays are being manufactured by the North American Rockwell Microelectronics Company for several major retail concerns. The company began mass-producing liquid-crystal displays last year.



INCANDESCENT-FILAMENT display devices have very bright, sharp numbers. These devices, produced by the RCA Corporation, consist of a seven-stroke array of a tungsten wire enclosed in a glass tube. Brightness can be adjusted by altering the voltage. The filament gives off a wide-spectrum yellowish-white light that can be filtered to produce a numeric display of almost any color.

plied to certain semiconductor crystals, light is produced by a recombination of electrons and holes near a *p-n* junction.

Light-emitting-diode displays are fabricated as seven-segment numbers or as four-by-seven or five-by-seven dot arrays. The materials and the fabrication processes required are expensive. Therefore LED's are most often used in small sizes, which mates them well to small portable devices such as pocket calculators. Most LED's emit red light, although yellow and green are becoming available and blue is being worked on. LED's have a sharp threshold and a bright output and are well suited for multiplexing. They operate on voltages around 1.7 volts, which makes them compatible with integrated circuits. Their efficiency, however, is low: they convert only about 1 percent of the electrical input into emitted light. Since LED's require substantial current and have an output that increases rapidly with increasing voltage above the threshold, most LED drive circuits are designed to supply a well-defined limited current. The devices are mechanically rugged and should have a long life.

Because of strenuous price competition the cost of LED displays is dropping rapidly. The lowest published price at present is about $1.60 per digit for large orders. Prices of $3.50 to $5 per digit are more typical, but price competition with other display types is intensifying.

Vacuum-fluorescent indicators are small special-purpose cathode ray tubes. A thin heated wire emits electrons that are drawn to any of the seven segments of a digit by applying a positive voltage to the desired segments. The segments are coated with a low-voltage phosphor that emits a blue-green light. Other colors can be produced by putting appropriate filters in front of the tube. The devices require power in the milliwatt range and about 80 milliwatts to heat the filament. The anode typically calls for about 25 volts. Because of this low voltage the phosphor must be viewed from the same side that the electrons hit it. The filament is thus between the lighted segments and the viewer. It is so small and operates at such a relatively low temperature, however, that it is unnoticeable [*see illustration below*]. Indicators of this type are made in a variety of sizes and formats and have been used successfully in small portable calculators. They are low in cost, they multiplex well and their low voltage requirements make them compatible with integrated circuits often without the addition of separate drive circuits.

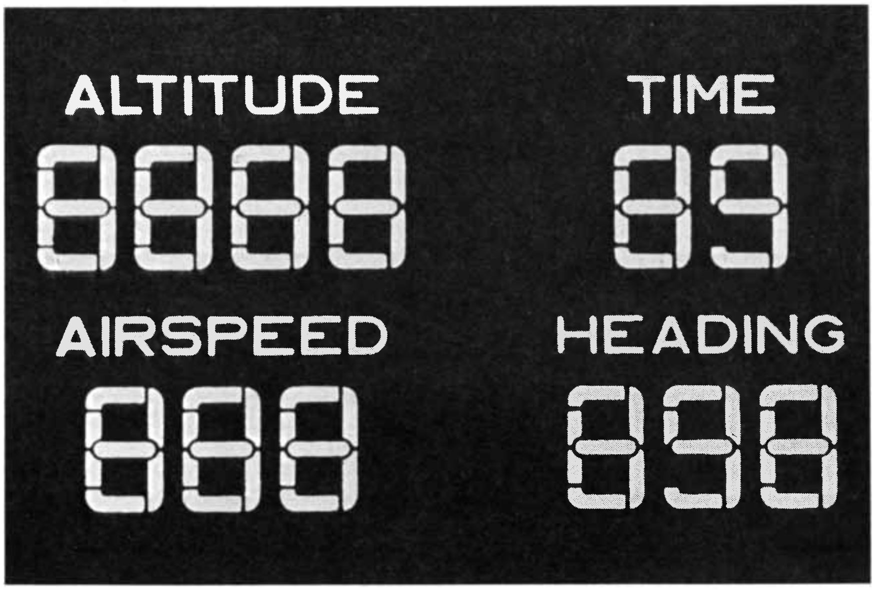
Electroluminescent materials were one of the early competitors to the Nixie tube. Powdered electroluminescent numerics are still made, and there have been substantial improvements in their performance. They do not, however, multiplex well without additional circuit elements.

A newer form of electroluminescent display has been developed, consisting of a thin film of phosphor deposited onto a glass base. The phosphor and the front electrodes are transparent. A black layer of dielectric material is deposited on the back of the phosphor (the side away from the viewer) and absorbs most of the incident light. The contrast of lighted numbers against the black background is excellent even in brightly lighted environments. Power requirements are in the 50-milliwatt range, but both forms of electroluminescent devices need alternating currents in excess of 100 volts for good performance.

VACUUM-FLUORESCENT numeric indicators are basically small cathode ray tubes. A thin wire cathode at the front of the device emits electrons that are drawn to phosphor-coated segments. The light emitted by the phosphor peaks in the blue-green region but spans a wider spectrum, thus making possible multicolor displays with the appropriate filtering. A new flat, rectangular fluorescent display made by the Wagner Electric Corporation is shown in this photograph. Fluorescent devices are also packaged in glass tubes.

Because packaged driver circuitry is not generally available users must produce their own circuits. This has tended to discourage use of these displays, even though their cost is as low as 50 cents per digit. A new, direct-current form of powdered electroluminescent material is being developed.

Incandescent tungsten filaments in vacuum tubes have also been used in display devices, usually arranged in the seven-segment format. These devices produce sharp, high-brightness displays. They emit a yellowish-white light that can be filtered to any desired color. The brightness is adjustable with simple voltage controls from zero output to a level that is easily visible even in very strong ambient-light conditions. The voltages required are low, making these devices directly compatible with integrated-circuit drivers, but multiplexing requires additional components. Power requirements are comparable to those of LED's. Because the filaments are operated at a relatively low temperature their life should be long. Costs are about $2 per digit and are expected to drop. Incandescent-tungsten devices have been used in testing and aircraft instruments and in applications such as cash registers with numeric displays.



ELECTROLUMINESCENT DISPLAY utilizing a light-emitting phosphor is made by Sigmatron, Inc. The phosphor, a transparent polycrystalline material, is deposited as a film on a glass plate previously coated with transparent electrodes. A black insulating layer is applied to provide a dark background. When the phosphor is excited by an alternating electric field, it emits yellow light. Contrast is excellent even in bright environments.

There are other types of numeric indicator in service or being developed. I have discussed here only the most important of those currently available. One new device harnesses electrophoresis to move small colored particles that are suspended in a liquid of contrasting color. Another incorporates tiny incandescent tungsten filaments, fabricated by integrated-circuit techniques, arranged in dot arrays. The technology of producing numeric displays is still evolving. We can anticipate that there will be new applications of electronic numeric devices, particularly in automotive dashboard instruments, home appliances and other consumer items.

There is no one display device that is clearly superior for all purposes. In fact, for many applications it is difficult to choose one device among several competitors. In addition to the obvious requirements of size and legibility the design engineer must take into account the environment (heat, cold, vibration and so on) in which the instrument will operate. Keeping down the cost of the display usually calls for juggling the cost of the various components, including the numeric indicator itself, the required drivers and interface circuits, the mounting of the display and the power supply. In order to make the best selection it is often necessary to design the display system several times, taking different approaches.

There are many instances where this kind of detailed engineering study is not justified or the time for it is not available. Clearly it would not be sensible to mount as large a design effort for the production of a few items as for the production of hundreds of thousands. (This is one of the reasons that equipment for the space program is so expensive. Only a few units are made, but an enormous amount of engineering must be lavished on those few.) For the usual low-volume-production design the engineer will be guided largely by his judgment. That judgment, in turn, will be affected by the ease with which he can obtain information, samples and the production quantities he needs. If one type of display is familiar and readily available, and if its driving circuitry can easily be produced from standard components, the designer is likely to select it over another device whose appearance, say, is more satisfactory but whose circuitry would be more difficult to incorporate into the final instrument.

Thus the choice of a display to be used in a particular product is often influenced by more than the strictly technical considerations of price and performance. The display manufacturer's field engineering may be as important as his quality control. This tends to favor devices manufactured by large companies with a good reputation and with the resources to provide advertising, marketing, applications literature and troubleshooting service to customers. Furthermore, it tends to favor established devices over newer ones. This is one of the important reasons that Nixie tubes are still being made in large quantities, even though the newer devices are growing in popularity. By the same token, the fact that so many new devices are being incorporated into production equipment is evidence both of the need for new solutions and the willingness of many engineers and executives to work with new technologies, particularly in circumstances where they can afford to do detailed engineering design.

Inventors have long been aware that the world does not beat a path to the source of the better mousetrap. That mousetrap must be sold to prospective users, and if it is to be incorporated into a mouse-control system, its advantages over tomcats must be convincingly demonstrated. Many engineering decisions concerning which device to use where are based not on exhaustive technical analysis but on the designer's judgment and intuition about which device will best meet his needs. His information will come from many sources: other users, trade literature and so on. Engineers usually find that they must make decisions with less information than they would really like to have. This is one of the challenges of design engineering; good engineers meet it successfully.

FOREST OF DOUGLAS FIRS in western Oregon dwarfs a field investigator suspended from a spar 100 feet above the ground (*at lower left of photograph, between first and second tree trunk*). Use of modified rock-climbing techniques allows such closeup studies.

LIFE IN TALL TREES

The high forest canopy consists of more than branches and leaves. Entire communities of other plants and animals dwell within this unique ecosystem, helping to provide the trees with nitrogen

by William C. Denison

A treetop in a forest, like a mountain peak or a deep canyon, is a remote world that is plainly visible but not easy to explore at first hand. Yet there is a strong urge to visit such private domains. They challenge both our curiosity and our skill. With respect to tall trees there might even be an echo of a past when our ancestors were at home in treetops.

It is perhaps a little surprising, therefore, that although investigators have been drawn to the direct exploration of mountain heights and sea floors, relatively little study has been given to the treetop world. It offers intriguing questions to a biologist. The forest canopy is a distinctive habitat, providing its own special conditions of moisture, light, temperature and other qualities. What kinds of plant and animal life does it support? Does it harbor an integrated community? What roles does it play in maintaining the forest ecosystem as a whole?

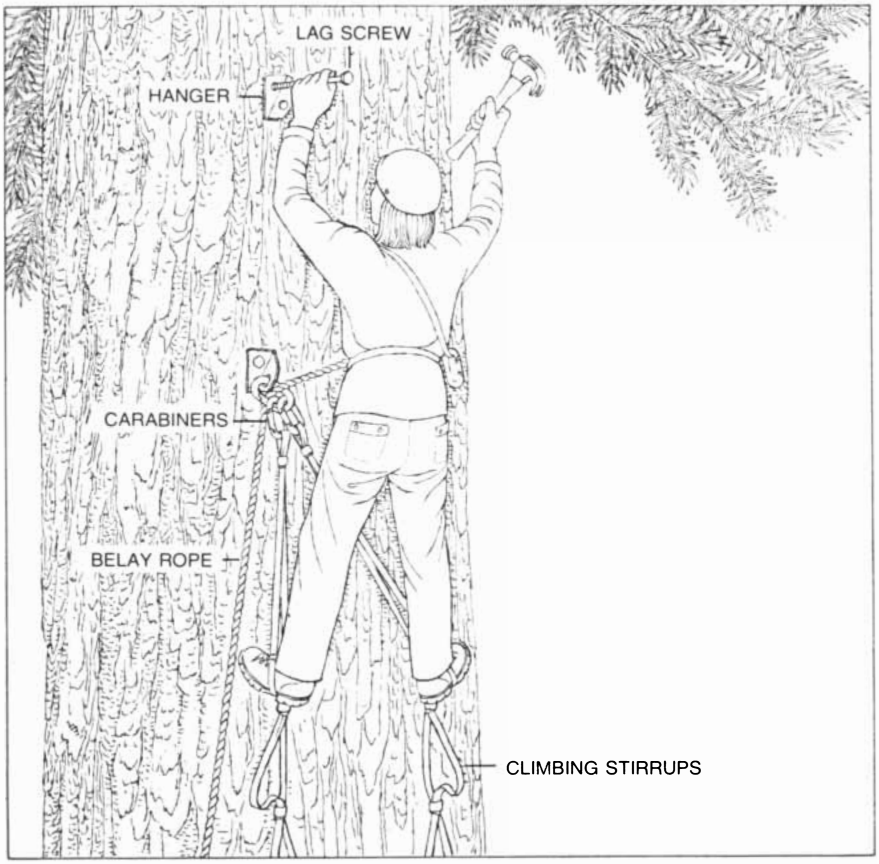
My colleague Lawrence H. Pike and I at Oregon State University undertook an exploration of treetop life as part of a U.S. study of ecosystems in western coniferous forests under the aegis of the International Biological Program. We enlisted a group of physically fit university and high school students in our climbing party and set out for the top of a 450-year-old stand of Douglas fir in the H. J. Andrews Experimental Forest in western Oregon, where the trees soar to a height of 200 feet or more and the lowest branch is usually at least 60 feet above the ground. Since the project presented a problem roughly corresponding to scaling the rock face of a mountainside, we borrowed ideas from rock-climbing to develop techniques for climbing the tall trees.

As in mountaineering, our climbers work in teams, with a teammate on the ground belaying the climber on the tree with a safety rope. The team members use small radio transceivers for communication between those on the ground and climbers far above in the canopy. In the first ascent of a tree the climber carries a hammer and lag screws (which hold better in the bark and trunk of a tree than the pitons, or spikes, used in rock-climbing) to assist the climb. The climber drives a lag screw into the tree trunk, fastening a hanger as far above his head as he can reach, attaches a pair of stirrup-loop ladders to the hanger, climbs the ladder stirrups to install the next hanger and proceeds in this way to the top of the tree [*see top illustration on next page*]. The initial climb takes several hours and is hard work, even for a person in excellent physical condition.

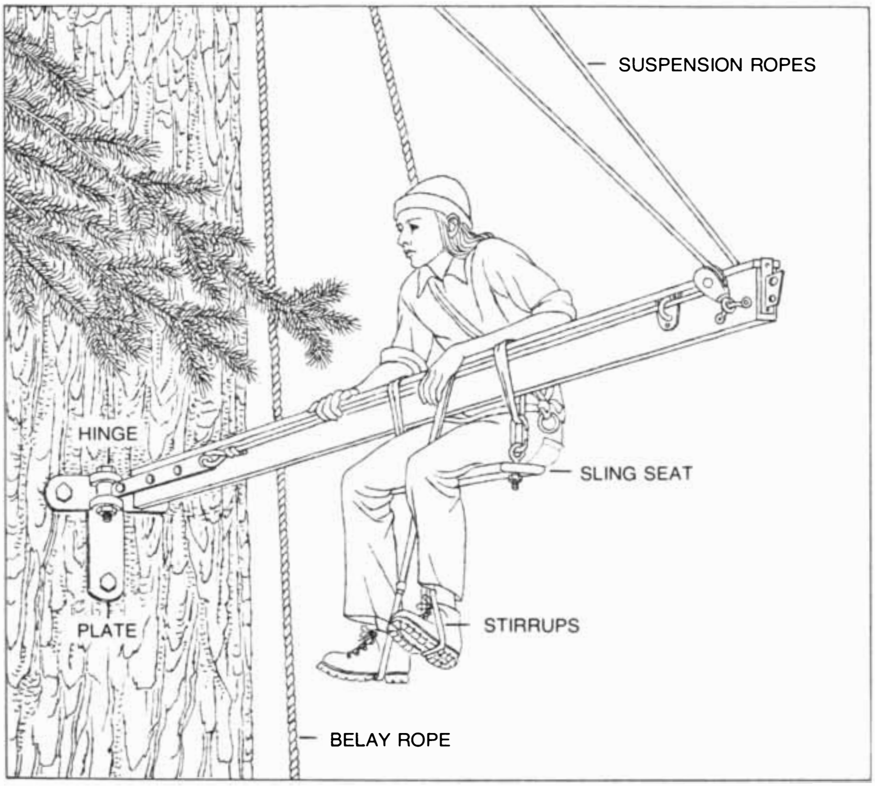
At the top of the ascent the climber attaches to the trunk a rope that will be used for subsequent ascents and a large pulley for the belaying rope that the climbers use for additional support. To ascend the climbing rope the climber holds in each hand a Swiss-made Jumar Ascender, which is a form of clamp that grips the rope tightly when weight is put on it and loosens when weight is removed. Slings of nylon webbing that serve as stirrups for the climber's feet are suspended from the jumars and the climber goes up the rope by taking the weight off one foot and stirrup, sliding the jumar up the rope with one hand, restoring the weight on the stirrup to tighten the jumar and then repeating the process with the other hand and foot. With practice one can climb the rope in this way with little more effort than it takes to climb an ordinary ladder. One can also rest one's feet from standing in the stirrups or free both hands for work at any point by sitting back on a seatlike sling that is attached to one jumar.

In addition to these devices for climbing and descending along the tree trunk, we have developed a 12-foot boom, which we call the "spar," for working in areas away from the trunk (including the ends of branches). The spar is attached to the trunk with a special hinge and is suspended at its outboard end by a pair of ropes [*see bottom illustration on next page*]. The occupant sits on a sling seat suspended from the spar and can move along the spar by alternately standing in stirrups while sliding the seat along and then moving the stirrups while sitting in the seat. By pulling one or the other of the two ropes supporting the outboard end of the spar the occupant can move horizontally around through an arc of 180 degrees. Thus we can work effectively anywhere in the canopy within 12 feet of the trunk.

After establishing access to the treetop, we carried out a systematic survey not only of the wood and foliage of the tree but also of the population of epiphytes: the many plants that grow on other plants but are not parasitic. We recorded the results of the survey on punch cards, a procedure that enabled us to store the data in a computer and to undertake computer analyses. The data were first used in preparing a "map" of the tree, showing the location and amount of living and dead matter on the tree trunk and on each branch system. The survey enabled us to estimate the biomass, or weight of living matter, of each branch system and of the tree as a whole. These estimates were compiled by considering separately the amount of foliage (obtained by observing what percent of the horizontal area within a branch system was covered by foliage), the amount of wood (in the branch and twigs) and the amount of epiphytic plant



ASCENT OF TREE is accomplished by means of the mountaineering methods designed for vertical rock faces, except that lag screws are used instead of pitons to secure hangers. The first climber up the tree rigs the climbing and belay ropes used by subsequent climbers.



HORIZONTAL SPAR provides a movable base among the tree branches. The investigator is supported by a sling seat hung from the spar and can study a 12-foot length of branch.

life inhabiting each branch. We checked our estimates of the biomass by making detailed measurements of typical systems. We found, for example, that in a Douglas fir 185 feet tall the foliage had an estimated total weight of 187 pounds and the epiphytic lichens and mosses growing on the tree weighed 38 pounds.

It was already well known, of course, that trees are inhabited by a great variety of plant and animal life. The plants that grow on tree trunks, branches and foliage include bacteria, algae, fungi, mosses, lichens and ferns. (In warm climates even advanced species of plants such as orchids are epiphytes.) The animals that live out their lives in trees show a similar range in size and diversity: protozoans, nematodes, higher invertebrates such as arthropods and mollusks, and various vertebrates (including primates in warm climates). As botanists, we focused our attention on the plant life in our exploration of treetops.

The forest canopy proved to be a subsystem of considerable complexity. The variety of its plant life reflects a variety in the habitats within the canopy. These habitats differ widely in the amount of available moisture and light, in temperature and in the age and surface texture of the supporting structure (twig or branch). As a twig grows and ages into a branch, its surface harbors a succession of different organisms, beginning with pioneering lichens and progressing through a series of complex communities. Diane Nielsen and Diane M. Tracy, two students who started with us as undergraduates and, pioneering our climbing techniques, were the first to ascend to the canopy, found that the diversity of habitats and the number of species of epiphytic plants increased the higher they went up the tree. Each habitat has a characteristic flora, with certain species predominating. For example, most of the large fir trees in the forest where we have been working are not strictly vertical but lean a little to one side. As a result their trunks have an upper side and a lower side, and the upper side is moister than the lower because rainwater streams down that side. The upper and lower sides therefore differ in the prevailing lichen species growing on them. On the branches aloft in the canopy certain large foliose (flat and leaflike) lichens predominate on the upper side of the branch, and the lower side tends to favor the lower plants known as liverworts.

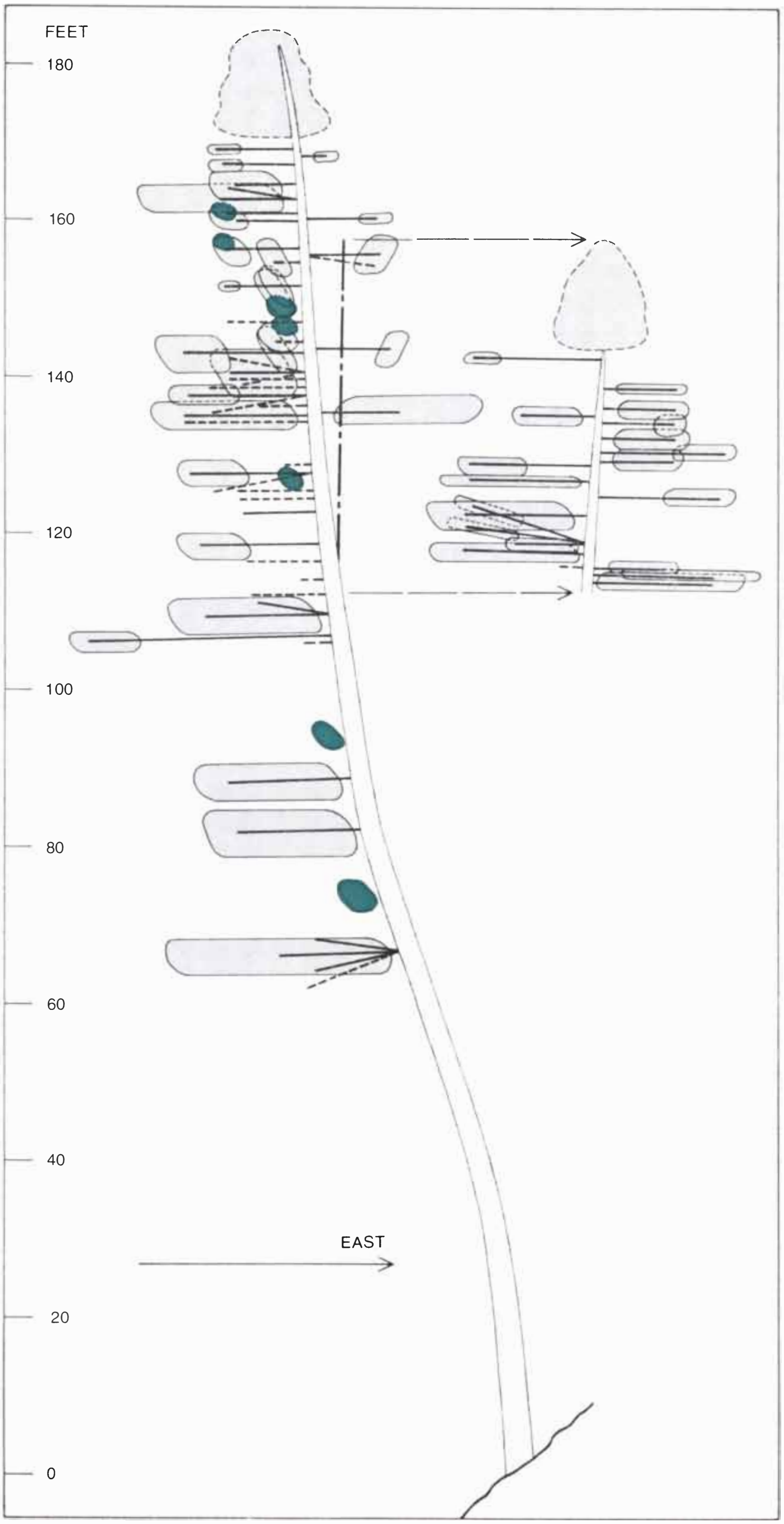
Habitat by habitat, we are cataloguing the characteristic epiphytic communities and obtaining a census of the many plant varieties. Pike has already noted 121 different species of lichens. The number will undoubtedly increase as we explore more trees.

We find, then, that the forest canopy is an active, well-populated system (comparable in many respects to the forest floor), and we must suppose that the epiphytes living in the canopy contribute substantially to the nourishment and viability of the forest as a whole. They undoubtedly take up water, minerals and other substances from the atmosphere. Through photosynthesis and other processes they produce nutrients that are released to the forest's animal dwellers and the trees. We have been particularly interested in investigating the epiphytic plants' role in capturing nitrogen for the forest's needs.

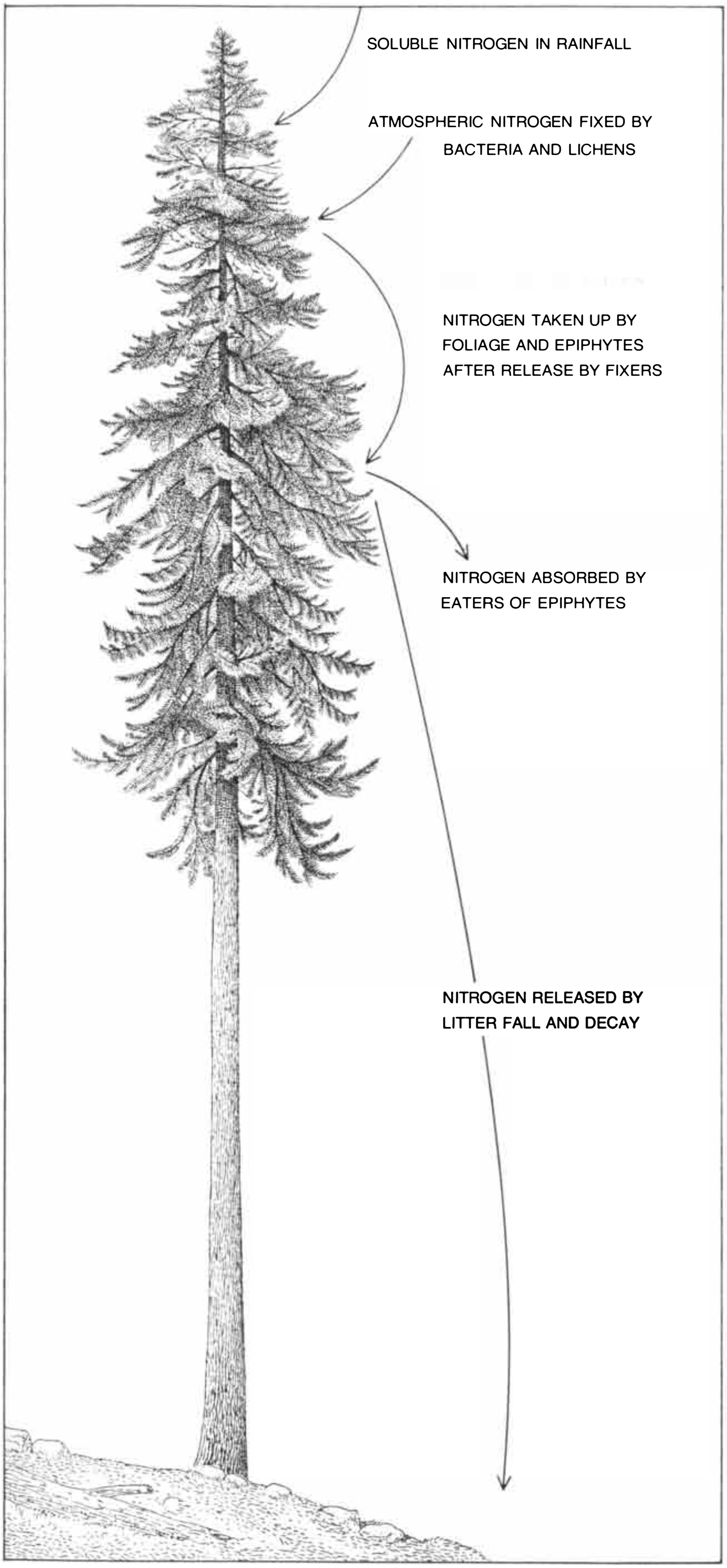
In an old-growth Douglas-fir forest the supply of available nitrogen is not abundant. Relatively little of this essential element is brought in directly from the atmosphere by rainfall, and the forest floor is largely barren of nitrogen-fixing plant life. It seemed likely, therefore, that lichens growing in the canopy, some of which were known to fix nitrogen, might be important contributors to the forest's nitrogen economy.

Our notice was attracted particularly to one lichen, *Lobaria oregana,* that is by far the most abundant species in the treetops. The forest floor is littered with fallen pieces of its green, lettucelike thallus (plant body). It has been established that *Lobaria* fixes nitrogen from the atmosphere, presumably through the agency of a blue-green alga that is embodied in granular packets within the thallus. In order to evaluate the importance of this lichen in the forest's overall economy, we worked out estimates of its probable annual contribution to the nitrogen supply.

Sterling A. Russell of Oregon State University, who has investigated *Lobaria's* nitrogen-fixing productivity, estimated that the lichen fixes nitrogen at a maximum rate of about 50 nanomoles (billionths of a unit of molecular weight) per hour per gram of the lichen's fresh weight, which is equivalent to 200 nanomoles per gram (dry weight). We estimate that in our Douglas-fir forest the amount of *Lobaria* growing on the trees is between 350 and 450 pounds per acre. If we took the 450-pound figure and assumed that *Lobaria* fixes nitrogen at the maximum rate throughout the year, we would arrive at a figure of slightly more than 10 pounds per acre as this lichen's



SCHEMATIC MAP of a Douglas fir more than 180 feet high is based on climbers' studies. The fir had a forked top; the shorter fork, on the east side of the tree, is offset for clarity. The map shows only branches that project east or west and that are more than 1.6 inches in diameter. Solid lines indicate living branches, broken lines dead ones. The enclosed area accompanying a branch is proportional to the weight of its foliage. Where branch is less than minimum diameter only the foliage area is shown (*color*). Treetops were not mapped.



annual production of fixed nitrogen. It is unreasonable to suppose, however, that the lichen sustains the maximum rate the year round.

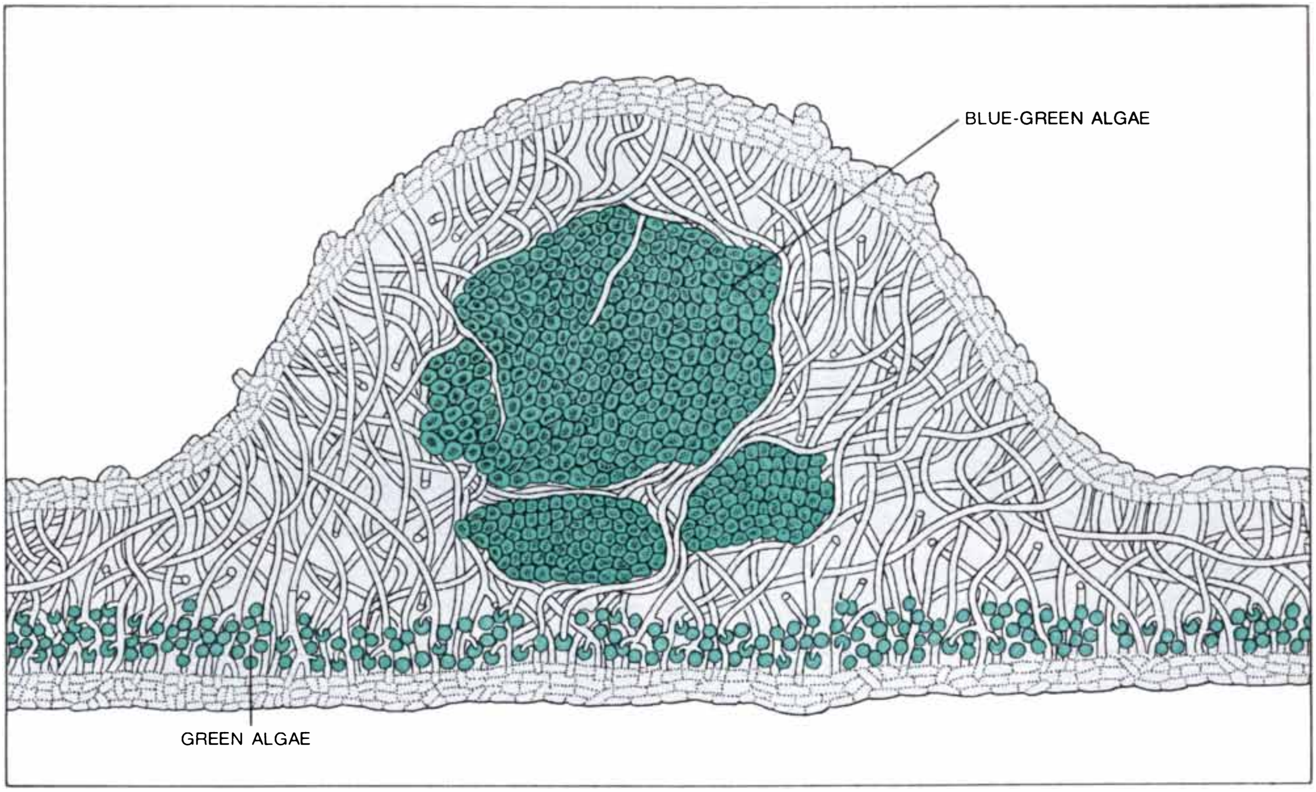
We decided to calculate a lower limit: the minimum amount of nitrogen *Lobaria* must fix to support its own growth. Its thallus, we estimate, adds about a fourth (in dry weight) in new growth each year. Taking the lower estimate of the amount of *Lobaria* in the forest (that is, 350 pounds per acre), the annual new growth would amount to 90 pounds per acre. Since the nitrogen content of the thallus is 2.1 percent of its total dry weight, the 90 pounds of yearly new growth per acre would contain roughly 1.8 pounds of nitrogen per acre.

Thus we conclude that *Lobaria oregana* contributes from 1.8 to 10 pounds of nitrogen per acre per year to the forest—certainly less than 10 pounds but probably substantially more than 1.8 pounds. The nitrogen trapped by the lichen is released in several ways for the eventual nourishment of the trees.

The chief route of the contribution is through *Lobaria*'s fall and decay. The annual fall of dislodged *Lobaria* from the canopy to the forest floor amounts to roughly 80 pounds per acre; most of this fall is peeled off by rain, snow and ice during the winter. Decomposing on the ground, the fallen thalli release about 1.8 pounds of nitrogen per acre per year to the roots of the trees and other plants.

Animals feeding on the lichens in the tree provide a second means of conveyance of the nitrogen. We have observed great numbers of invertebrates, including nematodes, mites and insects, eating away at *Lobaria* thalli. Certain vertebrates, such as the rodent called the red tree vole, supplement their diet with this lichen, among other plants. The animals consuming the lichen are

NITROGEN PATHWAYS in the forest canopy are indicated schematically. First (*top*) some nitrogen (less than one pound per acre annually) is washed into the canopy by rain. Far more nitrogen enters the system through the action of nitrogen-fixing bacteria on the fir needles and of blue-green algae in some lichens. The nitrogen-fixers pass the vital element along three pathways. Rainwater leaches some nitrogen from the living and dead tissue of the fixers; both the tree and its epiphytes absorb the rainwater. Fixed nitrogen follows a second path when herbivorous animals feed on epiphytes and then excrete. Finally, litter from the death and decay of both the epiphytes and the herbivores adds further nitrogen to the ecosystem.



PRINCIPAL NITROGEN-FIXER among the epiphytic lichens in the forest-canopy community, *Lobaria oregana*, is shown in cross section. Like all lichens, *Lobaria* is a symbiotic association of a fungus (*gray areas*) and two algae (*colored areas*). One of the Chlorophyta, or green algae, is the principal symbiont; it lives in the *Lobaria* thallus (*light color*). Populations of a blue-green alga, *Nostoc*, are also present (*solid color*) in bulges called cephalodiums. *Nostoc*, one of the Cyanophyta, fixes atmospheric nitrogen at a significant rate. As a result *Lobaria* contributes from 1.8 to 10 pounds of nitrogen per acre to the fir-forest ecosystem annually.

fed on in turn by predators, and the nitrogen is transferred to the forest soil through the predators' excreta.

Lobaria undoubtedly provides some transportable nitrogen even from its position high in the treetops. At its maximum rate of nitrogen-fixation it probably traps more than it uses for new growth, and this soluble excess of nitrogen may be leached from the thallus by rain and washed down the tree. The rain also picks up nitrogen from dead thalli decaying in the canopy. Rainwater flowing down over the branches and leaves loses nitrogen to mosses and other epiphytes that do not fix nitrogen, and perhaps some nitrogen is fed to the tree itself through its foliage. In any event, the epiphytes that take up nitrogen from the rainwater eventually release it to the ecosystem through their decay. We estimate that in our old-growth Douglas forest the contribution of nitrogen to the soil by epiphytes that do not fix the element amounts to roughly 5.4 pounds per acre per year.

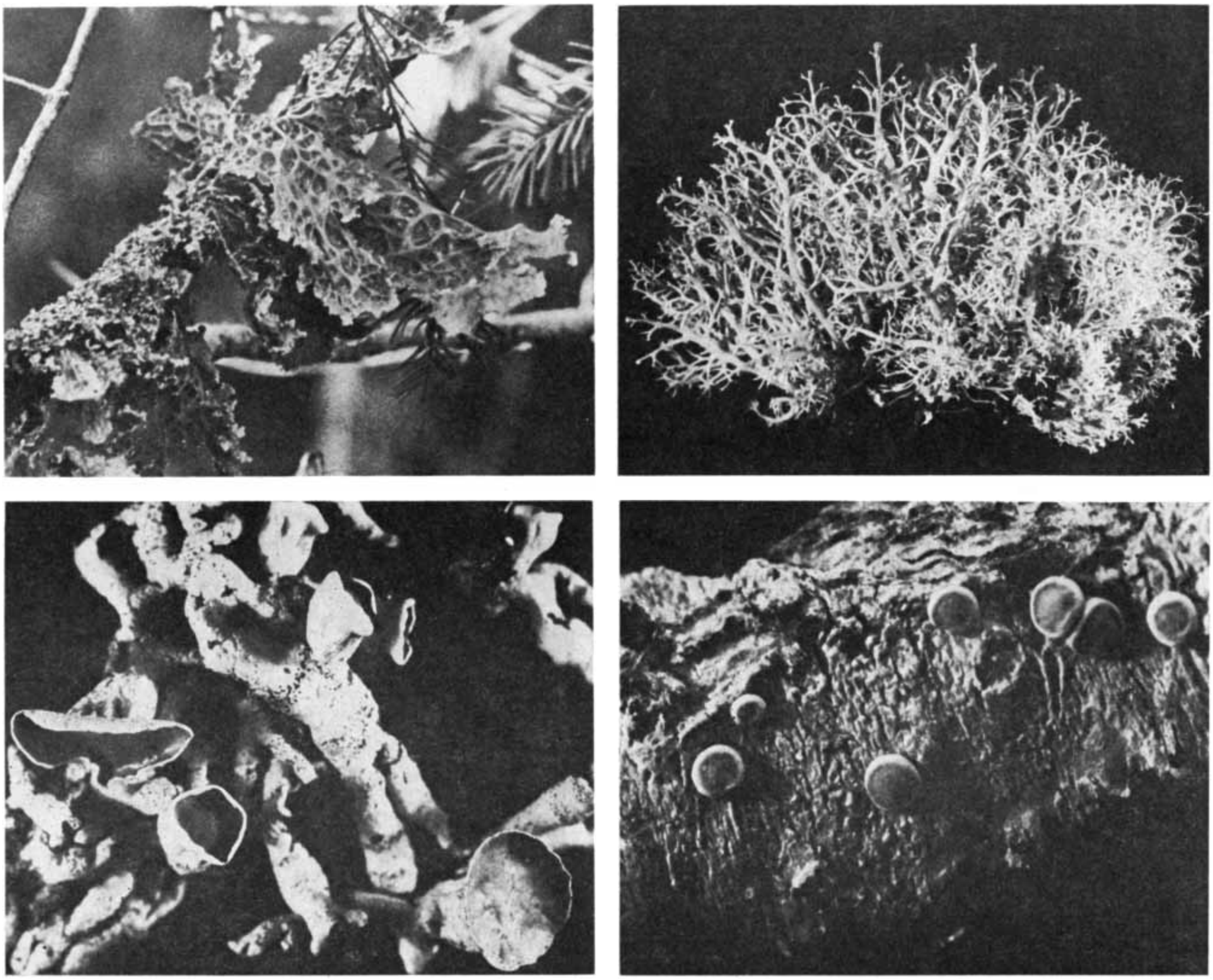
We do not yet have an estimate of the aggregate amount of nitrogen supplied to the forest by all the nitrogen-fixing epiphytes (of which *Lobaria* is certainly the most prolific). Probably the total is less than the 44 pounds per acre that is sometimes applied in the form of fertilizer to promote growth in Douglas-fir forests. Our survey indicates, however, that in an old-growth stand of Douglas fir the nitrogen-fixing plant life in the canopy can serve as the main pathway for the introduction of new nitrogen, an element required by all forest life-forms, small or large, plant or animal.

Epiphytic lichens such as *Lobaria* seem to be particularly susceptible to poisoning by pollutants in the atmosphere. In western Oregon these lichens are disappearing from forests as urbanization and industrialization of the land advance toward the woodlands. Here is further evidence that the atmospheric pollution that too often accompanies increasing density of population is inimical to nearby forested lands.

Our climbing around in the treetops so far can only be considered an early stage in the exploration of the microenvironment of trees. We hope to learn a great deal more about the communities of epiphytes living in the trees of our Douglas-fir forest and about how they are related to various tree environments.

The question of the differences in microclimates is particularly interesting. The climate at the top of the canopy is obviously very different from the climate low in the tree, so that an epiphyte high in the canopy is subjected to greater intensities of light and sharper fluctuations of temperature and humidity than one on a low branch. In the top of the tree lichens dry out within a few minutes after a rain, whereas those on the lowest branches may stay damp for months in the season of intermittent rains. As we have seen, there are marked climatic differences even between the upper and the lower sides of a branch, with resulting differences in the epiphytic communities of the two sides. Climatic differences within the tree also affect the growth of the branches themselves; for example, branches in the deep shade have a tendency to prune themselves. By installing meteorological instruments at various points in the tree we expect to obtain more specific information relating the growth of epiphytic communities and individual branches to factors in their immediate environment.

The trees to which we have given the



FOUR LICHENS commonly found on tree trunks and branches in fir forests are the nitrogen-fixing *Lobaria* (*top left*), a highly ramified species, *Sphaerophorus globosus* (*top right*), a species with cuplike fruiting bodies, *Hypogymnia enteromorpha* (*bottom left*), and a buttonlike species, *Ochrolechia oregona* (*bottom right*). Lichens are easily poisoned by urban and industrial air pollutants.

most study are the Douglas fir and the western hemlock. Viewed from a distance, both have about the same general shape: a slender cone consisting of relatively short branches extended laterally from a massive central trunk. Closer up, however, they are seen to have substantially different builds. An old-growth Douglas fir has relatively few branches and they are widely spaced, sometimes with gaps of up to 70 feet between branches on the shaded side of the trunk. Each branch is a complex system; often it is fan-shaped with a wide spread, and in many cases the system is a group of two or more young branches that have grown out from places where the original main branch broke off. In contrast, an old-growth western hemlock has more branches per length of trunk even in heavy shade; the branches are more evenly spaced up the tree, and they have most of their foliage concentrated at the end of the branch.

In his book *The Adaptive Geometry of Trees* Henry S. Horn has observed that the form of a tree can often be related to the tree's ability to take hold and thrive under specific conditions. One shape, for example, is characteristic of trees that spring up as the early settlers in place of a forest that has been felled by fire or clearing; other shapes allow the trees to invade an established forest at later successional stages. Examining our Douglas firs and hemlocks in this light, we find that the widely spaced, fan-shaped branches of the Douglas fir (forming what Horn describes as a "multilayer") fit his description of trees of the early successional type, and the western hemlock, with its evenly distributed foliage (forming a "monolayer"), corresponds to the late successional type. The old branches of a large tree show a pattern of developmental response to environmental factors that seems to be characteristic of its species and therefore probably is genetically influenced. The further exploration of the structure of big trees should provide useful information about the adaptive capacities of individual species.

I hope this brief description of our explorations in the treetops will encourage other investigators to extend the exploration to other forests and other types of trees. The investigation of living things in the forest canopy probably will also prove to be as attractive and rewarding to zoologists as it has been to us botanists. The study of small arboreal animals at home in the treetops should provide information that cannot be obtained by observing them in captivity.

The Human Lymphocyte as an Experimental Animal

Maintained through many generations in laboratory cultures, the cells that make antibody lend themselves to studies of cell differentiation and diseases involving the immune system, which may include cancer

by Richard A. Lerner and Frank J. Dixon

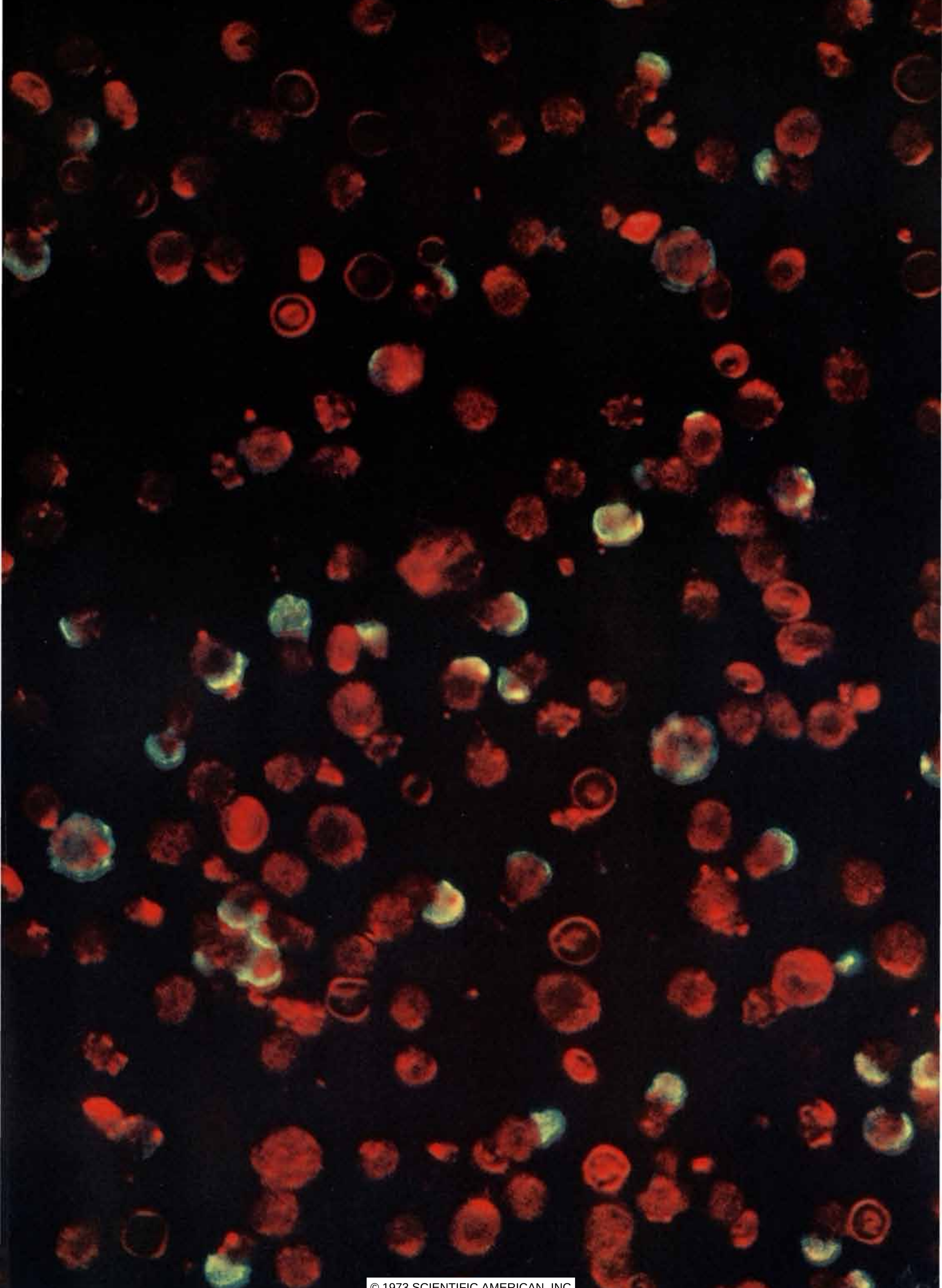
Some of the major questions in cellular biology today are these: How does a cell respond to its environment? How does that response alter the expression of the cell's genes? How do cells communicate with one another? How do cells differentiate? How do the specialized functions of a cell (such as secretion) relate to its general functions (such as gene replication)? There is a single class of cells that lends itself to the study of all these major questions. It is the lymphocyte, the cell that stands at the center of the immune response, the process by which an organism responds to its environment. The immune response is not only a defense against disease but also a cause of some diseases. Moreover, it seems increasingly clear that it is implicated in the chain of events that leads to cancer. Lymphocytes can now be cultured, or grown in laboratory glassware. Because of their special properties and the ease with which they can be manipulated, cultured lymphocytes are being exploited in a variety of investigations that are under way in a number of laboratories.

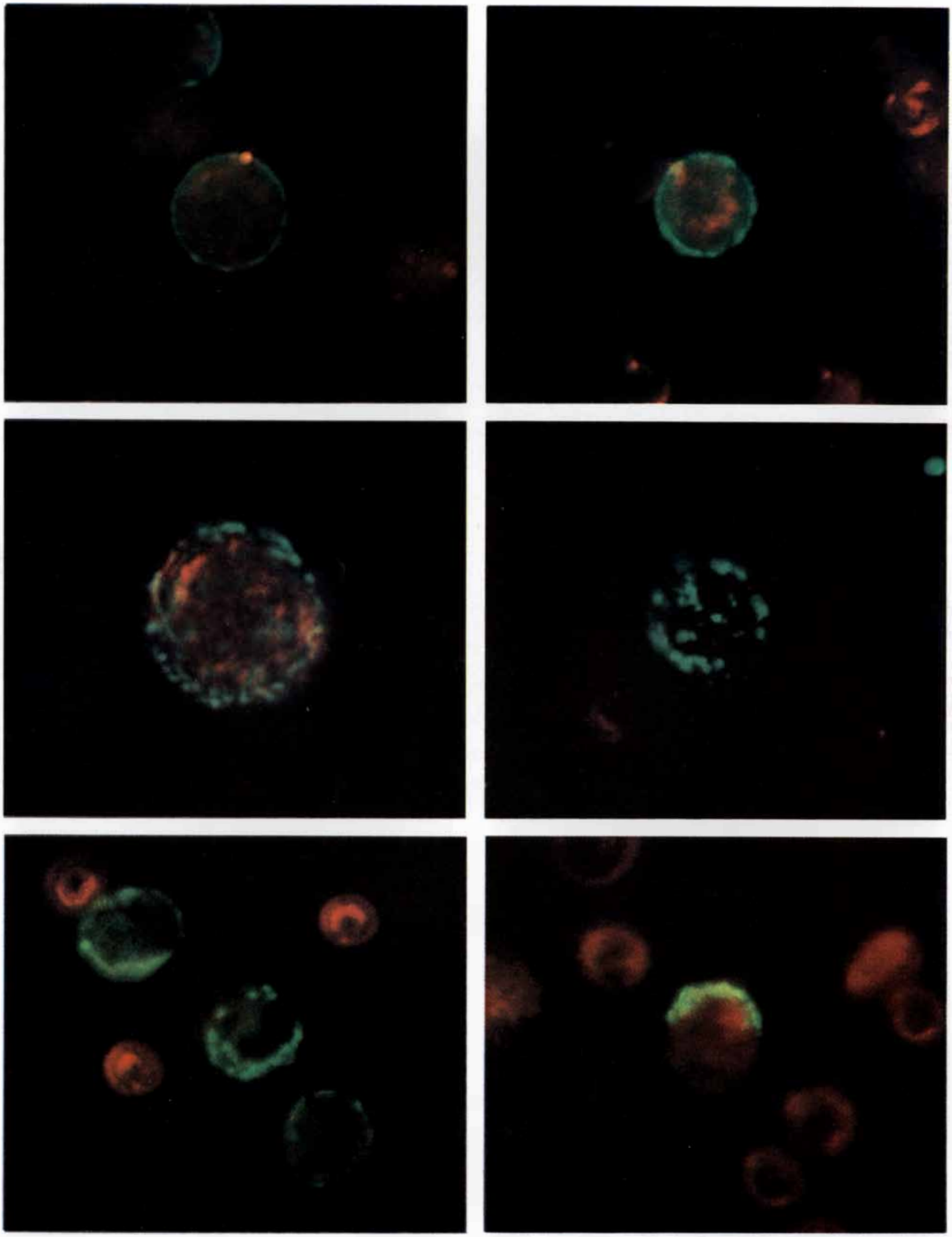
One way of characterizing these investigations is to say that they deal with the specialized functions of cells rather than the general functions (or, as Boris Ephrussi puts it, the luxury functions as opposed to the housekeeping ones). The housekeeping functions are those needed for the cell to stay alive and reproduce itself: energy transfer, the flow of genetic information from DNA through RNA to protein, the formation of cellular organelles. These functions are common to most cells. The various specialized functions, on the other hand, are peculiar to certain classes of cells and are reflected in the different shapes and duties of, say, brain cells and muscle cells, and the different secretions of, say, pancreatic cells and lymphocytes. When cells are maintained outside the organism in a culture, they can dispense with their specialized functions and still live, grow and divide. Indeed, it has been argued that a cell that can dispense with its luxury functions has a selective advantage over one that cannot, since it can put more of its energy into the housekeeping chores that are necessary for its survival. The differential expression of luxury functions in culture is illustrated by the human melanoma, a tumor that synthesizes the pigment melanin. When melanoma cells maintained in a culture are proliferating rapidly, no pigment is synthesized and the cells appear colorless. When growth ceases, however, the dark pigment granules are synthesized and the cells turn black.

Since about 30 years ago, when biologists began in earnest to remove cells from higher organisms and maintain them in a culture, a great deal has been discovered about what the housekeeping functions of cells are and how they are carried out. For example, we now understand rather well the structure of the genetic material, DNA, how it is transcribed into RNA and how RNA is translated into protein. It is time now to apply the rigorous procedures established over the past 30 years to the study of the luxury functions of cells. That will not be an easy task nor will it soon be accomplished, but it is the route toward a better understanding of the questions we listed at the beginning of this article.

Those questions have in common a relation to one of the central issues in biology: the problem of how events at the surface of a cell alter the expression of the cell's genes. The essential initial step in the immune response is the contact of an antigen—a foreign substance—with receptors on the surfaces of lymphocytes derived from the bone marrow, the so-called *B*-lymphocytes. (This contact may be altered by interaction of the antigen with the white cells called macrophages or with lymphocytes derived from the thymus gland, the so-called *T*-lymphocytes.) The contact alters the expression of the *B*-lymphocyte's genes, setting in motion a series of events during which the cells divide and differentiate [*see illustration on page 86*]. The cell type evolves from one that is

GREEN FLUORESCENT DYE reveals immunoglobulin molecules on the surface of lymphocytes (*opposite page*). An antibody against the immunoglobulin molecule is prepared by injecting some immunoglobulin from one species into an experimental animal of a different species. The antibody is coupled to fluorescein, a fluorescent dye, and is allowed to react with the cells. The fluorescein-labeled antibody binds specifically to immunoglobulin molecules on the cell surfaces (which in this case act as antigens). The resulting antigen-antibody complexes are revealed by the yellow-green glow emitted by the fluorescein under ultraviolet radiation. This micrograph and the ones on the next page were made by Curtis Wilson with a Zeiss dark-field microscope, mercury-arc lighting and special interference and barrier filters, one effect of which is to make unstained cells or parts of cells appear red.





IMMUNOGLOBULIN MOLECULES associated with the cell surface and visualized by fluorescent staining move over the surface of the cell. At first, thinly distributed over the surface, they appear as thin rims at the circumference (*top*). Then they become more concentrated, giving a granular appearance (*middle*). Finally they seem to collect in a restricted area, forming increasingly concentrated "caps" (*bottom*). The movement may represent a sweeping action in which complexes are removed from the surface. The way they move may lend support to the "fluid mosaic" model of the structure of the surface membrane, which sees the membrane as a fluid in which molecules are mobile. If energy supply is blocked (by low temperature, for example), only rim fluorescence is seen.

specialized to respond to an event in its environment to one that is uniquely adapted to the synthesis and secretion of antibody molecules. These two cells, the undifferentiated small, or "virgin," lymphocyte and the plasma cell, represent respectively the beginning and the end of the process of lymphocyte differentiation. Just how many times the cells divide in the course of this differentiation is not known, but many of the changes have been documented. There is a progressive decrease in the number of immunoglobulin molecules, or cell-surface receptors, on each cell. The irregular nucleus of the small lymphocyte becomes regular and round, and the nucleus becomes much smaller in relation to the surrounding cytoplasm. In the cytoplasm there is an increase in the number of such organelles as lysosomes, the Golgi apparatus and mitochondria. The organelles called polyribosomes increase in number and become associated with membranes to form an extensive network of endoplasmic reticulum, which is presumably involved in the greatly augmented synthesis and secretion of immunoglobulin molecules.

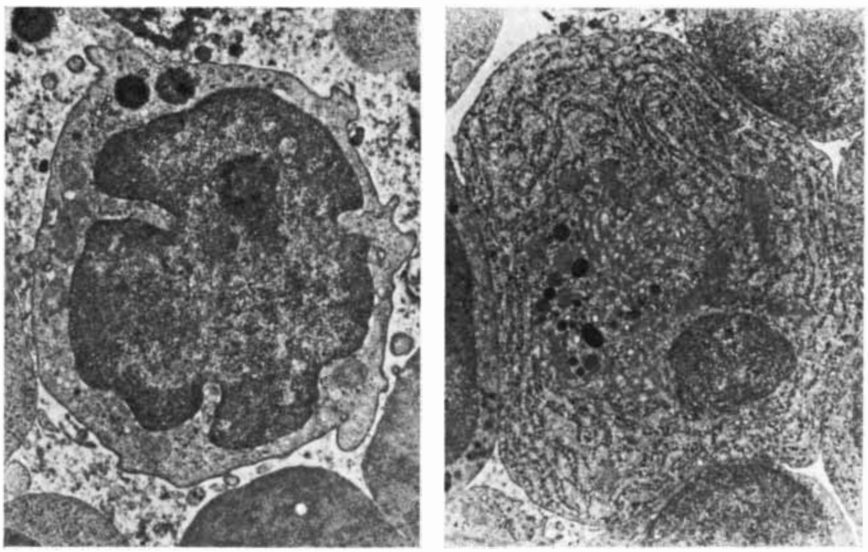
It would be convenient to be able to induce a culture of lymphocytes to enter the process of differentiation and then observe them as they move through the successive stages. Unfortunately this is not yet possible, but such an experimental system can in effect be simulated. Approximately 100 different lymphocyte clones, or cell lines descended from a single cell, are available. They cover the spectrum of differentiation, since the lymphocytes in any clone appear to be arrested at one point in the process of differentiation. By selecting the proper clone the investigator can therefore work with lymphocytes that represent a particular stage in the process of differentiation.

Lymphocytes in culture have a number of properties that suit them uniquely to the study of the special functions of cells and thus to the process of differentiation. One such property is their exquisite specificity. A single cell is capable of responding to only one antigen or at most to a limited number of antigens with a similar molecular structure. That means the cell-surface receptors for the antigens must be highly specific. What is the nature of the receptors and how many different ones must there be in order to accommodate the different antigens an organism must deal with?

Niels K. Jerne of the Basel Institute for Immunology suggested (in a modification of Paul Ehrlich's original theory of antibody response) that the remarkable specificity of the immune response could best be accounted for if the receptor for antigen were the antibody molecule itself, in which case the antibody would be both the receptor and the effector. The immunoglobulin, or antibody, molecule is well adapted to serving such a double function [see "The Structure and Function of Antibodies," by Gerald M. Edelman; SCIENTIFIC AMERICAN, August, 1970]. In its simplest form it is a protein composed of two heavy and two light polypeptide chains that are joined by several disulfide bridges. Each chain is divided into two kinds of regions: "constant" regions, in which the sequence of amino acids (the constituent units of the polypeptides) is essentially the same in all antibodies of a given class, and "variable" regions, in which the sequence is different in every antibody [*see illustration on page 87*]. The variable regions provide the specificity required for the receptor function: the binding of antigen. The constant regions are involved in such effector functions as the binding of complement or fixation to skin (an element in certain inflammation processes).

Studies in several laboratories have indicated that Jerne's hypothesis is correct, at least in its broad outline. At the Scripps Clinic and Research Foundation we have been able to measure the surface receptors in quantitative terms by developing a sensitive radioimmune assay that determines the number of molecules of immunoglobulin exposed on the surface of cultured lymphocytes [*see illustration on page 88*]. The number varies between 20,000 and 200,000 in different lines. As lymphocytes proceed along the line of differentiation the number of antibody molecules on the surface decreases; several studies have suggested that most plasma cells have few receptor molecules, if any at all. This is an interesting case of biological modulation. The plasma cell is in effect the end of the line: it is specialized for a high rate of synthesis and secretion of immunoglobulin molecules and has no need for receptor molecules.

Immunoglobulin molecules associated with the cell surface—the presumed receptor molecules—can be visualized by the technique of fluorescence microscopy. Antibody is prepared in one species against the immunoglobulin molecules of the species supplying the lymphocytes to be studied. The antibody is coupled to a fluorescent dye such as fluorescein. When the complex of the antibody and the dye reacts with cells, it binds to the immunoglobulin molecules associated with the cell surface and, when the cells are exposed to ultraviolet radiation, the characteristic green glow of the irradiated fluorescein reveals the location on the cell surface of the immunoglobulin molecules [*see illustrations on page 83 and on opposite page*]. The surface-associated molecules can be isolated and characterized by a combination of chemical and physical procedures. The exact structure of these molecules is still under study, but the available evidence suggests that

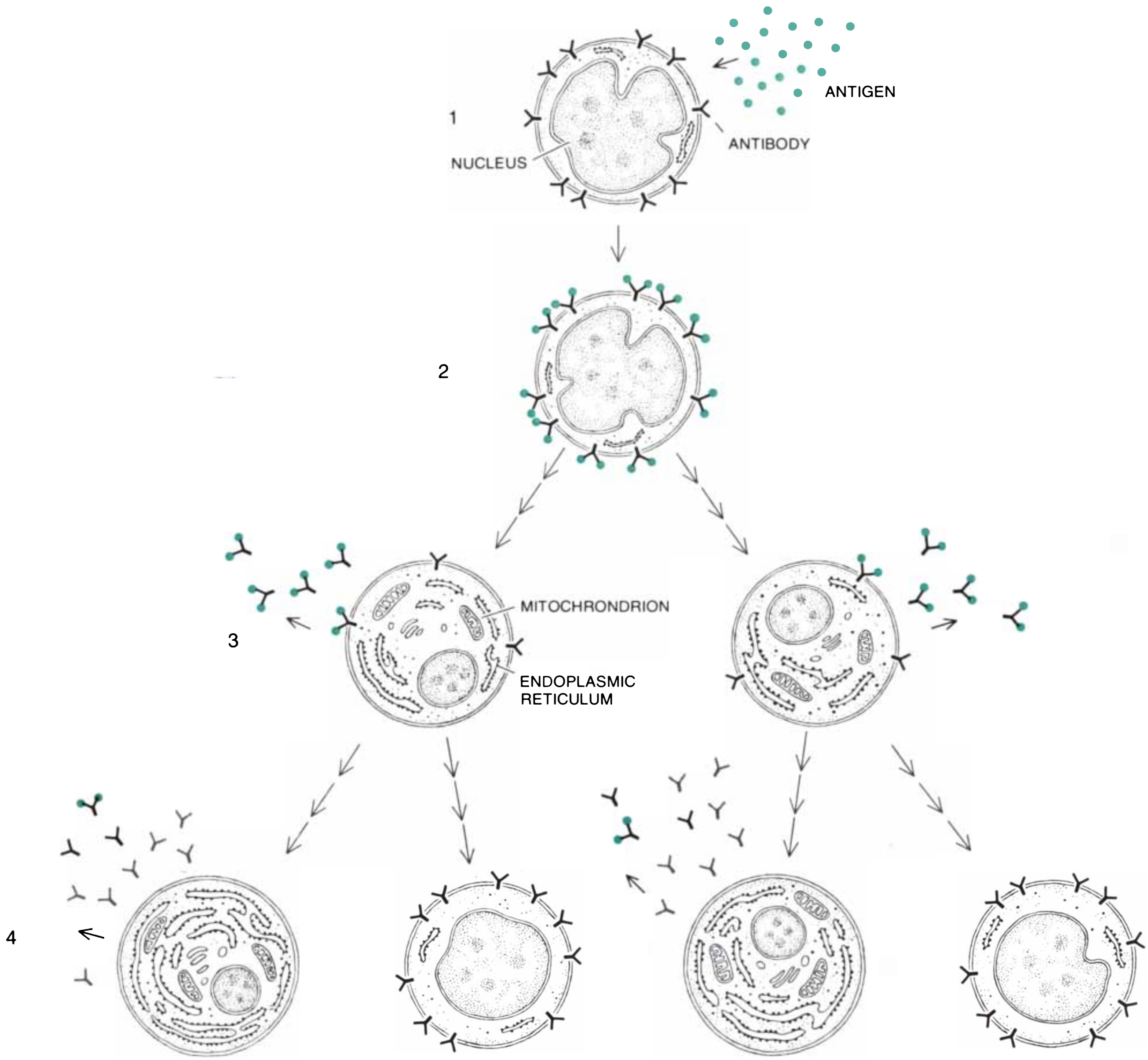


DIFFERENTIATION in the course of the immune response changes a lymphocyte (*left*) into a plasma cell (*right*). Noticeable changes include an increase in cell size, a relative decrease in the size of the nucleus and the formation of organelles involved in synthesis and secretion of antibody, including ribosomes (*small black dots*) arrayed along the membranous endoplasmic reticulum. The electron micrographs were made by Joseph Feldman.

the surface molecules may differ in some respect from the immunoglobulin that is secreted to become circulating antibody.

One of the characteristics of normal eukaryotic cells (the cells of organisms other than bacteria) is that they synthesize, grow and divide on definite cyclic schedules unless they retire from the cell cycle into a resting phase. Cell biologists conventionally divide the cycle into four phases called G_1, S, G_2 and M; the first three are defined on the basis of the different molecules synthesized in each of them and the fourth phase is mitosis, or cell division [*see illustration on page 89*]. The number of temporal segments into which the cycle is thus divided is arbitrary; the degree of detail is limited by incomplete knowledge of the regulation of synthesis in these cells. It is actually likely that cells regulate their activities minute by minute or even second by second, and as we learn more it should be possible to fill in the minutes and seconds on the cell-cycle clock.

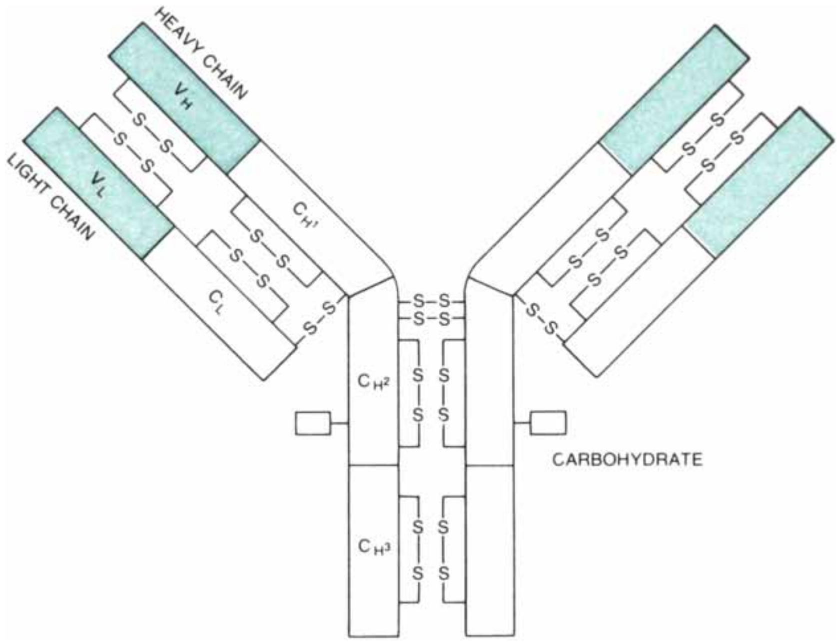
Studies in our laboratory and in that of David Pressman at the Roswell Park Memorial Institute and of Donald N. Buell and John L. Fahey at the University of California at Los Angeles School of Medicine have shown that secretory immunoglobulin is synthesized in lymphocytes during a limited part of the G_1 phase, as is the case for other luxury polypeptides in other cells. In contrast, the surface immunoglobulin receptor molecules appear to turn over throughout the cell cycle with a half-life of about 45 minutes. The reason for this rapid turnover of cell-surface immunoglobulin is not yet clear, but it may well be relat-



IMMUNE RESPONSE may begin when a small lymphocyte with antibodies (*black Y's*) on its surface encounters antigen molecules (*color*) against which its antibody is specific (*1*). The antigen is bound to the antibody (*2*); this event, together with a second signal involving another kind of lymphocyte, induces the small lymphocyte to divide and differentiate (*3*); surface receptor antibodies are lost and replaced by new ones. The lymphocytes differentiate to become plasma cells and memory cells (*4*). Plasma cells synthesize and secrete circulating antibody (*gray Y's*); memory cells, held in reserve, respond sensitively to reappearance of the antigen.

ed to modulation of the immune response, that is, to the requirement that immunoglobulin that has been "hit" by antigen not remain on the surface and that there be enough new receptor immunoglobulin on the surface to respond sensitively to any new hits by new antigen molecules.

Just as there is a temporal order for the synthesis of proteins during the mammalian cell cycle, so there are apparently limited periods during which the synthesis of enzymes can be induced. In liver cells, for example, David W. Martin, Jr., Gordon M. Tomkins and Daryl Granner of the University of California at San Francisco have shown that the enzyme tyrosine aminotransferase can be induced by corticosteroid hormone only during the last two-thirds of G_1 and at any time in S. It seems likely that we shall find a limited period of the lymphocyte cycle during which an antigen can induce cellular proliferation. The role of the antigen would therefore be not directly to trigger the synthesis of immunoglobulin but rather to induce specific resting lymphocytes to enter the cell cycle and so pass through a phase in which such synthesis is obligatory.



ANTIBODY MOLECULE is a protein composed of four polypeptide chains, two long "heavy" chains and two shorter "light" ones. Each chain has a variable region (*color*), whose amino acid sequence provides the specificity for binding antigen, and one or more constant regions. The four variable regions are homologous (their amino acid sequences match up), as are the eight constant regions. The chains are connected by disulfide bridges.

David Prescott of the University of Colorado has noted that when nerve cells and striated-muscle cells in living animals retire from the cell cycle, they do so during the G_1 phase; he has suggested that for most cells the control of further proliferation and differentiation is mediated by some event during this phase. An example of such an event occurring in cultured cells was reported by Karlin Nilausen and Howard Green of the Massachusetts Institute of Technology. They found that when certain mouse cells ceased to grow because of contact inhibition (that is, because they had formed a complete single layer in the culture dish), they did so during the G_1 phase. Moreover, some observations made by H. Becker of the Ontario Cancer Institute and his colleagues indicate that there are physiological differences between cells in the resting phase and cells in the G_1 phase. All of this tends to suggest that in the case of lymphocytes contact with antigen molecules may be the inducing event that initiates entry into the G_1 phase and subsequent cellular proliferation.

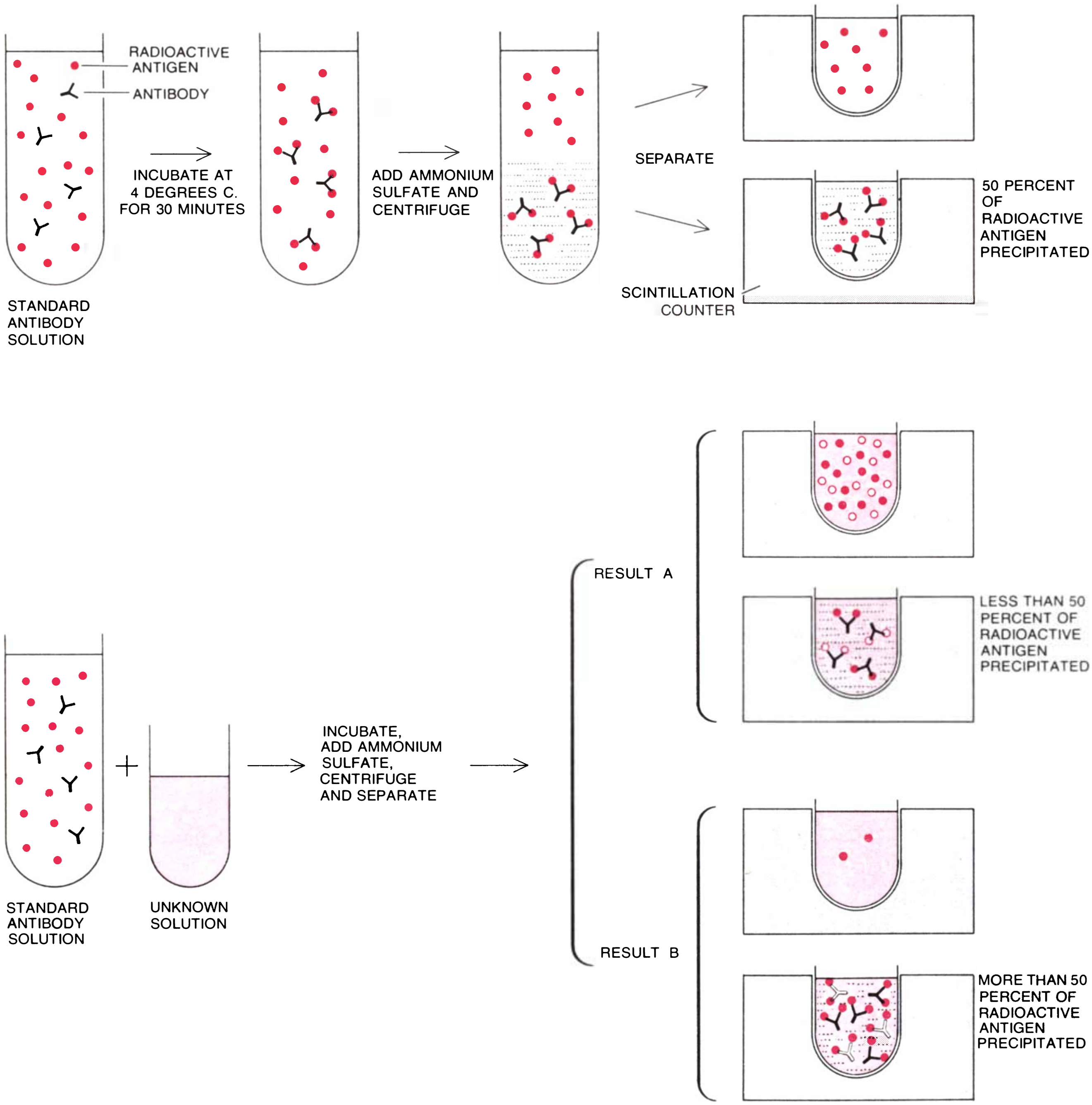
The rapid increase in antibody formation after an antigenic challenge is hard to explain simply on the basis of an increase in the number of cells that form antibody. Some investigators have proposed that additional cells are "recruited," and that in these cells the generation time is shortened, so that the cells divide with particular rapidity. Our data show, however, that as cells proceed through the G_1 phase the synthesis of immunoglobulin can increase as much as fivefold without cell division or even the synthesis of DNA. Recruitment, then, can be explained as the passage of an increasing number of cells to a point in the cell cycle where the rate of antibody synthesis is greatly increased. This contention is supported by the finding that during the induction of an immune response neither DNA synthesis nor mitosis is a necessary predecessor of antibody synthesis.

Like other cells in culture, lymphocytes can be used for studies of mutation and other kinds of variation. In this application lymphocytes have a number of advantages. They apparently have an infinite life-span in culture, during which their phenotypic expression persists, that is, their shape, physiological functions and other characteristics remain constant. They retain the normal diploid (double) set of chromosomes, whereas in most other cultured cells the chromosome complement changes. And they are easily cloned: a single cell with specific properties can be isolated from a culture and its progeny can be grown into a large population.

A number of inherited disorders afflicting humans are caused by inborn errors of metabolism such that the patient's cells fail to synthesize a necessary enzyme or some other kind of product. By obtaining and culturing lymphocytes from such patients one can track down the defect, quantify it and perhaps in time find ways to remedy it. Arthur D. Bloom and his colleagues at the University of Michigan Medical School have examined a number of such defects in this way. One was the Lesch-Nyhan syndrome, a serious defect in a metabolic pathway that converts preformed nucleosides (DNA precursors) into DNA. Bloom's group found that lymphocytes from these patients produce only 1.4 percent of the normal amount of HGPRT, a key enzyme in the pathway. Production defects of this kind may someday be corrected by manipulating or replacing the defective cells.

One way of manipulating would be through cell fusion: if two lines of cells are cultured together, some cells may fuse and form hybrid daughter cells that contain chromosomes, and thus genetic information, from both parent cells. It may be possible to remove cells from a patient, fuse them to cells that contain the gene for the missing enzyme and return the "corrected" cells to the patient. Henry Harris of the University of Oxford has shown that (at least in one system) fused cells may incorporate some foreign genetic material without bringing about the synthesis of foreign surface antigens, so that apparently one need not be concerned that the fused cells would be rejected.

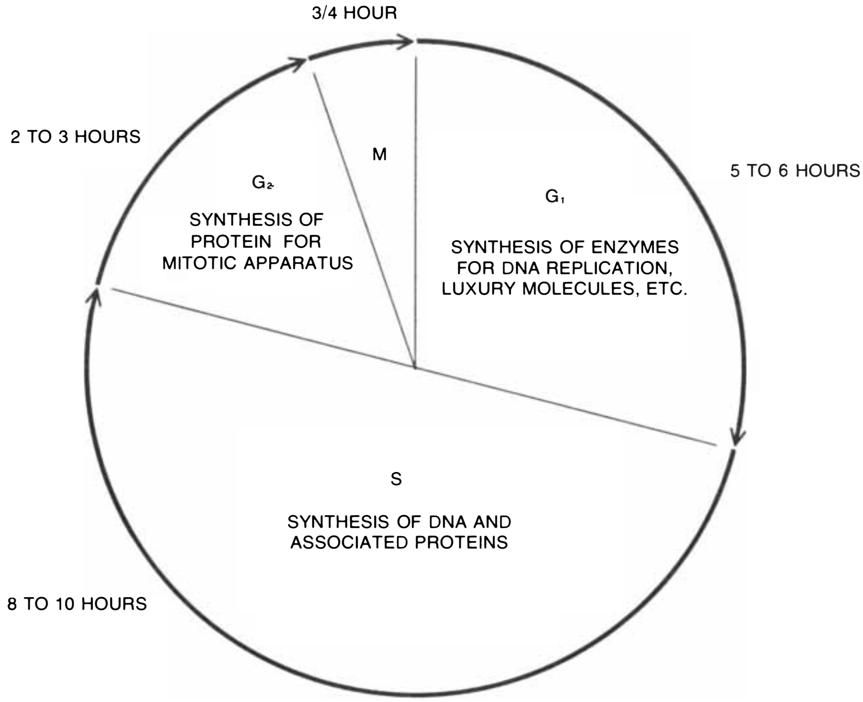
Cells that start out with the same genotype, or hereditary constitution, may come in time to have subtly different phenotypes. The immunoglobulins are ideally suited to the study of lymphocyte phenotypes because they are so readily detected, characterized and quantified by sensitive immunoassays. It has long been known that when cells from mouse plasma-cell tumors are cultured, some of the cells spontaneously lose the ability to synthesize the heavy-chain parts of the immunoglobulin molecule. (There is a much smaller tendency to lose the ability to synthesize light chains, presumably because any heavy



RADIOIMMUNE ASSAY determines the number of antigen or antibody molecules in a solution or on the surface of cells. It utilizes a standard solution (*top*) of antibody molecules (*black Y's*) and radioactively labeled antigen molecules (*colored dots*). The proportions are such that the antibody can bind 50 percent of the antigen. When ammonium sulfate is added and the preparation is centrifuged, the antibody and the bound antigen are precipitated, and measuring the amount of radiation in the precipitate and the remaining solution (*right*) shows how much of the antigen was bound. If now a solution to be tested is added to the standard solution (*bottom*) and the procedure is repeated, the assay shows whether there is antigen or antibody in the test solution, and how much. If there is unlabeled antigen (*open colored circles*) in the test solution (*Result A*), it will compete for binding sites on the antibody, so that less than 50 percent of the labeled antigen is precipitated (*right*). If there is antibody (*open Y's*) in the test solution (*Result B*), it will provide extra binding sites for antigen, so that more than 50 percent of the labeled antigen is precipitated.

chains without light chains are quite insoluble and are precipitated within the cell and destroy it; the loss of light-chain synthesis may be a lethal mutation and one that is therefore not seen.)

At the Albert Einstein College of Medicine, Matthew D. Scharff and his colleagues have devised an ingenious method of measuring the rate of loss of heavy-chain synthesis in individual mouse plasma cells [*see top illustration on next page*]. They have established that cells lose the ability to synthesize heavy chains at the rate of one loss per 900 cell generations. The event that causes this loss is not yet known, and it will be hard to track down because the appearance of immunoglobulin outside the cell entails the completion of a complex process involving transcription, translation, assembly, transport and secretion of the molecule; an alteration in any molecule or structure that plays a role in this process could cause the failure of secretion that Scharff detects.



CELL CYCLE is conventionally divided into four phases on the basis of molecules synthesized in each. The times given here are for the HeLa cell, a human tumor cell. The cycle begins with G_1 and ends with M (mitosis), when the cell divides to form two daughter cells.

The cause may or may not be a simple mutation, in which the substitution of a single amino acid subunit can alter the structure and/or the configuration of a single polypeptide and thus lead to a change in phenotype. Scharff has therefore called the rate of loss he measures a variation frequency rather than a mutation frequency. As a matter of fact many of the phenotypic properties of cells in higher organisms may involve the synthesis and secretion of polypeptides as the result of just such complicated processes. Although variation frequencies may not yield the precise information about the organization of genes that mutation frequencies provide, they may be more to the point in studies of the alteration of certain phenotypes in eukaryotic cells.

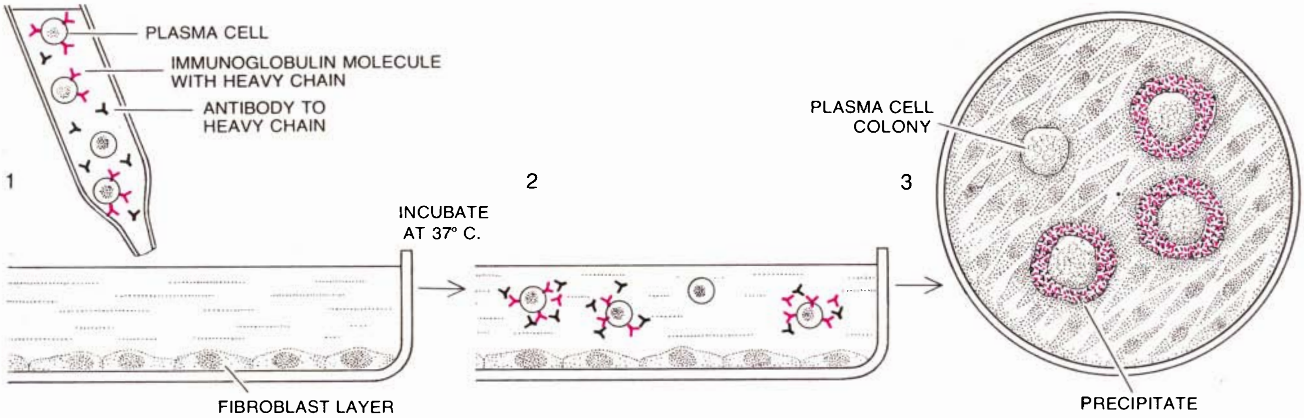
In the study of disease, lymphocytes have an important added advantage in that they are part of the system that generates the immune response and may therefore reflect abnormalities peculiar to that system. These include failures of the immune response, certain viral infections, autoimmune disease (in which the organism makes antibodies against components of itself) and perhaps cancer.

In the case of systemic lupus erythematosus (SLE), a serious autoimmune disease, the study of lymphocytes may have yielded important information not only about the disease itself but also about its relation to other conditions. In human SLE (and in a similar disease in the New Zealand strains of mice) the immune system synthesizes autoantibodies—often against the victim's own red blood cells and regularly against a variety of specific antigens found in the nuclei of the victim's cells. In many cases the most serious effects of these diseases have been shown to be caused by highly inflammatory antigen-antibody complexes in which at least some of the nuclear antigens are present.

Why, we wondered, do such patients and mice make antibodies almost exclusively to red blood cells and those particular nuclear antigens? What set of events would present those particular antigens to an affected mouse? One clue lay in the fact that many strains of mice harbor *C*-type RNA tumor viruses. Such viruses have RNA (rather than DNA) as their genetic material, they replicate by transcribing their genetic information into DNA (which can presumably integrate into the host's own genetic material) and they cause tumors in animals they infect [see "RNA-directed DNA Synthesis," by Howard M. Temin; Scientific American, January, 1972]. Could the mice be infected by a virus that was replicating and could the intermediates in that replication constitute the antigens in question? When we investigated, we found that the mice were in fact making antibodies that reacted with virtually every polypeptide or nucleic acid intermediate involved in the replication of the viruses [*see bottom illustration on next page*]. Was it possible that the immune system of the mice had been mobilized against an RNA tumor virus and that in fighting the war against the virus the mice had developed inflammatory antigen-antibody complexes and autoimmune disease?

An answer to that question might shed light not only on the origin of at least one autoimmune disease but also on the possible interrelation of viruses, the immune response and cancer. The connection between viruses and some animal tumors and leukemias has been known for many years. More recently it has come to seem likely that the immune system is somehow implicated in cancer. Since the affected mice had autoimmune disease and made antibodies to viral intermediates, it was obviously important to try to isolate the virus in question. And, for theoretical as well as technical reasons, it seemed to us that the cultured lymphocyte was a better place to find it (perhaps in large amounts) than the very occasional tumors these mice develop.

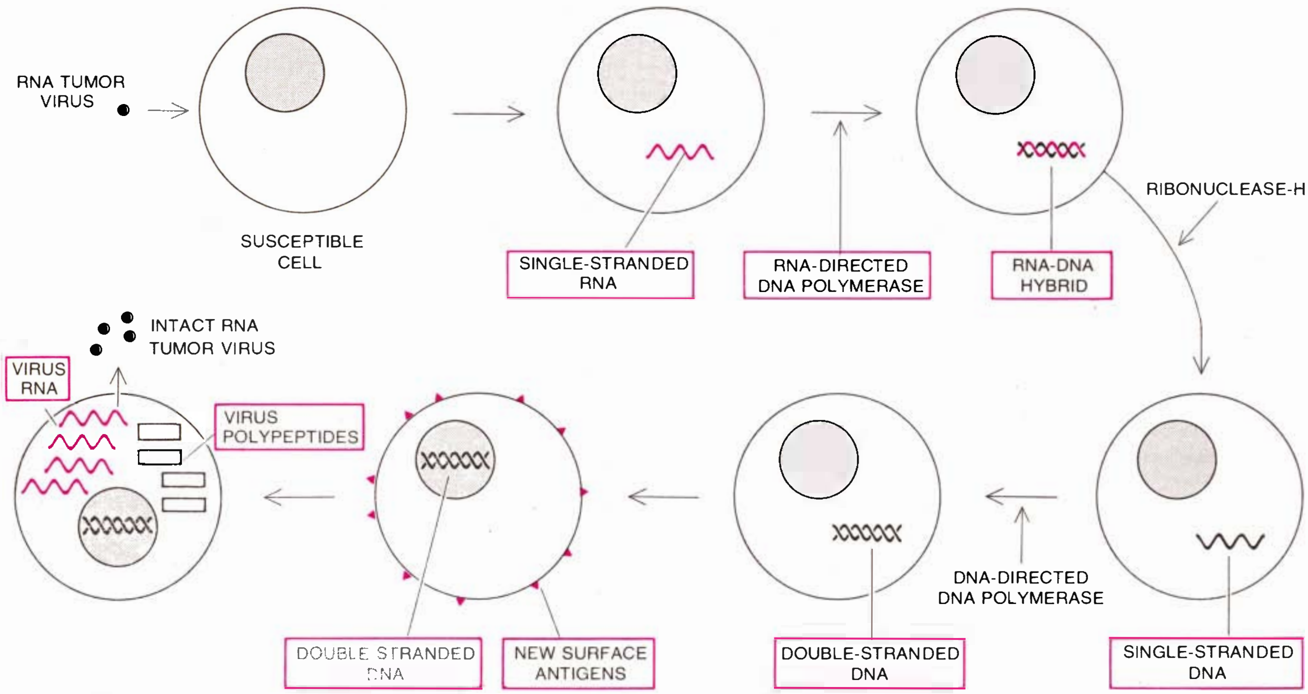
Let us outline the reasoning. Since a virus is an obligatory parasite that can only survive and multiply in living cells, a tumor is of no advantage to the virus; after all, it may kill the host and the virus as well. Certainly a tumor is of no advantage to the host. Moreover, a tu-



VARIANT CELLS that fail to synthesize the heavy chain of an immunoglobulin molecule are detected and cultured by a technique developed by Matthew D. Scharff of the Albert Einstein College of Medicine. The plasma cells to be tested are pipetted into a culture medium in which a "feeder" layer of fibroblast cells has been established (*1*). Plasma cells, most of which contain heavy chains, are introduced, along with antibody to heavy chains. The antibody (acting as antigen) becomes bound to the heavy chains, forming antigen-antibody lattices around the cells that contain heavy chains (*2*). After a few days each plasma cell has proliferated to form a colony of cells (*3*). The colonies descended from those of the original cells that produced heavy chains are surrounded by an antigen-antibody precipitate; the colony from the cell that did not make the chains has no precipitate. The variant colony is easy to distinguish and its cells can be cultured for further studies. (The third drawing is done at a much smaller scale than the first two.)

mor may not be a valid guide in the search for an oncogenic, or tumor-causing, agent. The reason is that the very formation of the tumor suggests that the host's defenses have been inadequate. The oncogenic agent (perhaps a virus) and the antigens it induces in tumor cells have failed to alert the host's defenses; they may well also evade the investigator.

How might the failure have come about? One possibility is that in a cancer that results from virus infection some process (an immune response?) causes a selection of cells with viral genomes, or gene complements, that are incompletely expressed—genomes in which only a few genes, whose products render the host cells cancerous, are expressed and other genes are not. From the point of view of selective advantage for the tumor this would be an ideal situation, since the unexpressed genes might have coded for more and/or stronger anti-



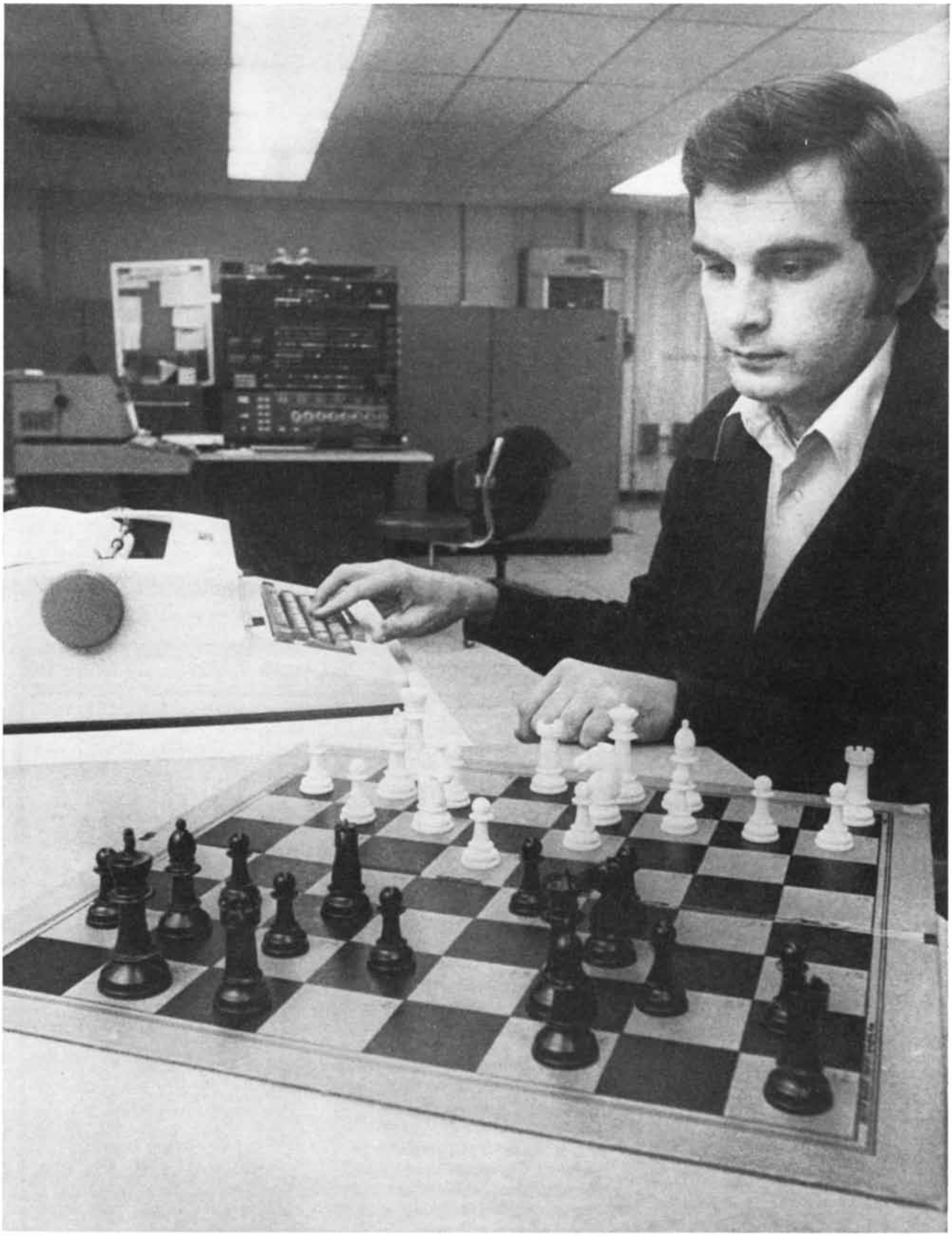
RELATION BETWEEN AN AUTOIMMUNE DISEASE and a tumor-inducing virus was suggested by the fact that human patients suffering from systemic lupus erythematosus (SLE) or mice with a similar disease make antibodies to a variety of antigens, almost all of which are related to similar molecules formed during the replication of an RNA tumor virus. The diagram shows the successive steps in the replication of such a virus, which is unusual in that its genetic information is carried in RNA that is transcribed into DNA in the cell by an RNA-directed DNA polymerase. This significantly broadens the range of foreign nucleic acids that is presented to the host. The molecules against which specific antibodies were found to be made by affected mice are indicated by the colored boxes.

genic polypeptides that would stimulate the host's defenses.

If this proposed chain of events should be correct, then tumors would not be the most profitable place to look for tumor viruses. It seems reasonable to assume that the growth of a tumor may be the rare and extreme result of the incomplete expression of the viral genome, and that the same virus or similar ones may be present in other disease states or even in apparently well individuals. In cancer, that is, the war has been lost and the virus has gone underground. In other situations the immunologic war may still be being fought against a detectable virus, and a disease (such as SLE) may develop as a result of the battle. We would suggest that the time and place to search for an oncogenic agent is when and where there is immunologic evidence that the host is defending against such an agent.

These concepts are predictive ones and they can be tested. Our prediction is that if a virus is isolated from the New Zealand mice, it should, on being injected into other mice, cause both autoimmunity and cancer. Together with Fred Jensen, we established a number of lines of cultured lymphocytes from New Zealand mice with the autoimmune disease. When we examined those lines, we found that each of them produced a unique RNA tumor virus of the *C* type. In addition, all cells in each of the lines had a distinctive extra chromosome, and some cells had a deletion—a missing segment—in one of the two X chromosomes. The extra chromosome was also found in eight of 38 cells taken from New Zealand mouse fetuses, indicating that the mice are "mosaics" for this chromosomal marker.

Byron Croker and Bert Del Villano isolated the virus produced by one of these lymphocyte lines and injected it into newborn mice of two New Zealand substrains. Their preliminary results suggest that the concepts we outlined above are correct, but many more mice and different combinations of strains remain to be tested before we can be sure that, through an immunologic reaction, autoimmunity is the price of protection against a tumor. The patient may lose either way, but the concept of such a trade-off is nevertheless significant. The immunologic tracks of autoimmunity may turn out to have the same predictive value for a putative (and hitherto elusive) human tumor virus as classical antigen-antibody tests have had for the identification and isolation of the agents of other diseases.



MAN V. MACHINE GAME OF CHESS is played between Charles I. Kalme, rated as a senior master, and the advice-taking chess machine, an IBM 370/155 programmed by the authors at the University of Southern California. Kalme, a member of the mathematics department at U.S.C., is the most highly ranked player engaged in the development of a computer-chess program except for Mikhail Botvinnik of the U.S.S.R. The U.S.C. program enables a chess expert with no previous computer experience to guide the machine to a more masterful game. Kalme's official rating is 2,445; machine currently plays at a sound novice level between 1,200 and 1,500.

AN ADVICE-TAKING CHESS COMPUTER

It has now been 24 years since Claude Shannon outlined how a computer could play chess. Here is a description of the first machine able to take lessons from a master

by Albert L. Zobrist and Frederic R. Carlson, Jr.

Last August, while Bobby Fischer was winning the world chess title from Boris Spassky in Iceland, several computers were quietly contending in Boston for the U.S. computer-chess championship. The winning machine, a Control Data Corporation 6400, was not actually present in Boston; its moves were relayed from Northwestern University in Evanston, Ill., where its program, known as Chess 3.5, had been written by David J. Slate, Lawrence R. Atkin and Keith Gorlen. Chess 3.5 is the fifth embodiment of a program that was initiated in 1968 and that in one version or another has played several thousand games on computers all over the world. In commenting on last year's final match in Boston, in which Chess 3.5 defeated a program written by James Gillogly of Carnegie-Mellon University, the former U.S. champion Samuel Reshevsky wrote: "The computer-chess match in Boston proved one thing: computers have a long way to go before they become international grand masters. But their game was an interesting experiment nonetheless."

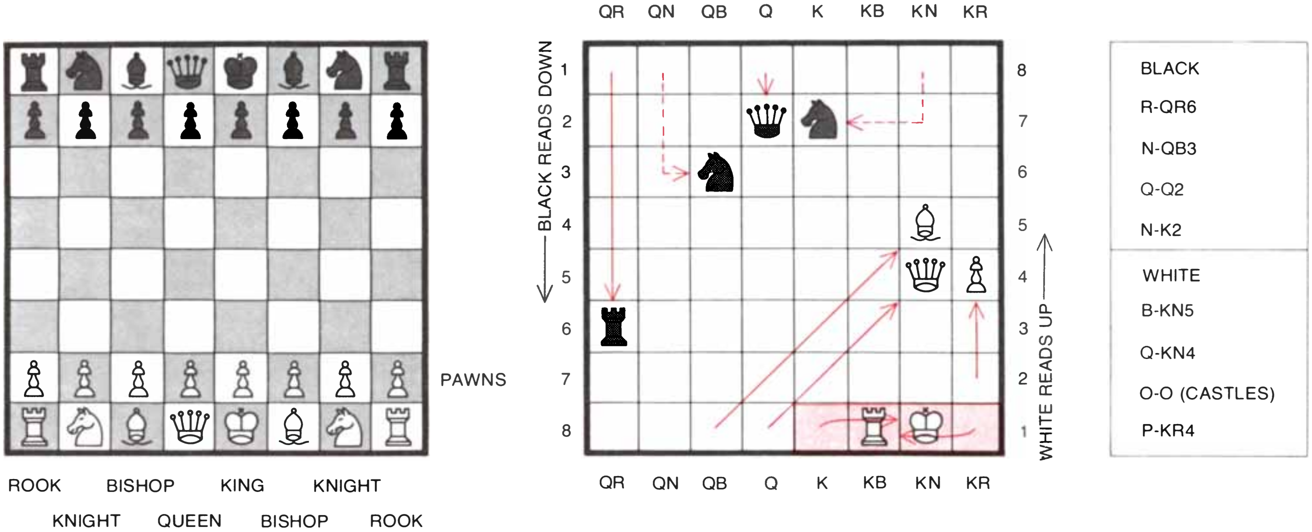
Among the victims of Chess 3.5 were programs written at Bell Laboratories and Columbia University as well as our own program, which we had begun developing at the University of Southern California less than a year earlier. Before the Boston tournament the U.S.C. program had played only about a dozen games and still had much to learn. We consider our program to be fundamentally different from all previous programs in that it is designed so that it can take advice. To help us in this endeavor we have persuaded a U.S.C. mathematician, Charles I. Kalme, to be our program's chess tutor. Kalme is a senior master with a rating of 2,455 on the International Chess Federation scale, on which 2,200 denotes a master and Fischer's rating is 2,785. Apart from Mikhail Botvinnik of the U.S.S.R., Kalme is the highest-ranking player currently engaged in the development of a computer-chess program. As a result of Kalme's tutoring we believe the U.S.C. program is much stronger than it was last August, and we are looking forward to this year's computer-chess championship. A game between Kalme and the U.S.C. program is presented at the end of this article, together with an analysis by William Lombardy, the international grand master who was Fischer's second in Iceland, and a separate analysis by Kalme himself.

Why should anyone bother to program a computer to play chess? Apart from the sheer intellectual challenge of the task there are several significant justifications. Specific techniques developed in programming a computer to play chess have already been used in writing programs for solving other types of problem, particularly those involving a search among many alternate pathways, as in a telephone-switching system or an electric-power grid. As new methods are uncovered they too will become generally useful. There is also the fascinating possibility that methods for the recognition of structural patterns on a chessboard will eventually help computer robots to see. At a deeper level there is the hope that in writing chess programs we shall eventually gain important clues to how the human brain works: how it analyzes patterns and quickly abstracts what is important from what is less important.

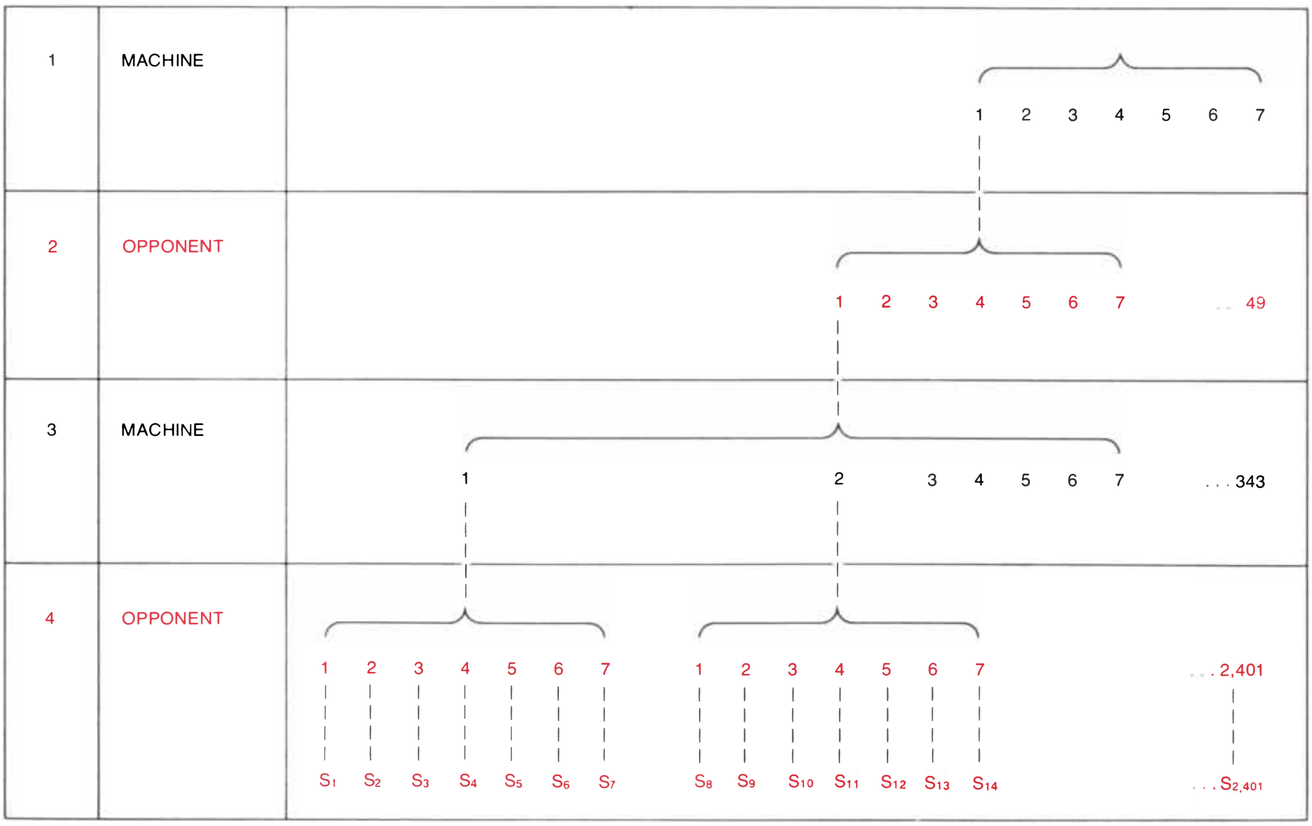
The first comprehensive description of how an electronic computer could be programmed to play chess was given by Claude E. Shannon, then at Bell Laboratories, in a talk before the National Institute of Radio Engineers in 1949. Writing in *Scientific American* in February, 1950, Shannon explained why he thought the chess-playing problem merited attention: "The problem is sharply defined, both in the allowed operations (the moves of chess) and in the ultimate goal (checkmate). It is neither so simple as to be trivial nor too difficult for satisfactory solution."

In his 1949 talk Shannon proposed that the numbers stored in a computer's memory could be used to represent various chess positions, and that the arithmetical and logical operations of the computer could be combined into procedures for playing chess. Starting from a given board position the computer tries out hypothetical moves. It then examines the opponent's possible replies to each move. For each reply the machine again considers another set of hypothetical moves, and so it goes. With the computers then available, Shannon thought it would be possible to look ahead four half-moves, or two full moves for each side. Since the number of legal moves available to a player at each turn averages about 30, a full look-ahead to a depth of four would require examination of about 30^4, or 810,000, moves. Shannon therefore proposed that the computer should not consider all possible moves from each position but should prune out the most obvious of the bad moves.

When this scheme was actually implemented in the late 1950's with an IBM 704 computer, a look-ahead search of two full moves took eight minutes. Even then the number of branches searched at each fork in the tree was pruned to no more than seven, so that only 7^4, or 2,401, discrete moves were inspected [*see bottom illustration on next page*]. Evidently a large stock of



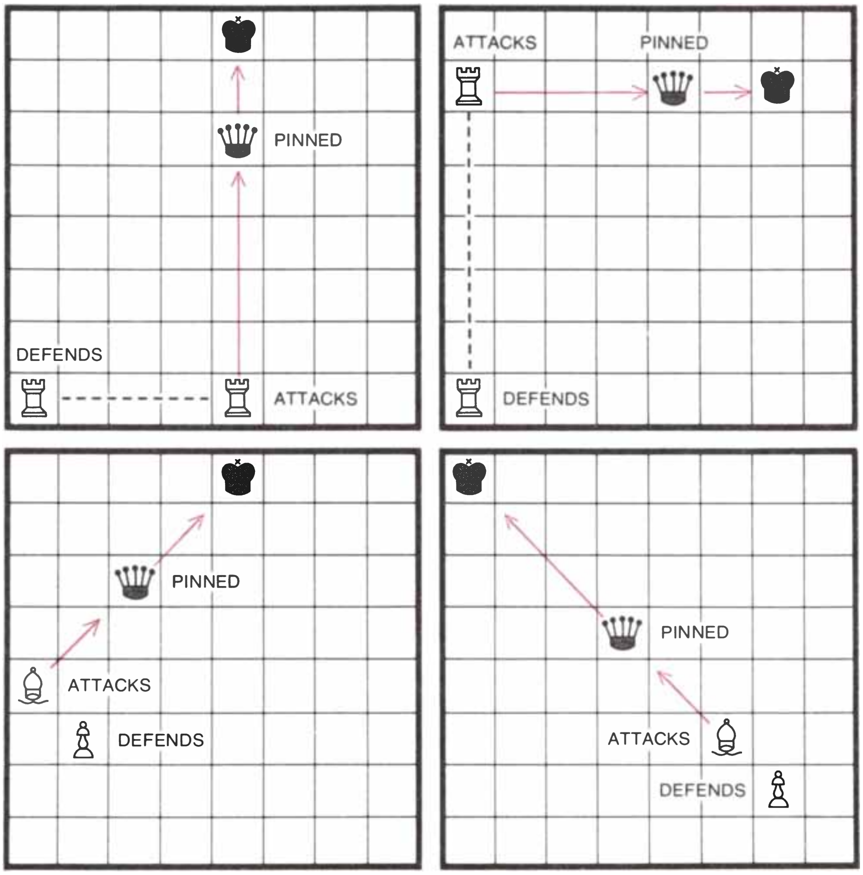
CHESSBOARD AND BASIC MOVES are probably familiar to most readers. The conventional symbols for the six kinds of white and black pieces are shown at the left. The board at the right illustrates typical legal moves. The notation for specifying each move appears at the far right. The files, or vertical rows, are designated by the initials of the pieces that originally occupy the first, or back, rank; thus the file at the extreme left is the queen-rook (QR) file. The ranks, or horizontal rows, are numbered so that each player's first rank is No. 1. Only the knight can jump over intervening pieces; the others require a clear path. A one-time exception in each game is castling, a king-and-rook leapfrog maneuver (*colored squares*). If performed on the king side, it is symbolized by O-O; if performed on the queen side, by O-O-O. A capture is indicated by x; thus in the final board position shown at the right whichever queen had the move could capture the other: QxQ. The white bishop could capture the black knight on the square K2 (BxN), but the black knight could not capture the white bishop because the knight can only jump over the two intervening squares in a dogleg.



SHANNON LOOK-AHEAD PROCEDURE is named for Claude E. Shannon of the Massachusetts Institute of Technology, who proposed it in his original description of a computer-chess program in 1949. From a list of all reasonable moves the machine selects, say, seven as the most logical on the basis of criteria written into the program (*1*). The machine then selects the seven most logical replies the opponent can make to each of these seven moves (*2*). At this level it must examine 49 moves. For each of these the machine selects seven countermoves (*3*), or a total of 343 moves (7 × 49). Continuing in the same fashion, the machine selects the seven most logical responses the opponent can make to each of the machine's 343 moves (*4*). Finally the machine assigns a figure of merit, or score (S_n), to each of the 2,401 moves the opponent might make. If the analysis is discontinued here, the computer will have looked ahead four half-moves, or two full moves. In current programs of the Shannon type the routine may examine half a million hypothetical moves, penetrating to an average depth of five half-moves. Authors' program is a major departure from Shannon approach.

chess knowledge is necessary in order to achieve good pruning. Chess masters can easily look ahead more than six half-moves because they have learned how to select only the most promising moves and countermoves for mental examination. In current programs of the Shannon type the look-ahead search usually averages five half-moves, but in special cases it can go as deep as eight half-moves, examining as many as half a million moves in all. The quality of play achieved with programs of the Shannon type can be traced largely to the programmer's skill in translating his specific knowledge of chess into the mathematical notation that computers understand. This has proved to be an immensely difficult task.

The difficulties encountered in nearly two decades of trying to improve programs of the Shannon type can be contrasted with the comparative ease of "programming" human beings to play chess. A single rule such as "Keep your knights off the edge of the board" can quickly and concisely convey chess knowledge from a teacher to a student. (The student will also appreciate that such rules are not immutable but must be violated at times.) Putting this kind of knowledge into a program of the Shannon type is so difficult that most programs have fewer than 20 such pieces of information with which to play, which is a minuscule amount by human standards.



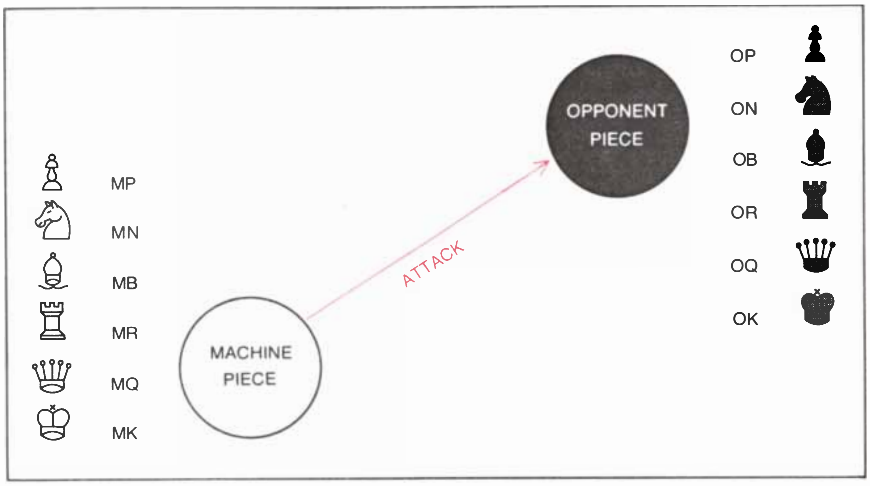
TEACHING COMPUTER TO GENERALIZE is one of the major goals in writing a computer-chess program. As these four diagrams illustrate, the basic structure of chess is somewhat independent of the actual positions of the pieces. In the two top diagrams a white rook pins the black queen against the black king, which means that the queen must either stay put or else capture the attacker. If the queen elects the latter course, she will herself be captured by the second rook. The two bottom diagrams show variations on the same theme except that here a white bishop does the pinning while being defended by a lowly pawn.

These considerations prompted us to consider a program by means of which a computer would be able to generalize from particular mistakes. The exact board position that led to a bad move will rarely be repeated, so that the computer must learn general principles that will enable it to correct itself in positions that are similar to, but not the same as, the position in which the mistake was made [*see illustration on this page*]. Since it appears very difficult to have the computer learn chess principles by itself, one possibility is to have it learn from an expert teacher, much as a child learns in elementary school. Such a chess machine would not have to induce or deduce new principles but would simply accept advice on how to play chess. This approach to machine learning, under certain restrictions that we shall describe, we have called advice-taking.

Because the approach involves the communication of knowledge, a language must be created to transfer knowledge from man to computer. There are four basic requirements. First, the language must be powerful enough for the description of significant chess strategies. Second, it must be simple enough (or natural enough) so that it can be understood by chess players even if they are not expert programmers. Third, the advice-taking part of the program must be able to interpret advice in the language. Finally, of course, the program must be capable of being implemented on a digital computer. Our purpose was not simply to study chess but to study advice-taking mechanisms in general. For our investigation we have had access to an IBM 370/155 computer at U.S.C.

The basic components of a language are its primitives and its syntax, that is, its words and its grammar. We have taken as primitives the 64 squares of the chessboard and the 12 types of chess piece. Also taken as primitives are an assortment of numbers that represent the details of the chess position, such as the number of elapsed moves, the number of white or black pawns remaining at any given time and whether or not either side has castled. The syntax of the language enables one to describe structural combinations of primitives. We use the term "pattern" to refer to statements in the chess language that describe general combinations of pieces that occur frequently in human play. For example, an elementary pattern might describe the general case of a machine piece attacking an opponent piece [*see top illustration on next page*]. We use the term "instance" to describe a particular configuration of pieces that fits one of the patterns.

The key mechanism of the U.S.C. program is a routine that can search a given board position for all instances of the patterns and record them internally. We call such an internal recording a snapshot. Only a few lines are needed to write down a pattern in the chess language [*see left half of bottom illustration on next page*]. The first two lines are "selectors" that can choose individual pieces



GENERALIZED ATTACK PATTERN is an essential part of the authors' program. The diagram emphasizes that the pieces do not have to occupy any particular square of the board. The program specifies, of course, how the pieces are permitted to move in making an attack. Thus a pawn can attack only pieces located one rank ahead on an adjacent file.

on the chessboard; thus OP stands for "opponent pawn" and MP stands for "machine pawn." Several such lines acting together can select combinations of pieces in specified configurations. At the end of the selector lines in parentheses are the words "ATTACK" or "FREE." "ATTACK" tells the computer to search for a piece on that line that is attacking some piece selected on the other line. "FREE" tells the computer to consider not only the actual location of pieces but also hypothetical locations that the pieces can reach in a few moves. The selectors are designed to iterate all possibilities, so that the attack pattern will usually find hundreds of instances of real or potential attacks [*see top illustration on opposite page*].

Under the selector lines there are a number of short commands that cause the programs to take a snapshot of the selected combinations of pieces and to code the snapshot internally together with a weight, or value. The parameters of the "WEIGHT" command simply tell the program what type of chess information to use in calculating the value of a snapshot. For example, the weight of an attack increases if (1) the attacked piece is valuable, (2) the attacking piece is cheap, (3) the attacked piece is close to its king and (4) the attacked piece is close to the center of the board. The total weight is intended to measure the worth of a particular snapshot; the weights of all relevant snapshots are included later in the computer's calculations when it attempts to evaluate the merits of various moves. The attack snapshots are of two types: those showing machine attacks and those showing opponent attacks, each being weighted appropriately.

The attack patterns and their resulting snapshots are involved in every move. Typically the computer creates, weights and stores some 1,500 attack and counterattack snapshots in roughly six seconds. Other patterns are less general, and some are quite specific. For example, one pattern is devoted to getting the knights and bishops off the back rank in the opening [*see right half of bottom illustration on this page*]. It works by giving negative weights to those pieces as long as they sit on the first rank. The first and third lines of the pattern, called predicates, act as barriers that cannot be passed unless certain conditions are met. The first predicate checks to see whether or not the game is in the opening phase between move No. 3 and No. 10. The second predicate determines whether or not a selected knight or bishop is on the back rank. Such predicates are a powerful means of precisely describing a chess situation.

The back-rank pattern played a significant part in the historical development of our program. When the initial version was finished in April, 1972, we decided to exhibit it at a computer laboratory open house. The first two games were a fiasco, because in each game the machine failed to develop one knight and one bishop, preferring instead to be hyperactive with its other pieces. To correct this flaw the back-rank pattern was typed into the computer, and the program has not had the problem since. The corrective procedure took less than five minutes; its ease and effectiveness convinced us that we had indeed achieved an advice-taking machine. Even minor changes in the chess "knowledge" of a program of the Shannon type call for many hours of conventional programming and debugging.

```
1          OB,ON,OQ,OK,OR,OP(FREE)
2 (1,A)    MB,MN,MQ,MK,MR,MP(ATTACK,
           FREE)
           PREPARE
           CODE(1,2)
  (1,F)    WEIGHT(1,-100%)
  (2,F)    WEIGHT(3,25%)
  (1,A)    WEIGHT(2,50%)
  (1,A)    WEIGHT(11,-50%)
           CREATE
           END
```

```
           COND(7,3,10)
1          MN,MB
  (1,A)    RANK(1,1)
           PREPARE
           CODE(1,1)
  (1,F)    WEIGHT(1,-500%)
           CREATE
           END
```

COMPUTER INSTRUCTIONS are written in the chess language devised by the authors. In the basic attack pattern (*left*) the first line selects an opponent piece (OB stands for opponent bishop and so on). The second line selects a machine piece that can attack the opponent's piece. "FREE" tells the computer to consider not only the actual location of pieces but also hypothetical locations that can be reached in a few moves. "WEIGHT" tells the computer how to assign weights, or values, to different kinds of attacks. "CREATE" causes a snapshot of the situation to be taken and stored in the computer for later use in determining the move. The routine at the right is an early piece of advice that helps the computer to move its knights and bishops off the back rank in the opening between the third and tenth moves. It works by giving negative weights to the knights and bishops on their original squares.

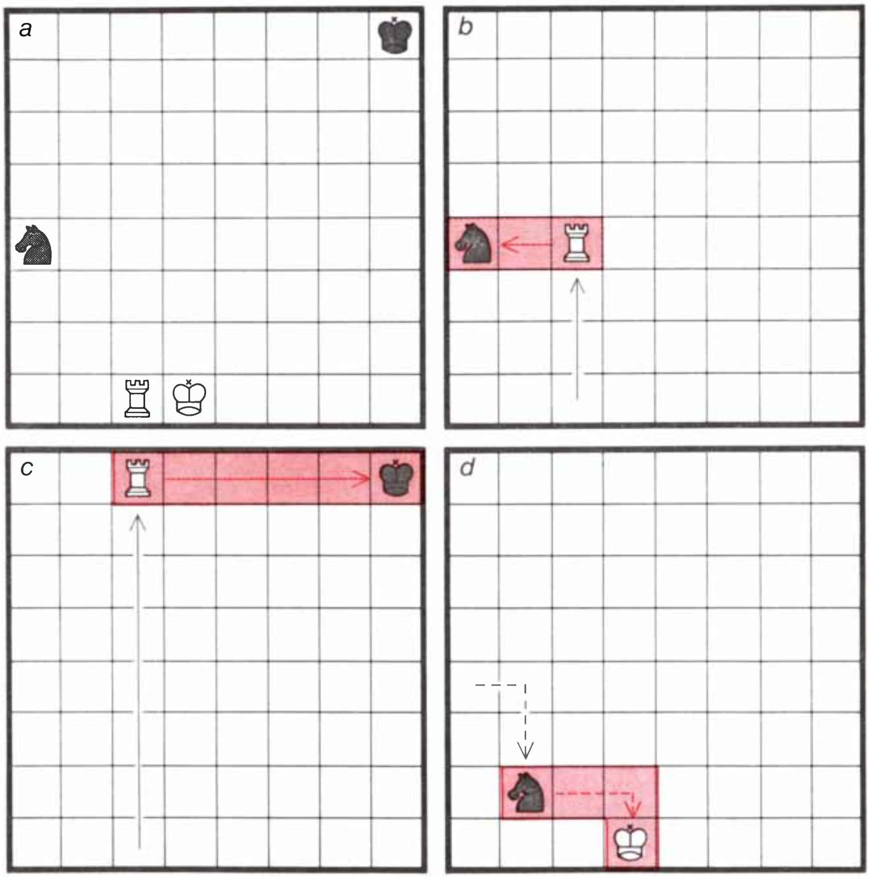
Our chess language is somewhat richer than these examples indicate. At present the language has three types of selector, 15 types of predicate and 15 other statements that can be used in patterns. The additional selectors and predicates can be used to describe a wide variety of chess situations, such as pins, forks, "holes" in pawn formations and simple mates.

The latest generation of computers has made the mechanics of computer chess quite simple. By dialing the proper telephone number one can connect a teletypewriter to the U.S.C. computer. One then types in the proper password, requests execution of the chess program and the machine is ready to play. After a move is typed in (say P-K4) the computer checks its legality, calculates a move in reply and types the move out. The human contestant can also request a printout of the chessboard and can give the computer some advice if he so wishes.

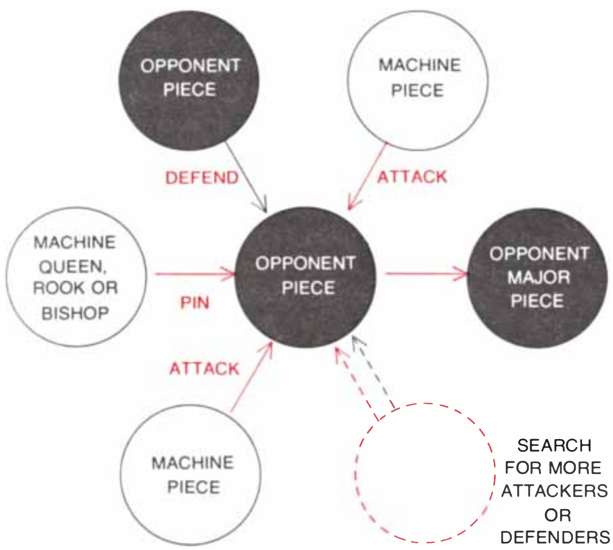
The machine replies to opening moves almost instantly because we have stored internally about 200 sequences of moves covering about a dozen popular openings. Thus it immediately selects a time-tested defense against, say, a Ruy Lopez opening or an Evans Gambit. Players who challenge the computer for the first time are likely to find its one-second replies to their first half-dozen moves rather intimidating. We claim no credit for this feature. Since machine time is costly, it has become standard practice in the computer-chess league to provide the computer's memory with a plentiful stock of opening moves.

Indeed, the reader may wonder why this technique should not be carried further. Given sufficient memory capacity and a fast search routine, why not store the first 15 or 20 board positions from several thousand of the best games played in the past century? The computer would either play the ultimate winner's move at each step or use a random strategy in selecting its move if more than one move evidently led to a winning game.

A little combinatorial mathematics will tell why this is not a promising approach. Men have certainly played fewer than 10^{15} chess games. By way of contrast there are about 10^{43} possible board positions and about 10^{125} ways of moving pieces to reach those positions. This means that only an infinitesimal fraction of all possible chess positions have yet been seen. Thus a rote-memory machine could be defeated by making



VARIETY OF ATTACK SNAPSHOTS are generated by the computer from any given board position. Playing White in the situation shown in *a*, it could attack the opponent's knight by moving its rook to B4 (*b*), or it could put the opponent's king in check by moving the rook to B8 (*c*). The attack pattern would give the first move a weight of about 30 and the second a weight of about 100. Another snapshot would show that the opponent's knight could put the machine's king in check by moving to N7 (*d*). This would be given a weight of about minus 100. Machine will usually find hundreds of instances of real or potential attacks and will make use of their aggregate weights in selecting its ultimate move.



MORE COMPLEX PATTERN, depicted graphically, enables the computer to discover opportunities to pin an opponent piece against a major piece, such as a rook, a queen or a king. Pattern is flexible not only as to the positions of white and black pieces involved but also with respect to the numbers of pieces attacking and defending the pinned piece.

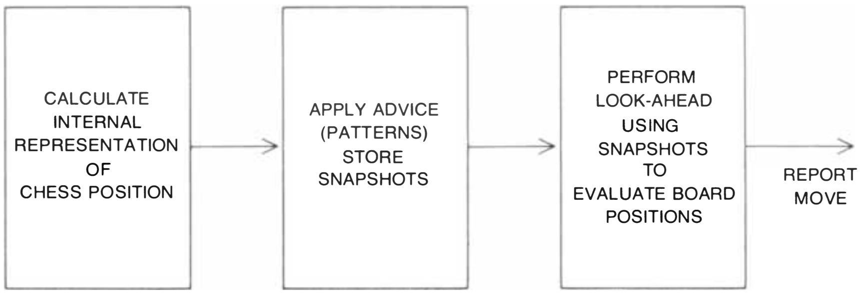
an unusual move. In fact, this can usually be done as early as the third move of any opening.

After the first half-dozen moves or so the computer switches over to its "thinking" mode. It then takes about 25 seconds to decide on its move. The calculation is performed in three basic steps [*see upper illustration on this page*]. In the first step the program calculates an internal representation of the chess position after its opponent's last move. These calculations begin with an eight-by-eight matrix of integers in which the value of the integer denotes the type of chess piece. From that another matrix is calculated in which the value of the integer gives the value of the piece. The usual value of a pawn is 1, of a knight or a bishop 3, of a rook 5 and of the queen 9. The kings are arbitrarily assigned a value of 20 even though they are priceless and are never traded. Other matrixes calculated by the machine represent such esoteric matters as holes in the opponent's pawn structure, open files on the board or distance from a king. In all, 20 such matrixes are calculated and stored in the computer's memory in two seconds.

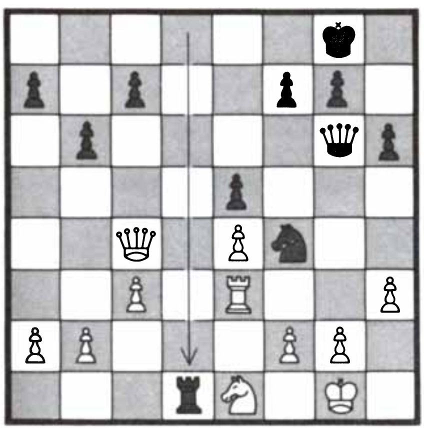
The second step is the one distinctive to our program: the application of advice, in the form of patterns, to the internal representation. The machine is now provided with 50 patterns, including the attack pattern and the back-rank pattern we have described. During the applications of these patterns between 1,000 and 3,000 snapshots will be coded and saved. The snapshots include not only selected pieces as they actually sit on the board but also snapshots of hypothetical positions after one or more moves and countermoves. Each of these snapshots has its own weight and value [*see illustration on opposite page*]. With the aid of the snapshots the computer makes a preliminary selection of the 10 "best" moves and lists them in descending order of merit. The elapsed time up to this point is eight seconds. The third step is most like the standard Shannon program: it applies a look-ahead procedure to calculate the machine's move. The important difference in our program is that the snapshots are used at every fork of the look-ahead to select moves for further examination just as they were used to select the 10 moves at the first stage of the look-ahead. The program uses a consensus of all the snapshots in determining the relative values of moves. The basic assumption is that a single piece of advice is somewhat unreliable. For example, "Keep your knights off the edge of the board" is usually good advice, but in some cases a snapshot of a knight on the edge of the board will carry a higher weight than a snapshot showing a knight in the middle of the board. Everything depends on the location of the opponent's pieces; a knight on the edge can sometimes lead to a checkmate. Our look-ahead routines typically examine about 10,000 moves in 15 seconds. Because the look-ahead is correlated with the stored snapshots, which represent significant chess knowledge, our program has to examine far fewer moves than programs of the Shannon type do. Our typical look-ahead proceeds to a depth of from five to 10 half-moves. Sometimes, however, the computer will penetrate as deep as 14 half-moves [*see bottom illustration on page 100*]. Because our approach attempts to make a reliable decision from somewhat unreliable information most computer workers would describe it as heuristic.

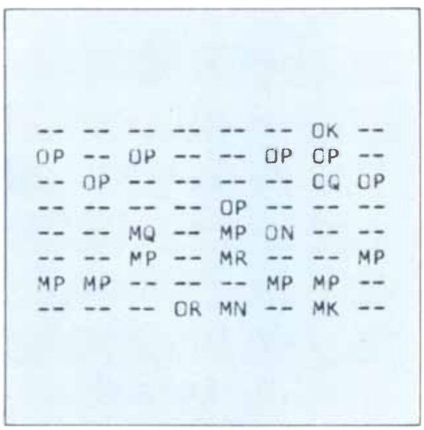
Theoretically speaking we feel that our work has achieved a considerable degree of success. We have been able to express a wide variety of patterns for openings, the middle game and the end game in the chess language. Giving advice to the program is so easy that we feel as if we can "talk" with the computer. Our chess-master collaborator, Kalme, had no difficulty learning the chess language in spite of a total lack of knowledge of computer programming. For the first time the computer is no longer completely dependent on expert programmers for the acquisition of chess knowledge. We see in our approach the fascinating possibility that the game's greatest players, such as Fischer, could record their chess technique for posterity.

Although our program embodies a

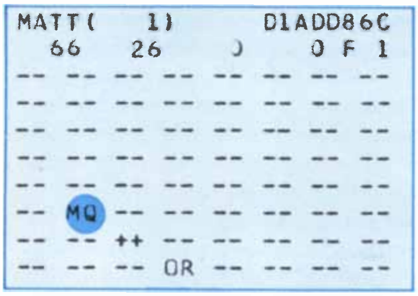


THREE BASIC STEPS are employed by the advice-taking chess machine in calculating a move. Normally the first step takes about two seconds, the second about six seconds and the last about 15 seconds. The computer performs some one million operations per second, or roughly 25 million operations per move. When the board position is complicated, however, the computer may take up to 90 seconds for a move. It took about 40 seconds after the 23rd move in the game that provides the material for the following series of illustrations.

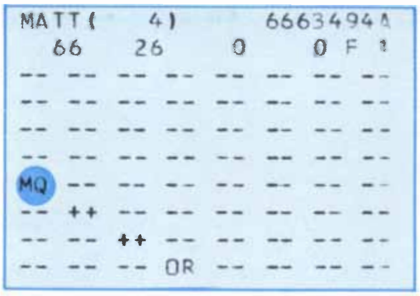




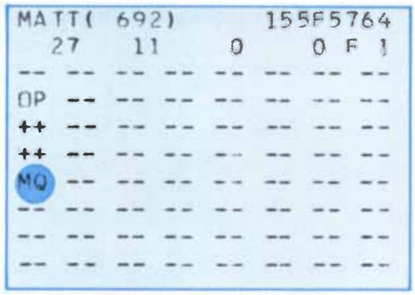
BOARD POSITION AFTER BLACK'S 23RD MOVE is depicted in conventional form at the left and in the computer's printout version at the right. The black pieces are being played by the computer's tutor (Kalme). In the printout the letter *O* identifies an opponent piece (one of Kalme's black pieces) and *M* signifies a machine, or white, piece. Kalme has just intensified his attack on the computer's king by advancing his rook to Q8. (The preceding moves can be found in the analysis of the game that begins on page 101.) To discover the best reply the computer begins by generating some 1,500 snapshots, 15 of which are shown on opposite page. Snapshots are used in selecting and evaluating look-ahead moves.



1. Q-N3 ATTACKS ROOK: 66 + 26 POINTS

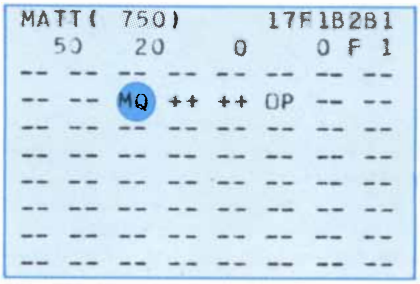


2. Q-R4 ALSO ATTACKS ROOK . . .

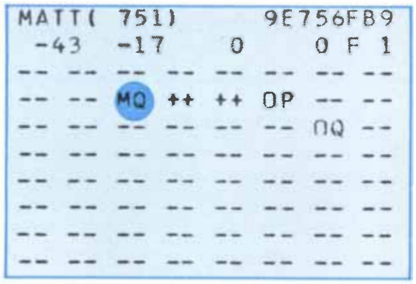


3. . . . AND ROOK PAWN AS WELL

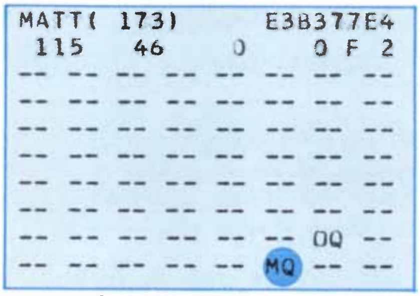
4. QxP YIELDS ANOTHER ATTACK ON ROOK PAWN . . .



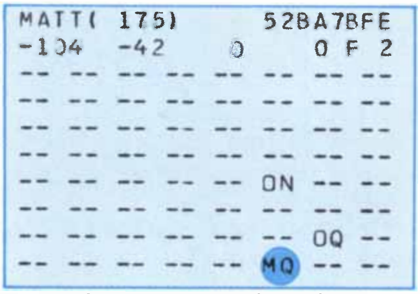
5. . . . AND ALSO ATTACKS BISHOP PAWN . . .



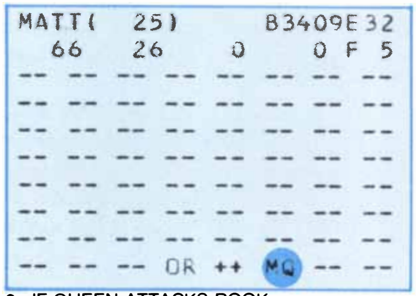
6. . . . BUT, ALAS, IT IS DEFENDED BY THE QUEEN



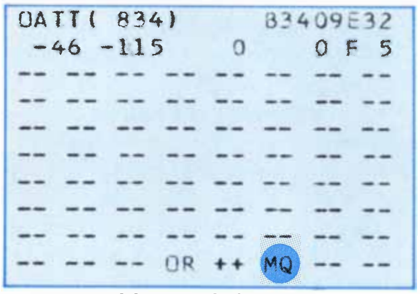
7 Q-B1 WOULD PREVENT THE MOVE OQ-N7 . . .



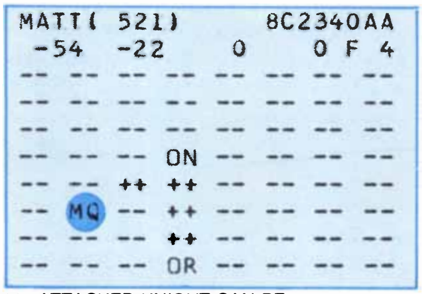
8. . . . EXCEPT THE KNIGHT STANDS BY TO DEFEND IT



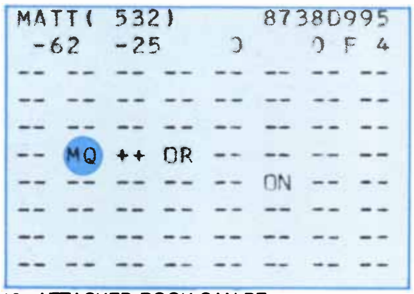
9. IF QUEEN ATTACKS ROOK: 66 + 26 POINTS



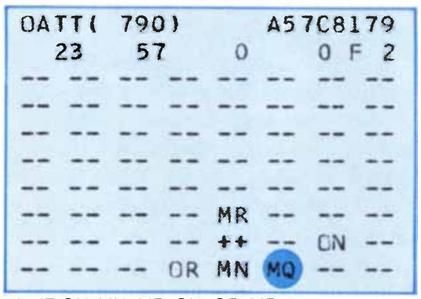
10. BUT IF ROOK ATTACKS QUEEN: −46 −115 POINTS



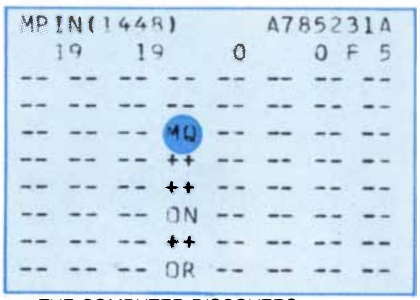
11. ATTACKED KNIGHT CAN BE DEFENDED BY ROOK



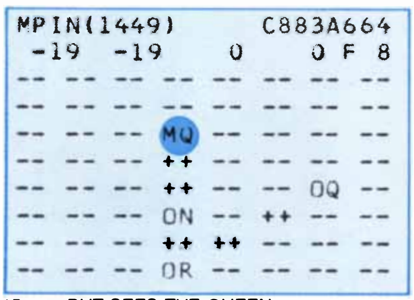
12. ATTACKED ROOK CAN BE DEFENDED BY KNIGHT



13. IF ONxMN, MRxON, ORxMR, MQxOR: 80 POINTS

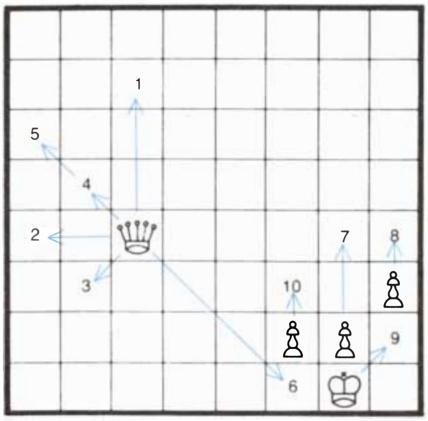


14. THE COMPUTER DISCOVERS A POSSIBLE PIN . . .



15. . . . BUT SEES THE QUEEN CAN REFUTE IT

SAMPLE OF 15 SNAPSHOTS OUT OF 1,500 suggests the variety of chess positions examined by the computer after Kalme's 23rd move. "MATT," "OATT" and "MPIN" stand respectively for "machine attack," "opponent attack" and "machine pin." The numbers in the second line are weights: positive if they favor the computer, negative if they favor the opponent. The first two numbers reflect immediate gain (or loss); the third number reflects longer-range values. All but the last two snapshots show positions the machine's queen (MQ) can reach in one move and the opponent pieces it can then attack or be attacked by. Note the sharp changes in weight that occur between pairs of similar snapshots (5 v. 6, 7 v. 8 and 9 v. 10) when the machine discovers it is vulnerable to an opponent piece. The last snapshot, a pin that is refuted by the black queen, appears in the design on the cover of this issue of SCIENTIFIC AMERICAN.



	MOVE	"PRUNING" VALUE
1	QxP	33
2	Q-R4	-53
3	Q-N3	-60
4	Q-N5	-67
5	Q-R6	-79
6	Q-B1	-79
7	P-N4	-81
8	P-R4	-81
9	K-R2	-88
10	P-B3	-91

PRELIMINARY SELECTION OF 10 BEST MOVES for the computer's move No. 24 is made on the basis of some 1,500 snapshots. The moves are assigned a "pruning" value that sums up the weights associated with all the snapshots in which the machine's pieces occupy a square reached by the hypothetical move. At this stage of the calculation the machine has discovered that a queen move is the most desirable but the move ultimately selected on the basis of the look-ahead procedure, Q-B1, still looks inferior to at least four other moves.

a	24.	QxP	RxN
	25.	RxR	QxP MATE
b	24.	Q-R4	RxN
	25.	K-R2	QxP MATE
c	24.	Q-N3	RxN
	25.	K-R2	RxR
	26.	K-R1	R-K8 CH
	27.	Q-B1	RxQ CH
	28.	K-R2	QxP MATE
d	24.	Q-B1	NxRP
	25.	K-R2	N-N4
	26.	K-N1	NxBP
e	24.	Q-B1	NxRP
	25.	K-R1	N-KB5
	26.	K-R2	NxBP
	27.	QxN	QxQ
	28.	NxQ	R-Q7
f	24.	Q-N5	RxN
	25.	K-R2	RxR
	26.	K-N1	R-K8 CH
	27.	Q-B1	RxQ CH
	28.	KxR	QxP CH
	29.	K-K1	Q-B8
	30.	KxQ	NxP

FOR NEXT 25 SECONDS MACHINE LOOKS AHEAD, testing one hypothetical move against another. The 48 half-moves shown here are selected from more than 15,000 examined. On each half-move the machine refers back to each of the 1,500 snapshots for guidance. It soon discovers the fatal flaw in QxP, Q-R4 and Q-N3. (The computer has since been taught how to find the one-move mate available to Kalme after any move but Q-B1.) The computer finally sees that Q-B1 avoids mate (*d, e*) and in a few cases looks ahead 14 half-moves (*f*). The authors emphasize that many of the lines studied by the computer may look unreasonable even to a novice, but that the cumulative effect is what counts. The computer finally selects Q-B1 as its move. In his analysis Kalme calls it "the only defense."

number of innovations, implementation has proceeded rapidly enough so that it now plays at a sound novice level. Chess 3.5, the Northwestern program that holds the present national computer title, has a rating of around 1,500. We believe our program, although it is not officially rated, plays in the upper part of the 1,200-to-1,500 range. When the program wins, it appears quite clever and is usually very aggressive.

Computer-v.-computer games give us a crude understanding of the strength of different programming approaches. Games between the computer and a human expert provide a still more penetrating test of our program's strengths and weaknesses. Kalme shares our optimism that the advice-taking approach promises to produce computer-chess games at a level of sophistication attained by serious players. In five years' time our program may be sufficiently advanced to challenge international master David Levy of Britain, who wagered 1,000 pounds that no computer will be able to beat him in a 10-game series before August, 1978. (Botvinnik is reported to have told Levy: "I feel very sorry for your money.")

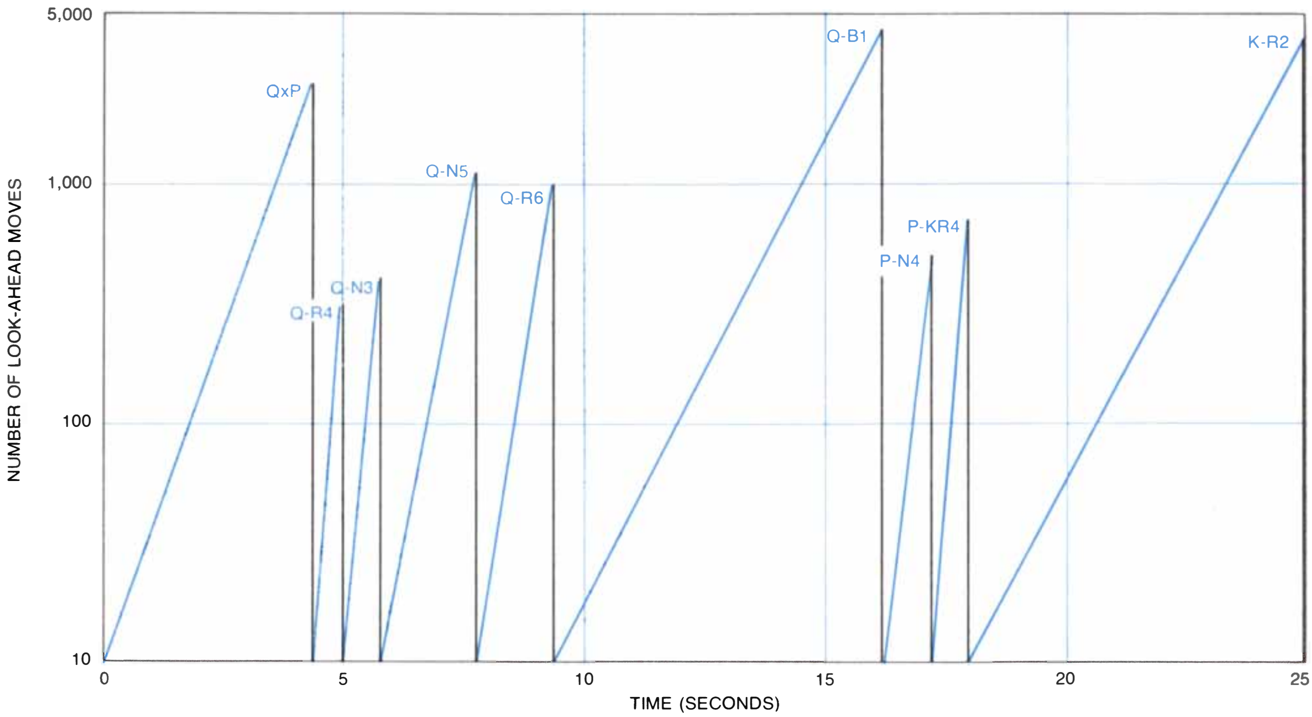
If an advice-taking machine demonstrates an aptitude for the complex intellectual task of chess, it may have similar value when applied to other problems of more practical interest. Our work suggests the possibility that computers of the near future will use advice-taking, as we have defined it, for the performance of tasks now believed to demand more "intelligence" than is needed for simple clerical chores and repetitive numerical calculations, which is the kind of work commonly done by computers today.

We readily concede that the burden is on us to demonstrate that a machine can beat a man at his own game. In the past those who believed a computer could never play master chess argued somewhat as follows: "For a machine to beat man at chess the machine must have a greater knowledge of the game. But if the machine gets its chess knowledge from man, whether by straightforward programming or by some advice-taking procedure, the machine will always have less knowledge than man." That argument usually leads to the conclusion that computers must be provided with some form of active learning process to attain championship proficiency in chess. In other words, one must somehow duplicate inside a machine the mental processes that enable a Bobby Fischer to become a better chess player than the people who taught him.

Even without active learning, however, the computer has some powerful advantages. Because of its memory capacity a computer could be set up to accept huge volumes of advice from teams of experts. In time the computer could know more than any one individual. Because of its tremendous speed the computer can see nontrivial consequences of even trivial advice. Furthermore, the computer is not subject to lapses of memory or concentration; it will not be bothered by poor lighting or brood over a blunder. Once advice is given, the computer will not forget it and will never fail to apply it.

A deeper and perhaps more serious question involves the use of language for the expression of chess knowledge. One might argue that many of the world's greatest players have already tried to describe their techniques in books and articles, but that no reading of their words is sufficient to convert an average player into a player of the first rank. The failure, of course, could be with the person reading such works.

On the other hand, it is possible that a significant portion of human chess technique cannot be expressed in words. For example, a grand master might be quite unable to explain the reasoning



TIME ALLOCATED TO HYPOTHETICAL MOVES in a search for move No. 24 indicates that the computer spent more time evaluating Q-B1 than any alternative. Although K-R2 was hopeless, it could have been interesting under slightly different circumstances. The computer had the good sense not to waste any time examining P-KB3, the least promising of the 10 moves selected on the basis of the snapshots. Programs of the Shannon type may examine as many as 500,000 look-ahead moves, a task that may take several minutes.

behind a particularly brilliant move. It is not just a question of whether one can have thoughts without words but of whether one can have thoughts of much greater content than the words one can find to express them. It is possible that chess thought depends heavily on spatial perception and that the perceptual processes involved are so subtle and rapid that only the final outcome reaches conscious expression. If so, there may be no adequate language for conveying chess knowledge to a computer.

This problem involves several areas of psychology and warrants much further study, but our tentative conclusions are as follows. First, artificial languages may actually be superior to English or any other natural language for the expression of chess ideas. For example, computer languages such as ALGOL have proved superior to the mixture of English and mathematical notation traditionally used to express algorithms. (Some computer scientists have even claimed that computer languages have enabled them to think thoughts that could not be expressed in English.) With our chess language it may be possible to communicate directly to a computer chess strategies that heretofore have been passed from master to master by means of example or inference.

Our second conclusion rests on the use of a pattern-based language for chess. Psychologists have evidence that the chessboard is perceived in "chunks," or combinations of pieces, but they can only speculate about the nature of these chunks. We feel that the patterns in our program operate as "chunk-detectors" and that our entire approach should be of interest to psychologists as a model of human chess perception. In effect we seek not only to rival man's ability at chess but also to simulate his methods of playing chess. Thus the early approach to computer chess, which reflected a fascination with mathematical algorithms for efficient play, is being replaced by an approach in which the computer becomes a tool for building models of the mind.

Analysis of a Game between U.S.C. Computer (White) and Charles I. Kalme, U.S. Senior Master (Black)

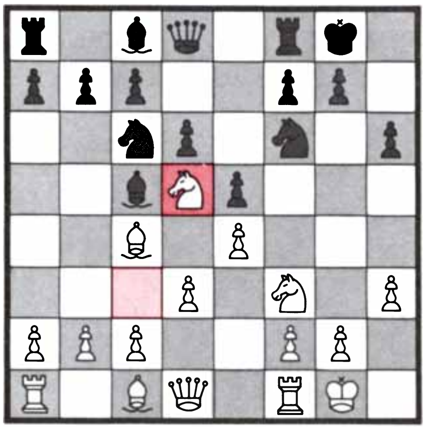
by William Lombardy
International Grand Master

I would describe the opening of this man-v.-machine game as a Giuòco Pianissimo. The chess opening Giuòco Piano is so named for its characteristically steady, purposeful plan for constructing an impressive pawn center. White's 4 P-QB3 signals the Piano. If by nature the Piano is slow, then the Pianissimo is slower. The Pianissimo makes no radical attempt to control the center with pawns. If, on rare occasions, that is the intention, then the buildup is more methodical and time-consuming (mostly the latter). The Pianissimo tends to reject P-QB3, supporting a pawn advance to Q4, in favor of subtle piece play. White jockeys positionally, gradually gaining for his minor pieces (bishops and knights) key central squares, springboards for future attack. Since such strategy is far more conceptual than mathematical, I would suggest that in its present stage of development the computer is at a distinct disadvantage when it is opposing the human mind.

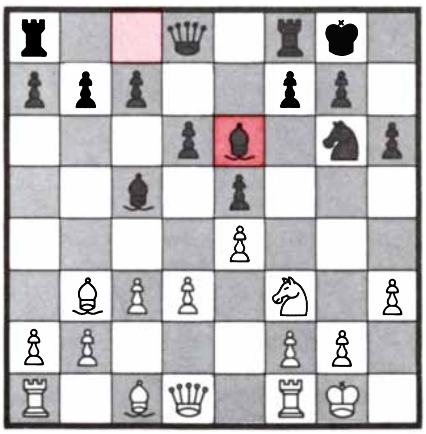
The British international master David Levy goes one step further. He has placed a considerable wager that a computer, given until 1978 to prepare for the match, will not survive a 10-game series against him. To this I would say, "Good luck, David!" [The diagrams on the following pages, keyed to the moves of the game by letters, refer both to this analysis and to the subsequent analysis by Kalme. Color identifies the piece last moved and the square it last occupied.]

a POSITION AFTER 4 O-O

b POSITION AFTER 5 P-KR3



c POSITION AFTER 8 N-Q5?



d POSITION AFTER 11 . . . B-K3!

1 P-K4 P-K4 3 B-B4 B-B4
2 N-KB3 N-QB3 4 O-O . . .
[*see diagram a*]

My analysis of this game is without reference to the identity of the opponents. At this point I surmised White to be the computer; by move 5 I was positive of its identity. Although White's 4 O-O is reasonable enough, the move is generally considered too passive to offer any real hope for an opening advantage. In all fairness, though, Black has at his disposal adequate defenses on whatever course White chooses to press the temporary initiative afforded by the first move.

Chess, however, is a game in which players must assess their chances, selecting lines most likely to apply maximum pressure on an opponent. The computer's move cannot be criticized as being totally incorrect. But let us say that chess grand masters accustomed to finishing well in but not winning tournaments are consistently plagued by excessive technical accuracy, this of course from the viewpoint of a human being opposing another human being. The inevitable sterility of coldly accurate play produces a string of draws with no victory in sight.

4 . . . P-Q3 5 P-KR3 . . .
[*see diagram b*]

This is blatantly overcautious. An amateur learns early in his career to beware the threatened pin (a maneuver designed to paralyze an enemy piece, any move of which might result in either checkmate or the fatal loss of material). The beginner fears any pin whatever, because he is unable to distinguish pins that represent genuine danger from those that are just a minor irritation, therefore the computer's P-KR3. The master recognizes those pins whose minor irritation may be endured for prolonged periods without ultimate disaster, which was the case with the pin threatened here.

Naturally the following line would not have occurred to the ordinary player, in this case the computer in an unimaginative mode: 5 P-B3 B-N5, 6 P-Q4 PxP, 7 Q-N3! Q-Q2!. Rather than save his queen knight pawn, Black defends the more pertinent threat against his king bishop pawn. 8 BxPch! QxB, 9 QxP K-Q2, 10 QxR BxN, 11 PxB PxP, 12 NxP QxBP, 13 B-K3 Q-N5ch, and Black squeezes out with a perpetual check. (The enemy king cannot escape check by an adversary's piece or pieces, and the game is therefore drawn.) Weighing his chances in a risky line, White embarked on a course that enabled him to exert powerful, even if temporary, pressure on his opponent. Only an experienced expert could have walked the tightrope to the draw.

5 . . . N-B3 6 N-B3 O-O

Without hastily castling, Black can play for the attack with 6 . . . P-KR3, 7 P-Q3 P-KN4, threatening . . . P-N5 and the disruption of White's king side. Also good is 6 . . . B-K3, neutralizing White's king bishop.

7 P-Q3 P-KR3

Black cannot allow the pin in this situation, because to break it he would need to dangerously weaken his own king position: 7 . . . P-QR3 (so that his king bishop has a retreat against the threat of 8 N-QR4), 8 B-KN5 (menacing 9 N-Q5) P-KR3, 9 B-R4 P-KN4.

8 N-Q5? . . .
[*see diagram c*]

Here it is obviously the computer at work. The strategic idea demands elimination of Black's king bishop with 8 N-QR4. Artificially aggressive, the computer's move is a futile waste of time. The computer has clearly been instructed in the adage: "A knight on the side cannot abide." It must also be taught exceptions to the rule.

8 . . . NxN 10 B-N3 N-N3
9 BxN N-K2

Black has attained a slight initiative owing to his comparatively well-posted knight, which will come thundering into KB5. The edge, however, is far from enough to win.

11 P-B3 . . .

White's pawns, ready to occupy the center, will be mere tenant farmers, since Black's pawn strongpoint at K4, holding back the tide, cannot be dissolved. The computer knows enough to occupy the center, but it clearly does not yet know that occupation is not equivalent to control.

11 . . . B-K3!
[*see diagram d*]

White cannot exchange bishops on K6, allowing Black operation along the king bishop file. The advantage of doubled pawns is often illusory. In this case the advantage would accrue to Black, as the newly doubled pawn would control vital squares (Q4 and KB4). Aware of these concepts, Black forces the exchange of White's only active piece, his king bishop. Question: Why does the computer not take the opportunity to double its opponent's pawns?

12 P-Q4 BxB 13 QxB . . .

If the computer is unwilling to capture doubling its own pawns unless specifically instructed, then it must learn that doubled pawns are at times an asset. Here, however, the move given is best.

13 . . . B-N3 14 PxP?! . . .
[*see diagram e*]

White should, temporarily at least, strive to maintain the pawn chain in the center with 14 B-K3; for example, 14 B-K3 Q-B3, 15 QR-Q1 KR-K1, 16 Q-B4 PxP, 17 BxQP! Q-K2, 18 KR-K1, and White is not worse off.

14 . . . PxP 15 R-K1? . . .
[*see diagram f*]

Development with 15 B-Q2 followed by QR-Q1 leads to a level game in which neither side wrests control of the queen file from a mutually vigilant opponent. Also playable is the more ambitious 15 N-Q2, intending N-B4 and P-QR4-R5, rooting out the annoying bishop. Notice that even if Black should occupy the queen file with his rooks, the occupation would be of limited value, since he could not immediately penetrate to the seventh rank. Amazingly enough, White is in no danger after 15 N-Q2 Q-Q6, 16 Q-B4 Q-N6, 17 N-B3! (ruling out 17 . . . N-B5 or R5) QR-Q1, 18 Q-K2!, with the well-concealed threat of 19 K-R1.

15 . . . Q-B3 16 B-K3 BxB

Not only is Black stymied but also White stands well after 16 . . . N-B5, 17 BxN QxB, 18 P-QR4 P-QR4, 19 QR-Q1 KR-Q1, 20 R-Q5!.

17 RxB• P-N3 18 R-Q1?! . . .
[*see diagram g*]

Correct is 18 P-N3 (limiting Black's knight) QR-Q1, 19 K-N2 (avoiding 19 R-Q1 RxR, 20 QxR Q-K3!, winning a pawn) R-Q2 (Q3 may be better), 20 R-Q1! KR-Q1, 21 RxR (or R(3)-K1) RxR, 22 Q-R4 P-B3, 23 Q-R6 Q-Q1, 24 P-QR4. White now has ample play in spite of Black's occupation of the queen file.

18 . . . QR-Q1 19 R(3)-Q3? . . .
[*see diagram h*]

The machine is still unaware of the necessity of restricting enemy minor pieces, preventing their invasion. Correct was 19 RxR RxR, 20 P-N3, and again White has little to fear.

19 . . . RxR 21 R-K3 . . .
20 RxR N-B5
[*see diagram i*]

The rook must vacate the queen file to guard brother knight. The Black knight, which now cannot be conveniently ousted, threatened NxPch. White would then have been unable to capture the intruder without losing his own knight. Of this tactical nuance the machine seems aware.

21 . . . R-Q1 22 Q-B4?! . . .
[*see diagram j*]

Bent on aggression, the machine cannot fathom the vital maneuver 22 K-R2 (defending the king rook pawn) and 23 P-N3 (prodding the enemy steed from its post).

Simply put, the move is a sophisticated blunder.

22 . . . Q-N3! 23 N-K1 . . .

The computer may have seen that 23 P-KN3 is no better, since then . . . NxPch, 24 K-N2 (if 24 K-R2 NxP) N-B5ch, 25 K-N1 (25 K-R2 Q-R4ch) R-Q8ch, 26 R-K1 RxRch, 27 NxR N-R6ch, 28 K-N2 N-N4. Black therefore wins just as easily as in the game.

23 . . . R-Q8 25 Q-B4 P-QB3
24 Q-B1 R-Q7 26 P-QN4 . . .
[*see diagram k*]

If the machine could be said to hope, then the last move could be considered the typical reaction of the ordinary chess buff who, realizing his king side is lost, preserves his queen side and hopes for a miracle!

26 . . . P-N4 27 Q-N3 . . .

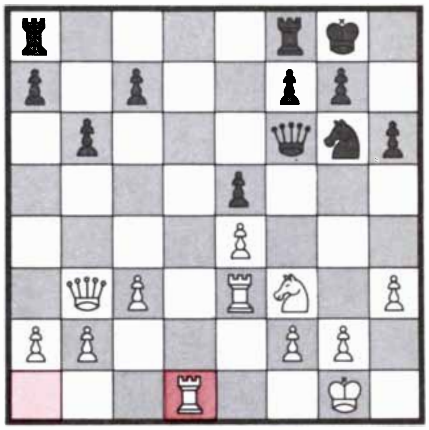
If 27 Q-B5, R-Q8 again!

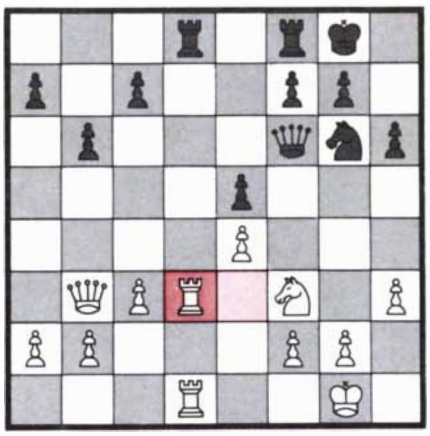
27 . . . Q-N4 28 P-B4? . . .
[*see diagram l*]

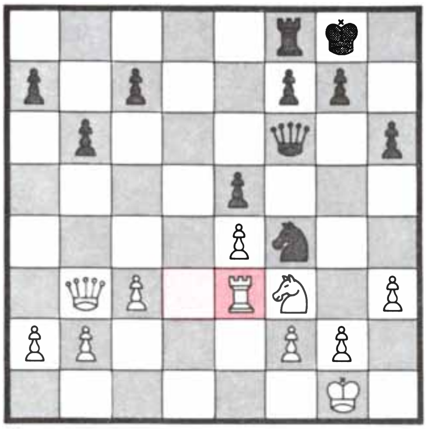
Although the computer hangs on doggedly, offering prolonged resistance, it still does not offer the best defense. Correct is 28 K-B1 Q-R5, 29 R-B3, with


e POSITION AFTER 14 PxP?!

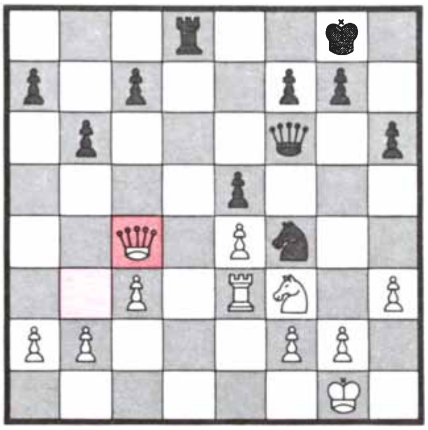

f POSITION AFTER 15 R-K1?


g POSITION AFTER 18 R-Q1?!

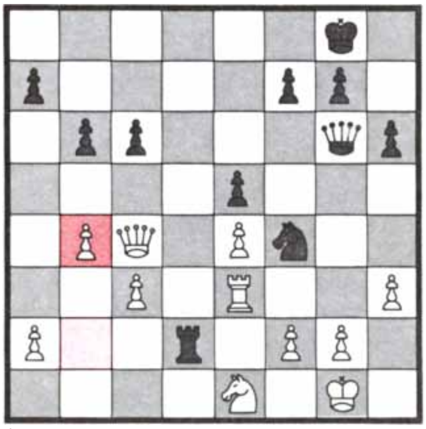

h POSITION AFTER 19 R(3)-Q3?



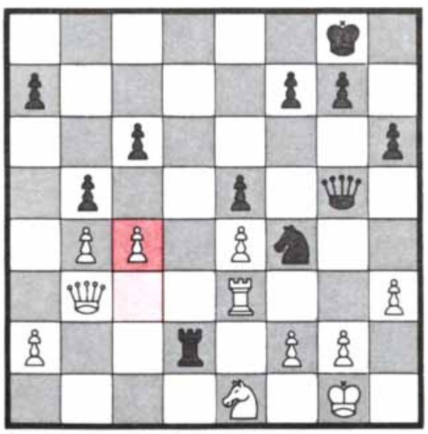
i POSITION AFTER 21 R-K3



j POSITION AFTER 22 Q-B4?!



k POSITION AFTER 26 P-QN4



l POSITION AFTER 28 P-B4?

P-B4 in the offing. Black would then have to take his courage in his hands and play P-KN4-N5, or find another means of applying the screws, such as 29 ... Q-Q1.

28 ...	Q-R5	29 R-KB3	...

Or 29 P-N3 NxPch, 30 K-N2 RxPch and so forth.

29 ...	R-K7	32 N-B2	Q-N4
30 Q-Q1	RxKP	33 Q-KB1	R-B5!
31 PxP	PxP	34 R-KN3	Q-R4!

[*see diagram o*]

At this stage a strong player would resign. Weaker players (the computer as well?) go on hoping for stalemate, even against a senior master.

35 N-K1	N-K7ch	37 KxN	R-B8
36 K-R2	NxR	38 P-B3	Q-N4ch

[*see diagram p*]

Did the computer resign here, or did the programmer perform this task?

It is a remarkable game for the computer, particularly against a topflight master. My guess is that David Levy will have his hands full in 1978.

Analysis of Game between U.S.C. Computer and Kalme

by Charles I. Kalme

1 P-K4	P-K4	4 O-O	P-Q3
2 N-KB3	N-QB3	5 P-KR3	...
3 B-B4	B-B4		

[*see diagram b*]

This is a move not generally played by a machine without special instructions. We do have a pattern that guards against getting into a pin that cannot be readily broken.

5 ...	N-B3	7 P-Q3	P-KR3
6 N-B3	O-O	8 N-Q5	...

[*see diagram c*]

As was pointed out above, machines like "aggressive" moves, and here it has nothing to strike at, so that 8 N-Q5 seemed to it as good as any.

8 ...	NxN	10 B-N3	N-N3
9 BxN	N-K2	11 P-B3	...

The program now is sophisticated enough to execute very simple positional ideas such as creating pawn chains, striving for control of the center and so forth.

11 ...	B-K3	14 PxP	PxP
12 P-Q4	BxB	15 R-K1	...
13 QxB	B-N3		

[*see diagram f*]

The machine apparently does not like 15 R-Q1 because it has instructions to clear the back rank of minor pieces and it would like to develop the bishop to K3 rather than Q2. In the process it gets maneuvered out of the queen file, but that point is too delicate for it to see.

15 ...	Q-B3	17 RxB	P-N3
16 B-K3	BxB	18 R-Q1	...

[*see diagram g*]

The machine is aware of the value of controlling open files, and now that its development is complete it goes after the queen file.

18 ...	QR-Q1	20 RxR	N-B5
19 R(3)-Q3	RxR	21 R-K3	...

[*see diagram i*]

The machine sadly has to abandon the queen file owing to the threat of ... NxPch if the White knight is left unprotected. This tactical point the machine sees.

21 ...	R-Q1	22 Q-B4	...

[*see diagram j*]

This is a reasonable move if one takes into account that the machine cannot anticipate that it will get tied down because of threats against its king. The maneuvers needed to achieve this state of affairs are too delicate for the machine to uncover, and it proceeds with an aggressive move, attacking the bishop pawn, where an untangling of its king side might be in order.

22 ...	Q-N3	23 N-K1	...

The machine does see that 23 N-R4 Q-N4 will not do.

23 ...	R-Q8	24 Q-B1	...

[*see diagram m*]

The machine is aware of the tactical threats and finds the only defense.

24 ...	R-Q7	25 Q-B4	...

Not being able to defend its knight pawn, the machine counterattacks Black's queen bishop pawn.

25 ...	P-QB3	26 P-QN4	...

[*see diagram k*]

This move together with the preceding one is a rather nice maneuver for the machine to execute in order to save the queen-side pawns, even though this may not seem like a great achievement with its king in danger. However, the machine probably does not feel too insecure there, although it should!

26 . . .	P-N4	28 P-B4	. . .
27 Q-N3	Q-N4		

[*see diagram l*]

This is a typical machine defense, which in principle would be good. It is aggressive and brings the queen back in touch with the other pieces but fails to meet Black's threat. The move 28 K-B1, which appears to be the only way to cope with the immediate threats, is probably far down on the list of moves to be considered, and such a move would be made only if it were necessary to meet a clearly perceived threat. The ensuing sequence apparently is too delicate for the machine to see as an immediate threat at this point.

28 . . .	Q-R5	29 R-KB3	. . .

This is the only move.

29 . . . R-K7

Now White's game crumbles.

30 Q-Q1	RxKP	32 N-B2	Q-N4
31 PxP	PxP	33 Q-KB1	. . .

The machine does see that 33 R-N3 loses the exchange to 33 . . . N-K7ch.

33 . . .	R-B5	34 R-KN3	. . .

[*see diagram n*]

By now, with every move losing material, it is not quite clear if the machine sees the ensuing loss of the exchange. It is aware that a knight move anywhere but to K1 allows . . . R-B8 winning the queen, and since 34 N-K1 leaves the QNP loose, temporarily the move chosen is the only one to keep the present balance.

34 . . . Q-R4

[*see diagram o*]

The double threat on the N and N-K7ch gains more material.

35 N-K1 . . .

The machine still sees that . . . R-B8 too is in the air.

35 . . .	N-K7ch	37 KxN	R-B8
36 K-R2	NxR	38 P-B3	Q-N4ch

[*see diagram p*]

The game was abandoned here as being hopeless for White.

Concluding Remarks

Of the many games played by the machine against the author this is by far the best and shows many interesting features, both positive and negative.

First for the positive features:

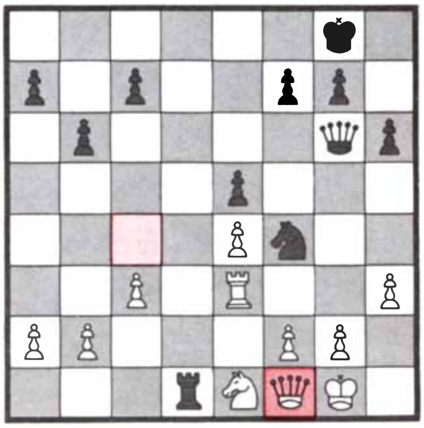
Taking into account that its opponent is a senior master with a rating in the same range as many grand masters, and that the opponent played as well as he could (in playing the game over I cannot see where I would have varied my play even in a serious tournament against a respected opponent), the machine put up a very good defense. It met every tactical threat perfectly, never losing material until it was maneuvered into a position where that was inevitable, and it actually succumbed only to quite sophisticated positional maneuvers. Indeed, an expert might have defended no better, except that at some stage he might have realized that getting so cramped was hopeless anyway and would have sought counterplay at the expense of, say, a pawn, whereas the machine knew no better than to hang on to everything no matter what the cost in position.

Lest the reader think we are claiming expert-level play for our program, let us now look at the negative features:

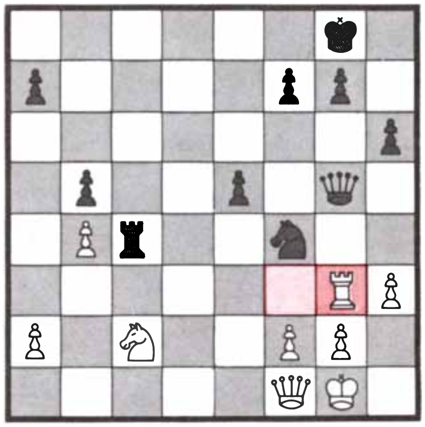
1. The machine was never really in the game and had no viable plan beyond the defense it was forced into. It is easy to make the best moves, in the sense of losing most slowly, if these entail meeting immediate tactical threats.

2. Delicate threats, where the real tactical threats come at the end of a sequence only a few moves deep, were beyond the machine's immediate perception and so it did not prepare its defense ahead of time, reacting only when it was too late. (It takes a good player, however, to create and execute such threats.)

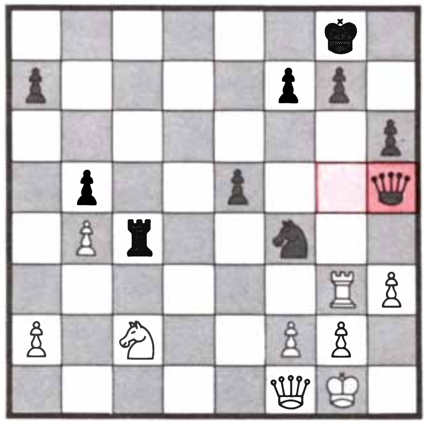
3. The method of attack by Black was particularly suitable for making the machine look good. Although, as I have mentioned, I would have made the same moves against a serious opponent, against a weak player I might have mixed it up more. Indeed, I can crush the machine tactically by choosing wilder lines of play. I still have to play the attacks fairly well; just any unsound attack will not do!



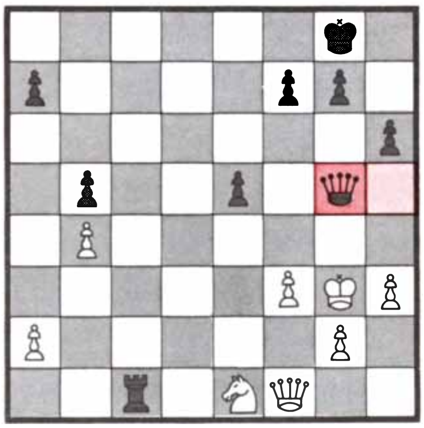
m POSITION AFTER 24 Q-B1



n POSITION AFTER 34 R-KN3



o POSITION AFTER 34 . . . Q-R4



p POSITION AFTER 38 . . . Q-N4 CH

MATHEMATICAL GAMES

Plotting the crossing number of graphs, and answers to last month's miscellany

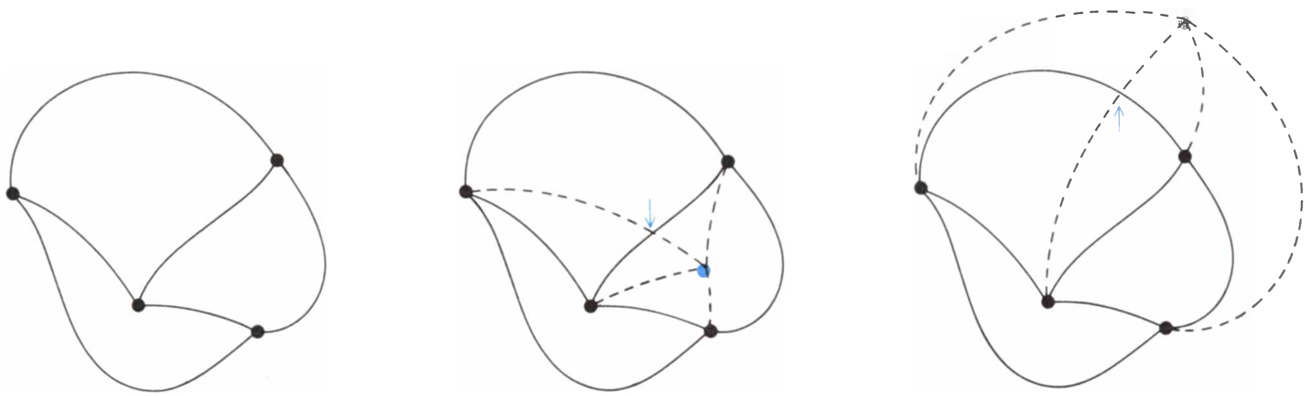
by Martin Gardner

Modern graph theory is raising many curious questions that appear to be simple but turn out to be extraordinarily complex. An entertaining class of such problems, some the basis of classic puzzles, are those that have to do with "crossing numbers." As Paul Erdös and Richard K. Guy wrote in "Crossing Number Problems" (*The American Mathematical Monthly*, Volume 80, 1973, pages 52–58): "Almost all questions that one can ask about crossing numbers remain unsolved."

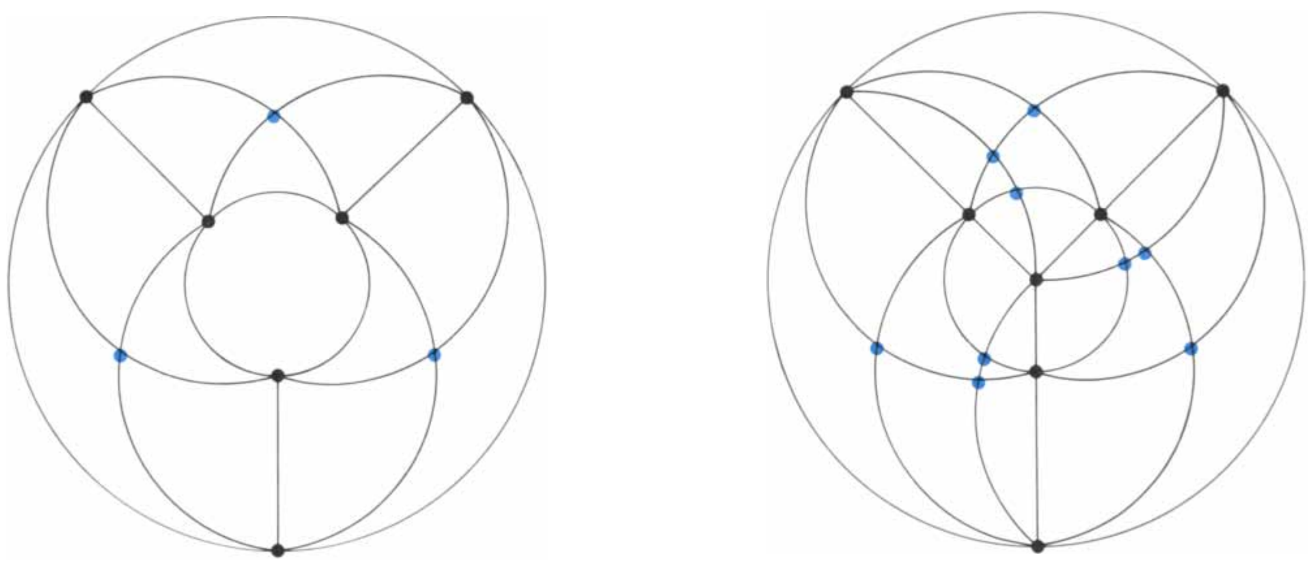
Before explaining what a crossing number is, a few fundamental terms must be defined. A graph is a figure consisting of points and lines connecting some of the points. The points are called nodes (or vertices) and the lines are called edges (or arcs). Only the graph's topological structure is significant. Think of the nodes as little spheres joined by elastic strings. Two graphs may look quite different, but if they represent two ways of placing the same ball-and-string model on a surface, they are considered identical.

Where two edges intersect at a point other than their nodes, the point in common is called a crossing. A graph can always be drawn so that no edge crosses itself or crosses an edge joined to one of its nodes, and so that no more than two edges go through any one crossing. Such a drawing is called a good drawing. When a good drawing is designed so that the number of crossings is as small as possible, the minimum number is called the crossing number of that graph.

To make this clearer, consider what is called the complete graph for n points. This is a graph on which every pair of nodes are joined by one edge. It is obvious that the crossing number for complete graphs of one, two and three points is 0, and it takes only a moment of pen-



Proof that the complete five-point graph has a crossing number of 1

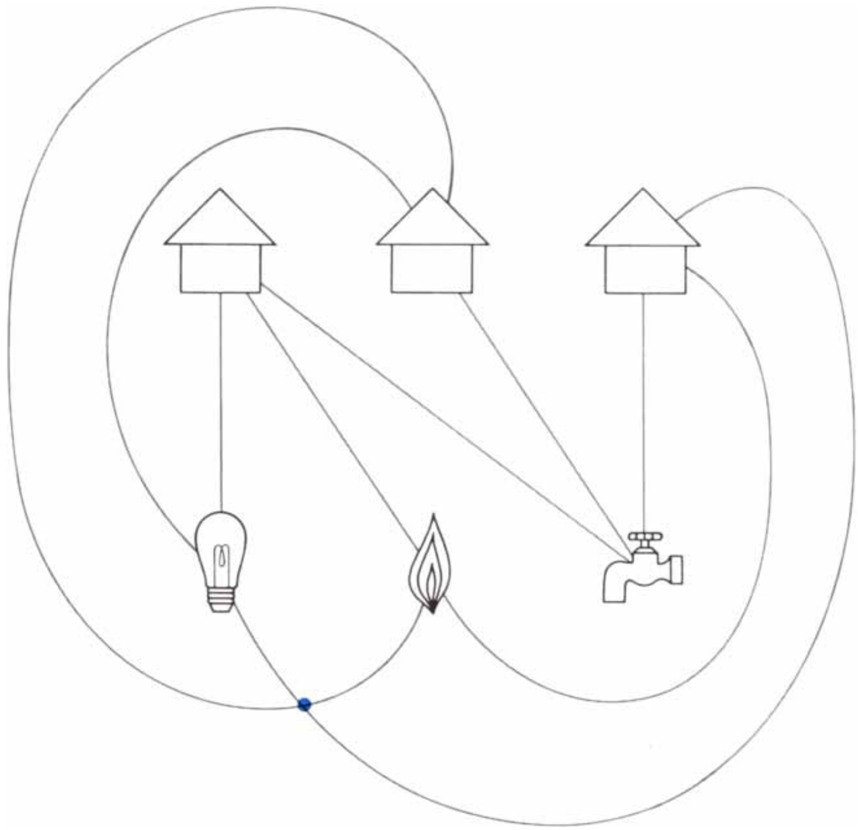


Six-point graph (left) *and seven-point graph* (right) cil doodling to find that it is also 0 for four points. A graph with a crossing number of 0 is called a planar graph. The simplest nonplanar graph is the complete graph for five points. It has a crossing number of 1. This means that, try as you will, you cannot join all pairs of the five nodes without producing at least one crossing. This can be proved informally as follows. All forms of the complete graph for four points consist of three mutually contiguous regions [*see upper illustration on opposite page*]. A fifth point (shown in color) must go either inside one of the three regions or outside the entire figure. When the fifth point is inside, you cannot connect it to the node outside its region without crossing an edge. When the fifth point is outside, you cannot get from it to the interior node without crossing an edge. The crossing is indicated by the arrow. (For a different proof see page 94 of *Graphs and Their Uses*, by Oystein Ore, an excellent introduction to graph theory.)

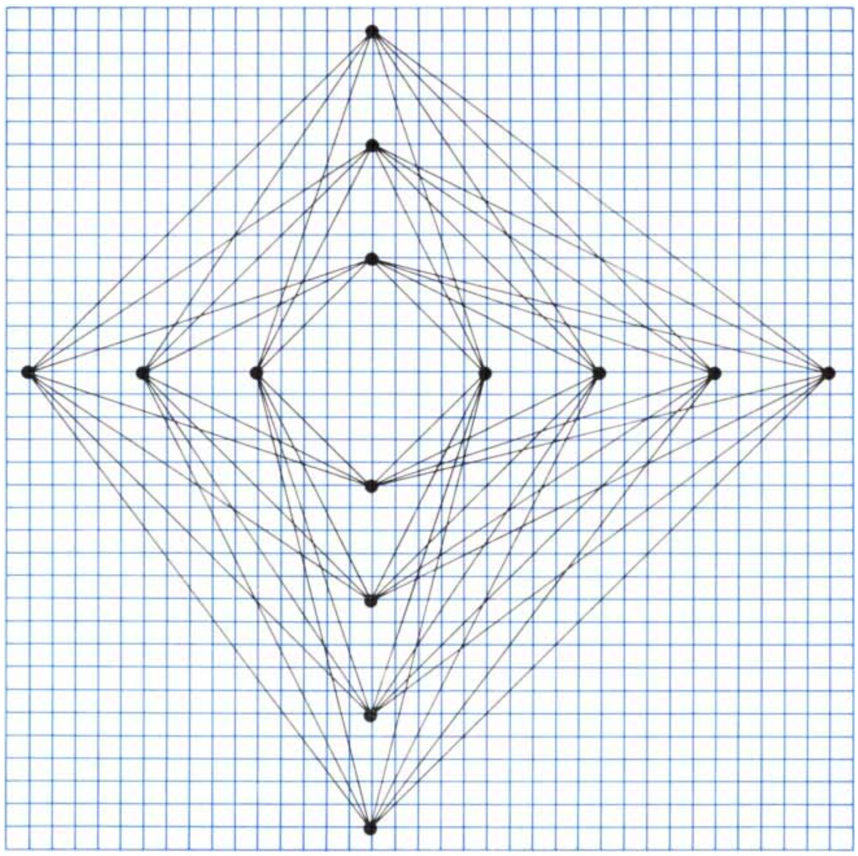
The fact that the complete graph for five points is not planar establishes that a map of five regions cannot be drawn so that every pair of regions share a boundary. If such a map could be drawn, we could put a point inside each region, connect each pair of points by an edge that crosses the border shared by the two regions containing those points and do this without creating crossings. In other words, we would be able to draw a complete graph of five points with a crossing number of 0. As we have seen, that is impossible. Unfortunately this does not prove the famous four-color map theorem. Several times a year someone sends me what he believes is a proof of the four-color theorem when actually it is no more than a proof, by way of the five-point graph, that no five regions can be mutually bordering.

It is true that for any map, say of many hundreds of regions, any specified set of five regions can always be colored with four colors without having two adjacent regions of the same color. It is conceivable, however, that five colors might still be required for the entire map. If you try to color it with four, there may always be a place where you run into trouble. Eliminate the trouble at that place by recoloring regions and the trouble always pops up at some other place. That five regions cannot mutually touch was established long ago, but the four-color map theorem, an altogether different matter, is still unsolved.

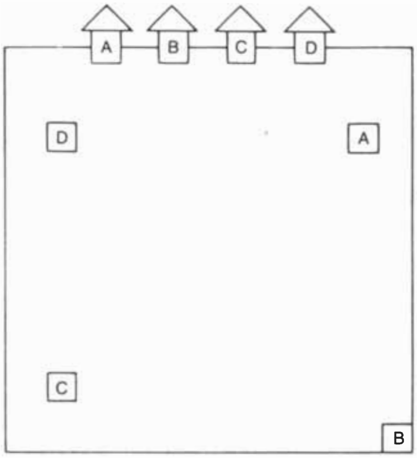
One might suppose it would be simple to write a formula for the crossing number of a complete graph of *n* points, but



The utilities problem



The complete 7,7 bipartite rectilinear graph (81 crossings)



The four-schoolhouses problem

this too is unsolved. In 1960 (writing on "A Combinatorial Problem" in *Nabla,* a bulletin of the Malayan Mathematical Society, Volume 7, pages 68–72) Guy conjectured that the formula is

$$\tfrac{1}{4}[\tfrac{1}{2}n]\,[\tfrac{1}{2}(n-1)]\,[\tfrac{1}{2}(n-2)]\,[\tfrac{1}{2}(n-3)]\,,$$

where the brackets indicate that the number inside is rounded down to the nearest integer. As far as I know this has been verified only for n through 10.

Complete graphs for six and seven points with crossing numbers of 3 and 9 respectively are shown in the lower illustration on page 106. The graph for six points (like that for five) is unique, but there are five variations of the seven-point graph. They are dissimilar in the sense that, if you regard the graph as being embedded in the plane, you cannot change one to the other (in terms of our ball-and-string model) without lifting a ball off the plane to carry it over an edge or a node.

Complete graphs of eight, nine and 10 points are known to have crossing numbers of 18, 36 and 60 respectively, as given by Guy's elegant formula. The eight-point graph has three variants and the nine-point graph has about 200. Several mathematicians have shown that Guy's formula is an upper bound for all larger n, but a proof of the formula is still elusive.

An interesting side question occurred to Guy. Can a complete graph with the minimum number of crossing points always be drawn by restricting the edges to straight line segments? He found that the answer is yes for seven or fewer points and also for nine points, but for eight points the rectilinear crossing number (as it is called when all edges are straight) is 19, not 18. Little is known about rectilinear crossing numbers for complete graphs of more than nine points, although it has been proved that for 10 or more points the rectilinear crossing number is greater than the crossing number. It has been conjectured that the 10-point graph has a rectilinear crossing number of 63.

Here is a pleasant little problem for which a very simple polynomial formula is readily available: What is the maximum number of edges that can be drawn, as part of a complete graph for n points, without a crossing? (Example: For the six-point graph the maximum is 12.) The answer will be given next month.

A formula for the crossing number of complete bipartite graphs of m and n points also has not yet been discovered. Such a graph has each point in set m joined to each point in set n, but no edges connect an m point to an m point or an n point to an n point. Complete bipartite graphs with points of 1,1, 1,2, 2,2 and 2,3 have crossing numbers of 0. The 3,3 graph, known as the Thomsen graph, has a crossing number of 1.

Students of recreational mathematics will at once recognize the 3,3 case as the old "utilities puzzle," so called because Henry Ernest Dudeney presented it with the following story line. There are three houses and three utility sources: water, gas and electricity. The puzzle is to draw lines connecting each house to each utility without any crossings. It cannot be done because the crossing number of this graph is 1 [*see top illustration on preceding page*].

The best conjecture (made by K. Zarankiewicz in 1954) for the crossing-number formula of a complete bipartite graph is

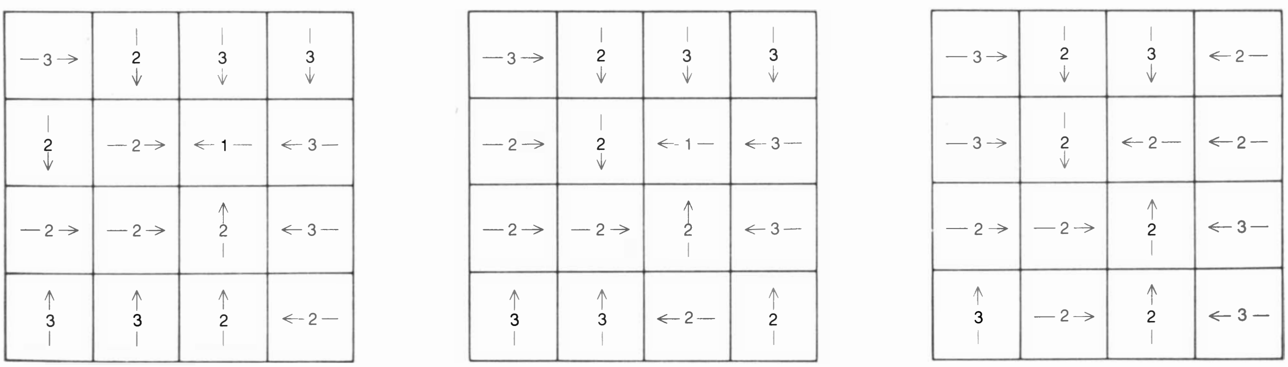
$$[\tfrac{1}{2}m]\,[\tfrac{1}{2}(m-1)]\,[\tfrac{1}{2}n]\,[\tfrac{1}{2}(n-1)]\,.$$

The formula applies to all values of m and n through 6, but beyond this all that can be said about this formula is that it gives a lower bound. This was proved by Daniel J. Kleitman in his paper "The Crossing Number of $K_{5,n}$" (*Journal of Combinatorial Theory,* Volume 9, 1970, pages 315–323). The crossing number for the 7,7 graph is not known. Using Zarankiewicz' formula, together with other arguments, Kleitman proved that the crossing number for the 7,7 graph must be 77, 79 or 81. He ends his paper with the one-word sentence "Which?"

A rectilinear graph from the article by Guy and Erdös mentioned earlier for the 7,7 case with 81 crossings is shown in the bottom illustration on the preceding page. So far this construction method (in which each set of points is arranged in a straight line, and the two lines are perpendicular) has produced rectilinear graphs that give the lowest crossing numbers known. No one has yet proved that it always does so, although Kleitman told me he believes it does.

As in the case of complete graphs, it is not hard to find a simple polynomial expression for the maximum number of edges that can be drawn, as part of a complete bipartite graph of m,n points, without a crossing. (Example: For the 3,3 graph the maximum is 8.) Can the reader give the formula before seeing next month's answer?

Some recent work has been done on the crossing numbers of other types of graphs, notably graphs on the plane for



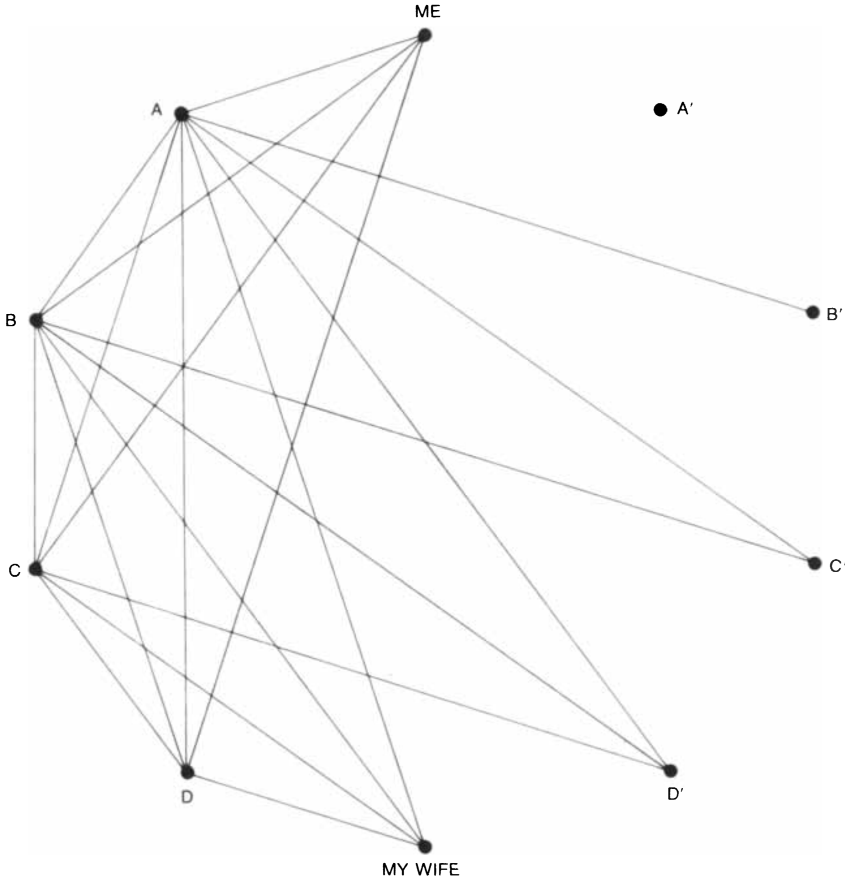
Answer to arrow tours

the skeletons of n-dimensional cubes, and complete graphs and complete bipartite graphs on such surfaces as the torus, the Klein bottle and the projective plane. (Graphs drawn on a sphere are the same as those drawn on a plane, because the sphere can be punctured at any spot not on the graph and flattened to a plane without altering the graph's topological structure.)

Richard K. Guy, Tom Jenkyns and Jonathan Schaer, in their paper "The Toroidal Crossing Number of the Complete Graph" (*Journal of Combinatorial Theory,* Volume 4, 1968, pages 376–390), prove that the toroidal crossing numbers for seven, eight, nine and 10 points are 0, 4, 9 and 23 respectively. (The zero crossing number for seven points on the torus corresponds to the fact that a maximum of seven regions, mutually bordering, can be drawn on the torus.) For 11 points, 42 is strongly believed to be the toroidal crossing number. The best results known for 12, 13, 14, 15 and 16 points are 70, 105, 154, 226 and 326 respectively. The paper gives upper and lower bounds for n greater than 9.

Guy and Jenkyns, writing on toroidal crossing numbers for complete bipartite graphs (*Journal of Combinatorial Theory,* Volume 6, 1969, pages 235–250), give upper and lower bounds for sufficiently large m and n. The toroidal crossing number for a 3,3 graph is 0, which means that the utilities graph is solvable on the torus. Indeed, the authors prove that it can be solved even if there are four houses and four utilities. The toroidal crossing number is also 0 for graphs of 3,4, 3,5 and 3,6. Complete bipartite graphs of 4,5, 5,5, 5,6 and 6,6 have toroidal crossing numbers of 2, 5, 8 and 12 respectively. An interesting classroom project is to find ways of drawing these graphs, and those in the preceding paragraph, on the surface of a large model of a doughnut.

Old puzzle books contain many problems based on crossing numbers. Here is an easy problem from one of Dudeney's books. Four boys lived in four houses and went to four schools. Show how the boy in house A can walk to school A, boy B to school B, boy C to school C and boy D to school D without any of their paths crossing one another or going outside the large square boundary [*see top illustration on opposite page*]. Of course there must be no tricks such as running a path through a house or a school. An answer will appear here next month.



Answer to the handshaking problem

Last month I presented a collection of miscellaneous problems and puzzles. Here are the answers:

1. The three ways of forming maximum-length arrow tours on the 4-by-4 field are shown in the bottom illustration on the opposite page.

2. Among the five married couples no one shook more than eight hands. Therefore if nine people each shake a different number of hands, the numbers must be zero, one, two, three, four, five, six, seven and eight. The person who shook eight hands has to be married to whoever shook no hands (otherwise he could have shaken only seven hands). Similarly, the person who shook seven hands must be married to the person who shook only one hand (the hand of the person who shook hands only with the person who shook eight hands). The person who shook six must be married to the person who shook two, and the person who shook five must be married to the person who shook three. The only person left, who shook hands with four, is my wife [*see illustration above*].

3. An 11-sided convex polygon can be formed with unit-sided squares and equilateral triangles, as shown in the top illustration on the next page. The angles possible for a convex polygon formed with the pieces are 60, 90, 120 and 150 degrees. To make a polygon with the maximum number of sides, all angles must be 150 degrees and the number of sides will be 12.

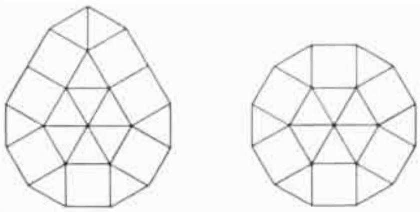
4. Only the ninth statement is true.

5. One solution for building a pentomino fence of arbitrary shape to enclose a "farm" of 128 square units is shown in the second illustration from the top on the next page. It is not unique.

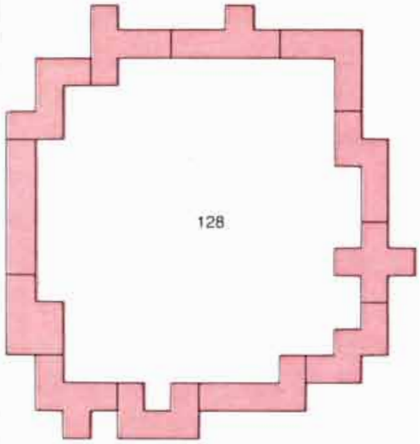
6. A square table can always be placed somewhere on a wavy floor with all four legs touching the floor. To prove this, put the table anywhere. Assume that only three legs, A, B, C, are on the floor and D is off [*see third illustration from top on next page*]. It is always possible for three legs to touch the floor because three points, anywhere in space, mark the corners of a triangle. Rotate the table 90 degrees around its center, keeping legs A and B always on the floor. This brings the table to a position where C is now the only leg that does not touch the floor.

During the rotation D has moved to the floor and C has left it. But D must have touched the floor before C left, otherwise there would be a position at which only A and B would touch the floor, and we know that it is always possible for three legs to touch. At some point in the rotation, therefore, all four legs must have been in contact with the floor. A similar argument can be applied to wobbly rectangular tables by giving them 180-degree rotations.

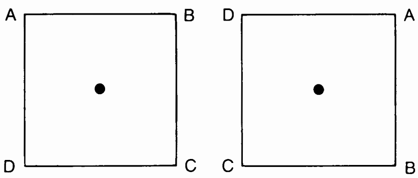
7. To put a chicken-wire pattern of creases into a small sheet of paper, first roll the sheet into a tube about half an inch in diameter. With the thumb and forefinger of your left hand pinch one end of the tube flat. Keeping pressure on the pinch with your left hand, your right thumb and forefinger, now pinch the tube flat at a spot as close as possible to the first pinch, making the pinch at right angles to the first one. Press firmly with both hands, at the same time pushing the two pinches tightly against each other to make the creases as sharp as possible. Now the right hand retains its pinch while the left hand makes a third pinch adjacent to and perpendicular to the second one. Continue in this way, alternating hands as you move along the tube, until the entire tube has been pinched. (Children often do this with soda straws to make "chains".) Unroll the paper. You will find it hexagonally tessellated in a manner that is most puzzling to the uninitiated.



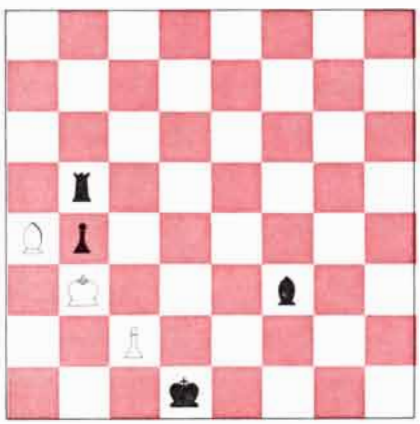
Eleven-sided and 12-sided convex polygons



The largest pentomino "farm"



The wobbly-table proof



Retrograde chess

8. To reach the chess position given last month, set the pieces as shown in the bottom illustration at the left and move as follows:

White	*Black*
1	B-Q4 (check)
2 P-B4	P takes P *en passant* (double check)
3 K takes P (check)	

Removing the white king will now leave the desired position.

9. The key to simplifying the three polypower equations is the basic law

$$(a^b)^c = a^{(b \times c)}.$$

Applying this to the first equation,

$$x^{(\overleftarrow{x\,x\,x})} = (x^x)^{\overrightarrow{x}}$$

$$x^{(\overleftarrow{x\,x\,x})} = x^{(\overrightarrow{x\,2})}$$

The two bottom x's are equal, therefore their parenthetical exponents are equal. Cancel the bottom x's, then repeat the procedure:

$$x^{(\overleftarrow{x\,x})} = x^2$$

The bottom x's again drop out, leaving $x^x = 2$, which gives x the value of 1.55961+.

The same procedure simplifies the second equation (down-4 equals up-4) to $x^x = 3$, and $x = 1.82545+$. Each succeeding equation increases the value of x^x by 1. The third equation reduces to $x^x = 4$, or $x = 2$.

The general procedure is to replace the down ladder by a number one less than the number of its x's and remove two x's from the up ladder. (Example: Down-5 = up-5 reduces to 4 = up-3.)

SCIENCE/SCOPE

Canada's Anik 2 communications satellite, launched in April, will provide additional capacity to meet the demand for Telesat Canada's commercial services. Like Anik 1, world's first domestic synchronous satellite, it has a capacity for more than 5,000 telephone circuits or 12 high quality color television channels. Hughes is building a similar series of satellites for two American companies that plan U.S. domestic systems to start in 1974 and 1975.

Several new composite materials developed for NASA space projects hold great promise for civilian use in such fields as biomedicine, agriculture, and furniture manufacturing, according to Hughes scientists who have just completed a technology utilization study for NASA. Medical applications include bio-compatible prosthetic devices of carbon fibers in an epoxy-resin matrix and braces of graphite-fiber-reinforced plastic (equal in strength to steel at one-eighth the weight).

Filament-wound graphite-glass fiber composites reduce the cost of grain silos and water tanks. Air-inflated barns and other structures made from fiber-reinforced plastics have a life expectancy of at least 20 years. Foams with fire-suppressant and self-extinguishing properties have been developed for furniture -- both hard foams for casting replicas of carved wood and soft foams for upholstering.

An Australian domestic communications satellite system is the subject of an extensive study by the Australian Post Office (APO). The system would distribute telephone, telegraph, television, and educational services throughout Australia's vast territory. Hughes was the successful competitor in a worldwide competition to provide consultant services to the APO for this study, which will initially determine the operation, benefits, and economic feasibility of the system.

Jacques Cousteau's oceanographic research vessel, Calypso, safely navigated the hazardous Drake Passage at the southern tip of South America with the help of an earth-orbiting sensor and a satellite built by Hughes. The multispectral scanner aboard NASA's Earth Resources Technology Satellite 1 photographed weather and iceberg formations along the route. The pictures were processed by Goddard Space Flight Center and relayed to the U.S. Navy's Fleet Weather Facility, which relayed the information to the Calypso via Applications Technology Satellite 3.

Hughes needs R & D Engineers to participate in the growth of our microwave and millimeter wave product lines. We have openings in transistorized microwave amplifiers, Gunn and Impatt diode sources, millimeter wave subsystems, traveling wave tubes for space applications, and high-voltage, high-efficiency microwave power supplies. Please write: B. E. Shryack, Hughes Electron Dynamics Division, 3100 W. Lomita Blvd., Torrance, CA 90509. An equal opportunity M/F employer.

1500 watts of electricity from solar energy are being produced by the FRUSA (Flexible Rolled-Up Solar Array) system aboard a U.S. Air Force Agena satellite. Its two 16-foot panels, each with 17,250 solar cells, were rolled into a cylinder at launch, then unfurled in space. Now Hughes is developing an advanced multi-mission version with an analog/digital voltage regulator. Welded aluminum construction will enable it to tolerate much higher temperatures during near-sun missions.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY

THE AMATEUR SCIENTIST

A laser beam and a photocell are used to measure the dirt content of water

Conducted by C. L. Stong

The purity of the water in rivers, lakes and ponds can be tested by determining the coliform-bacteria count or the presence of dissolved gases and of various compounds. Tests of this kind were described in this department in March, 1970, and February, 1971. During the past year a remarkably accurate instrument for measuring turbidity (the concentration of solid matter that is carried by the water in the form of solid particles) has been developed by a group of undergraduates at the University of Rochester. The group, which included Cathie Lubell, Thomas Barry and Gregory Hearn, worked in the Institute of Optics and Materials Science Program under the supervision of Edward M. Brody, assistant professor of optics. They describe the work as follows:

"It is possible to determine the turbidity of solutions by many techniques. With a microscope one can count the number of particles in a unit volume of the specimen. With a calibrated grid in the viewing field of the microscope one can measure the relative sizes of the particles in a polydisperse suspension (a suspension containing particles of widely varying size). Turbidity can be evaluated relatively by filtering the specimen and weighing the dried filtrate. All these procedures are time-consuming. Boredom and fatigue on the part of the observer can lead to error.

"Extinction turbidimeters are doubtless the simplest instruments that have been devised for measuring the concentration of solids in suspension. They are based on the principle that turbidity is inversely proportional to the minimum length that a column of fluid must have in order to extinguish at one end of the column a source of light at the other end. A major source of error in all such turbidimeters arises from the fact that changes in the range of particle sizes in the suspension affect the pattern in which the transmitted light is reflected from particle to particle. The effect is known as multiple light-scattering. It causes apparent turbidity to decrease as the length of the light path through the specimen is increased. The influence of the multiple-scattering effect becomes increasingly significant when the attenuation of light through the specimen exceeds 15 percent. Extinction turbidimeters must therefore be calibrated with suspensions of known particle size. Indeed, reliable measurements of turbidity cannot be made with extinction turbidimeters unless the distribution of particle sizes is known.

"The substitution in turbidimeters of lamps for natural illumination and of photocells for the eye has led to improved measurements, but many sources of potential error remain. Variations in voltage applied to the lamps, fatigue of the photocathode of photocells and the effects of space charge in photomultipliers can combine to introduce an error of at least 10 percent. Moreover, it has been shown that the inherent temperature sensitivity of photoelectric detectors causes the output to vary as much as 50 percent through the temperature range from zero to 30 degrees Celsius.

"To cope with such errors we built an instrument of the ratio type in which an error in one part of the system is in effect compensated for, keeping the ratio constant. We call the device a dual-beam turbidimeter. Apparatus employing dual-beam transmission is available commercially, but it is costly and needs modification for use as a turbidimeter.

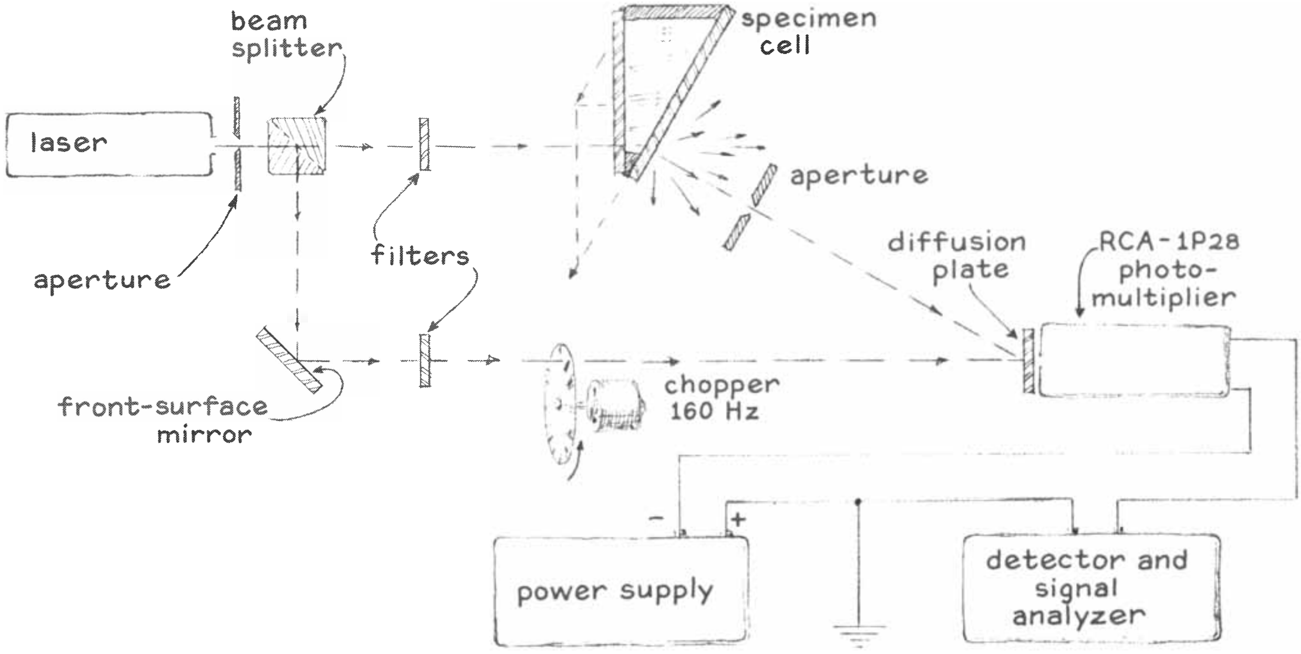
"In our device light from a single source is split into two beams. One beam traverses the specimen and the other, a reference beam, is transmitted through air. The two beams combine in a photodetector, where they are added algebraically. Experience has demonstrated that the instrument measures turbidity effectively whether the liquid contains suspended particles of widely varying size or suspended particles of uniform size. The results have been highly reproducible when checked against comparable measurements made by other procedures.

"A helium-neon laser serves as the light source [*see top illustration on opposite page*]. We elected to use the laser because it costs little more than a high-intensity lamp and provides parallel rays in the form of a monochromatic beam one millimeter in diameter. The output power of about two milliwatts exceeds the requirements of the photomultiplier detector by at least 100 times. We reduce the intensity by inserting neutral-density filters in the beams. An incandescent lamp could be substituted for the laser by inserting a color filter in the optical path (to select a range of wavelengths) together with simple collimating lenses to make the rays parallel.

"The output of the laser falls on a beam splitter, which is a rectangle of clear plate glass set at an angle of 45 degrees with respect to the beam. The transmitted portion of the beam passes through a neutral-density filter and enters the specimen cell. This vessel (a hollow prism made of sheet glass) has the form of a right triangle. The refracted rays emerge from the cell, pass through an aperture and fall on the photocathode of the photomultiplier tube. The intensity of these rays, which constitute the specimen beam, varies with the turbidity of the specimen.

"Rays that are reflected by the beam splitter fall on a front-surface mirror, from which they are reflected through a neutral-density filter and the apertures in a motor-driven chopper. The chopper interrupts the beam 160 times per second. Rays of the flickering beam recombine with those of the specimen beam at the photocathode.

"The output of the photomultiplier tube is a unidirectional electric current that varies in amplitude 160 times per second. A circuit separates the output into two parts: a direct-current signal and an alternating-current signal [*see bottom illustration on opposite page*]. The direct-current signal represents the sum of the current generated by light transmitted through the specimen plus



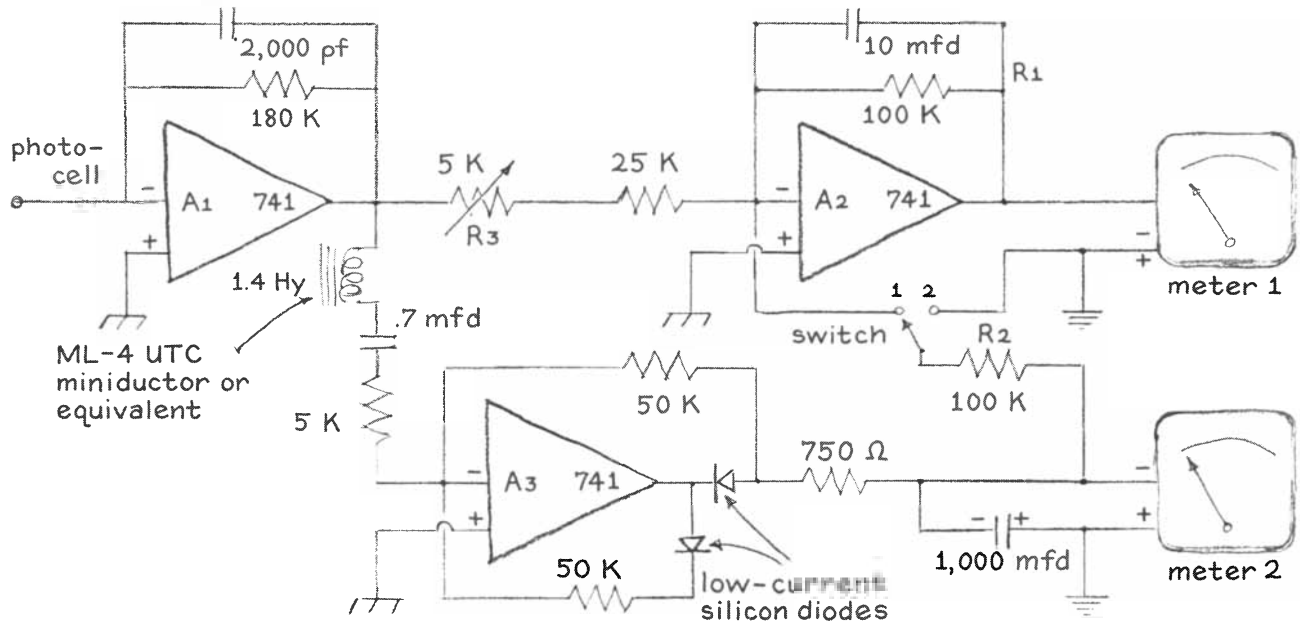
Top view of elements of the turbidimeter designed by a group at the University of Rochester

the direct-current component of the chopped reference beam. It is measured by meter No. 1. The amplitude of the alternating-current signal is proportional to the intensity of the reference beam. It is measured by meter No. 2.

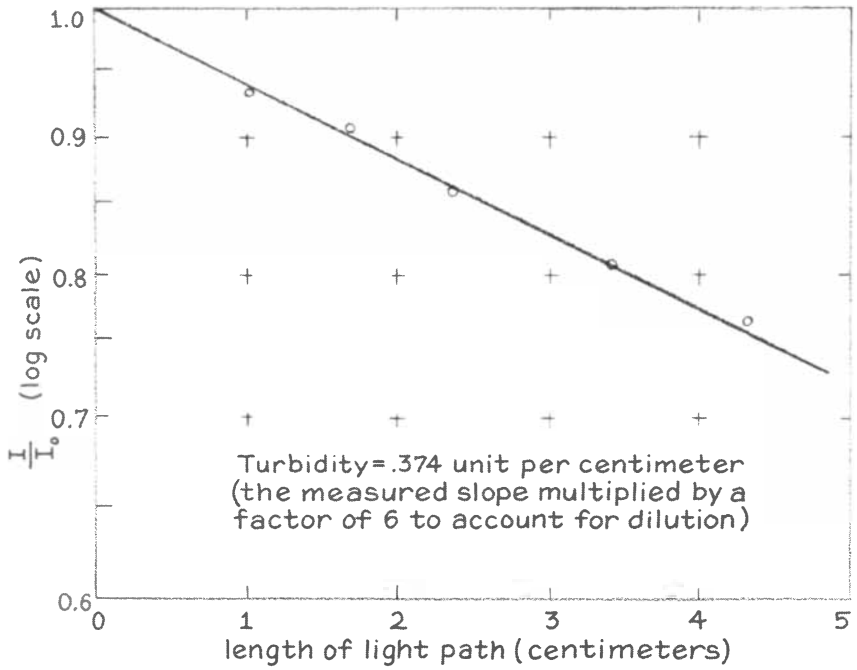
"The electronic circuit consists of a network that includes three operational amplifiers, which boost the output of the photomultiplier to a usable level. The alternating-current and direct-current components are amplified independently for display on separate meters. The filter that separates the alternating current from the direct current consists of an inductance of 1.4 henrys and a capacitor of .7 microfarad connected in series with the input of the reference-signal amplifier, A_3. The filter transmits maximum current at a frequency of 160 hertz. Both the capacitor and the inductor are available commercially.

"The frequency at which the chopper operates is not critical. Choppers can be improvised to operate at other frequencies. For example, a synchronous motor that operates at 3,600 revolutions per minute could be fitted with a disk that contains two apertures to chop the beam at a frequency of 120 hertz. The input circuit of the reference-beam amplifier could be tuned to this frequency by retaining the 1.4-henry inductance and increasing the capacitor to approximately 1.25 microfarads. The values of capacitors and inductances that resonate at other chopping frequencies can be calculated by the formula $1/6.28 \times (L \times C)^{1/2}$, in which L is the inductance in henrys and C is the capacitance in farads.

"The experimenter is primarily inter-



Circuitry of the turbidimeter



Turbidity of a water sample from Lake Ontario

ested in the intensity of the transmitted beam as it is displayed on meter No. 1. This meter, however, responds to both the direct current of the specimen beam and the direct current in the reference beam unless the instrument is appropriately adjusted. To make the adjustment connect switch No. 2 to the position shown in the circuit diagram, block off the specimen beam and vary the resistance of potentiometer R_3 until the pointer of meter No. 1 falls to zero. Unblock the specimen beam. The adjustment need be made only once a day.

"The electronic components of the circuit are quite stable. The switch is not strictly necessary. It merely serves as a check for determining that the electronic components are working properly. When the switch is in position 1 and the circuit is properly adjusted, the alternating-current portion of the reference signal is subtracted at the input junction of the A_2 amplifier, because operational amplifiers reverse the polarity of signals. When the switch is turned to the other position, at which the output of the reference-signal amplifier is connected to ground, meter No. 1 displays the sum of the reference and specimen voltages instead of the difference.

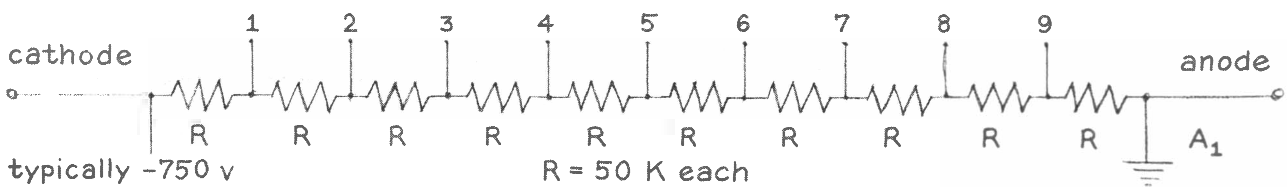
"After the instrument has been adjusted meter No. 1 indicates voltage proportional to the intensity of the beam transmitted through the specimen cell. Meter No. 2 similarly indicates voltage proportional to the intensity of the light source. An easy way to use the instrument is to fill the cleaned specimen cell with clear water and record the ratio of the two meter readings: M_1/M_2 (water). The specimen is then transferred to the cell and a second set of readings is made: M_1/M_2 (specimen). The transmission of the specimen, as calibrated against clear water, is then expressed by the ratio, (M_1/M_2) (specimen)/(M_1/M_2) (water). The apparatus automatically compensates for changes in the intensity of the light source, changes in the sensitivity or gain of the photomultiplier and most other variables that degrade the results of extinction turbidimeters.

"Ideally a specimen of water that is well mixed should be uniformly turbid, and a transmission measurement for one path length is sufficient to determine the turbidity. Experience suggests, however, that a series of turbidity measurements should be made through columns of fluid of varying length to improve the accuracy and as a check that only single scattering processes are involved. A plot of the logarithm of the transmission versus the length of fluid traversed should be a straight line. Systematic deviation from a straight line indicates multiple scattering and a difficulty in calculating the turbidity.

"Our specimen cell was designed expressly with this requirement in mind. As mentioned, it consists of a right triangular prism. The vessel is mounted to a traversing table with the faces of the prism perpendicular to the plane in which the transmitted and reference beams are located. The prism is moved in a direction parallel to the hypotenuse of the triangle. The light continues to enter the specimen cell at a right angle to the vertical face as the specimen cell is displaced in the plane of the hypotenuse, but the length of the path traversed by the light beam varies directly with the position of the cell. The cell is moved along the traversing table by a screw mechanism calibrated in increments of one millimeter.

"The four pieces of the cell were cut from ordinary window glass. Three rectangular sides were glued to a base piece to form a right-triangular prism. Before making the strips with a glass cutter of the wheel type we inspected the sheet and chose areas that were free of bubbles and striations. Surfaces of optical quality are not required because the measurements involve intensity, not image formation. Any convenient angle can be made between the hypotenuse and the side through which the beam enters. Our vessel was assembled with Sauereisen Adhesive Cement, an acid-proof cement that is distributed by the Fisher Scientific Company in Rochester, N.Y. Doubtless any epoxy cement would work as well if the cell is cleaned only with detergent.

"We clean the cell after most measurements with a solution of chromic acid made by dissolving from 30 to 50 grams of sodium dichromate or potassium dichromate in a liter of concentrated sulfuric acid. A comparable cleaning prep-



Circuitry of the photomultiplier tube

aration known as Chromerge is available commercially. It must be similarly diluted with concentrated sulfuric acid.

"To clean the cell wet the inside surfaces of the glass with the acid mixture, let the cell stand for five minutes and then rinse it thoroughly with filtered water. We prepare the rinsing water with a Millipore filtering apparatus. In our opinion cells need not be cleaned with chromic acid unless extreme accuracy is desired. For the routine assessment of turbidity in natural streams the vessel can be cleaned with any good household detergent.

"We found that the results of our transmission measurements were reproducible to within 1 percent. We also found that we had to dilute specimens from rivers and lakes about six to one, even when they were taken from calm water, to avoid significant multiple scattering of the light and a corresponding loss of accuracy. Usually severe multiple scattering occurs when turbidity attenuates the light more than 15 percent.

"Our results were checked by theoretical calculations of the turbidity of known samples and also by experiment. A typical experimental check was made with a specimen prepared by suspending polystyrene spheres two microns in diameter in filtered distilled water. We counted the number of spheres in a known volume of solution by means of a microscope and estimated the concentration to be between 100,000 and 200,000 particles per cubic centimeter. Subsequent measurements with our turbidity meter indicated an actual concentration of approximately 130,000 spheres per cubic centimeter. The test is based on the fact that concentrations of spheres of uniform diameter attenuate light by a predictable amount.

"Spheres of this kind are available commercially from a number of suppliers, including the Dow Chemical Company (Midland, Mich. 48640) and Particle Information Service (600 South Springer Road, Los Altos, Calif. 94022). The spheres come in the form of a concentrated solution. Dilute the solution by adding it to distilled water. This procedure prevents the spheres from clumping. We used one drop of the sphere solution in 250 milliliters of water for a 'stock' solution that was diluted further to conduct scattering experiments. We positioned the specimen cell near the bottom limit of its traverse for the maximum optical path length and diluted the sphere solution until the attenuation was less than 15 percent.

"To verify the concentration with a microscope we counted the particles against a translucent background in the form of a grid. If the focal length of the objective lens is known, the depth of the field can be computed. If the concentration is low, the microscope views particles in a single plane. We counted the number of particles per unit length of the field in one direction and cubed the result to find the concentration per unit volume.

"A polydisperse specimen was obtained at a depth of eight meters in Lake Ontario near the inlet of the Genesee River. Turbidity was measured after the specimen had been diluted six to one with distilled water. The results were expressed as a graph by plotting the logarithm of the ratio of the transmitted light (I) divided by the unattenuated intensity of the source (I_0) against the length of the path through the specimen that was traversed by the light beam [*see top illustration on opposite page*]. Turbidity is equal to the slope of the graph. In this example it amounts to .374 unit of turbidity per centimeter of path length.

"The apparatus can be used to measure the distribution of sizes and concentrations of particles, provided that the particles are assumed to be spherical in form so that they settle at a predictable rate. The rate at which the intensity would change with time could then be related to particle diameter.

"The physical details of the construction are left to the resources of the experimenter. The optical parts of our instrument were mounted to a metal base by rugged fixtures that were made with machine tools. Most of the parts could be made of wood or plastic. We enclosed the entire apparatus in a lightproof housing so that measurements could be made in a normally lighted room.

"Our apparatus is not readily portable. Moreover, the laser and the light chopper require a source of 115-volt, 60-hertz power. For this reason the apparatus cannot be used in the field. It should be possible to modify the apparatus for use with battery power by exchanging a certain amount of accuracy and precision for portability. For example, a photodiode or phototransistor could be substituted for the photomultiplier. High-intensity lamps that operate on 12 volts, together with color filters and collimating lenses, could replace the laser. It should also be possible to substitute for the motor-driven chopper an oscillating chopper improvised from a 60-hertz resonant switch of the kind formerly used in the inverter circuits of automobile radios."

BOOKS

The irrational uses of rational science, tropical monocot crops and bottled gases

by Philip Morrison

The Domination of Nature, by William Leiss. George Braziller ($6.95). The "cunning of unreason"—a Hegelian phrase turned upside down by a young critical philosopher—consists most fatefully in this: Modern science, cool, rigorous, wide-ranging, can explain the phenomena of the world in a way superior to any method available before the advent of its prophets, Francis Bacon and Descartes, and yet the hope that social development itself might come under rational control has consistently been frustrated. Instead the rationalism of science and technology remains "caught in the web of irrational social contradictions," from the B-52 to the Harlem gutter to the jammed freeway to the political prison camp.

This compact and tightly argued book (which counts among its virtues a brief exposition, if only by way of critique, of the chief positions of the German school of social philosophy connoted by such names as Edmund Husserl, Max Horkheimer and Max Scheler) lies at the boundary between the history of ideas and the critique of social thought. The author, who teaches political science in Saskatchewan, was armed for his task by his degree of philosophy under Herbert Marcuse at San Diego; he brings to this brilliant essay both the energy of deep engagement and a scrupulous honesty of argument. At a time when the conceptual source of the problems of our day is sought variously in Christian dogma, in the Galilean world view or in the size of the population, his book deals a more trenchant cut at the tangle than any currently at hand. The reader is warned: This book contains no statistics and no description; it centers on ideas, their origins and their inner meaning. For those who would follow the arguments the pages are intense. Here in review we can only suggest the development.

The Elizabethan sailors of Bacon's *New Atlantis* were lucky enough to witness the rare ceremonial visit to town of a director of the isolated research institute called Solomon's House, or the College of the Six Days' Works. Passing in "theocratic majesty," he is "a Man of middle Stature, and Age, comely of Person, and had an Aspect as if he pittied Men.... He was carried in a rich Chariott...Litter-wise...Horses...richly trapped in blew Velvett Embroydered.... Pannells of Sapphires... Emerauds of the Peru Colour.... He held up his bare Hand, as he went, as blessing the People, but in Silence." By implication the social arrangements of Bacon's Utopia became clear. The scientist is beyond the people, detached, aloof, bringing blessings by gesture (and in the rise of the gross national product, we may assume). He is an allegory of our troubled times, when "technological providence" brings blessings (and doom) that the passive citizens neither understand nor control. For Bacon the nature of the state was unaffected by science; research was no more democratic or progressive than the cathedral at Canterbury or the Vatican. The ethics of those researchers flowed not from science but from religion; the secrets of Solomon's House would not be given to the state save for right purposes, as judged by the austerely independent adepts, who "take all an Oath of Secrecy."

Science and technology ever since Bacon have come to seem the only path to the mastery of nature. It was not so earlier; man's cleverness was only the ape within his nature. Adam dominated the beasts of Eden morally by virtue of his closeness to God; the 17th-century watershed was crossed once science was understood as the objective study of objects that themselves were value-free. For Scheler "to conceive the world as value-free is a task which men set themselves on account of a value: the vital value of mastery and power over things." Thus science becomes *Herrschaftswissen;* positive science is the Nietzschean "knowledge...as a tool of power." For Leiss this view is mistaken, or at least incomplete. Science indeed progressively masters its subject, but technology does not therefore automatically master nature. Society must intervene.

Here enters the work of Husserl, who influenced Scheler in mid-career. Husserl saw—with Galileo—that there are two distinct types of nature: the nature we intuit and experience in everyday life and the mathematized atomic world that science perceives behind experience. Everyday nature has "been the object of mastery in every stage of human development." The Paleo-Indians killed off the big game; simple cultivators still burn the forests to clear land. The rational world of atoms does not rule but "submits to the conditions that define mastery of nature in the prevailing historical and social world."

If Horkheimer in his turn is correct, the world that man sees has become steadily "demythologized." Reason, the modern enlightenment, is as old as Athens. Slowly argument drives out all values except self-preservation, since the natural laws "possess no purpose man can fathom." Now the social context is left to determine the link between domination of nature and control over man. Social conflict binds the two elements; as long as our species had little material choice, human conflict remained self-limited. Today "the cunning of unreason" takes its revenge. "After centuries of benign neglect the population of the Sudan, for example, is suddenly offered advice and martial gifts by a medley of agents from Egypt, the Soviet Union, the United States, Libya, the German Democratic Republic, and Israel, while its neighbors in Ethiopia battle each other with American and Chinese armaments." Technology has unified the world!

We have seen such a degradation of ideals before. The natural rights of man suggest a great human goal, but by centuries of conflict they acquire another, more ideological identification. One side finds in them justification for corporate domination of the market, seen in terms of the half-vanished entrepreneur; the other sees only a cover-up for the ingressions of economic exploitation. On the one hand we find the tiger cages of "free"

South Vietnam; on the other, the psychiatric treatment given some Russian critics. In its turn "mastery of nature seems less a grand enterprise of the species than a means of upholding the interests of particular ruling groups." Which groups? Just about all of them.

It is clear to Leiss that it is not capitalism or Genesis that is the source of our trouble, nor is it Galilean science alone. Neither Shamanism up to date nor the peasants' ways will solve our modern problems. Socialist states, like capitalist ones, east and west, find easy convergence on the same view: "Mao Tse-tung Thought guides us in conquering Nature," says the *Peking Review*. We need new structures that reflect human realities; man is after all part of nature, albeit a special part. It is the relation between nature and humanity that we need to master. For success the majority of the species must become engaged in institutions that "encourage critical faculties among all men and women," this committed critic writes. "The dazed onlookers and the magnificent potentates alike" must transform themselves. Religion might provide traditional values, but the universality of the problems and the secularity of behavior today appear to make likely its continued ineffectualness. "To control their scientific and technological ingenuity men must first cease to be astonished by it and to request blessings which it is incapable of bestowing on them." Science is not our ruin, as it is not our panacea. Freedom remains, as another German philosopher wrote a century ago, the recognition of necessity.

TROPICAL CROPS: MONOCOTYLEDONS 1 AND 2, by J. W. Purseglove. Halsted Press Division, John Wiley & Sons, Inc. ($24). "The whole thing is a fascination," these columns remarked about what was only the first half of the full work. Now we really have the whole thing, and the fascination remains. These two volumes complete the story of tropical crops by treating the monocots just as the first two volumes, published nearly five years ago, discussed the dicots. Again a lucid and up-to-date set of articles is placed before the reader, spanning about 50 crops, from Job's tears and the Panama-hat plant to the sweet-smelling, orange-red fleshy fruit of the pandanus, long cultivated on the low coral islands of the Pacific to eat with the bland flesh and milk of the coconut.

Centering on a botanical account, each crop is fully described (and 32 of them are figured in detail in beautiful full-page line drawings done in Trinidad after living material by Marjorie Wong) with its husbandry, uses, economic importance and world trade, pests and the like. The author's interest in the origins and distribution of each crop is clear. "An author should be entitled to his foibles," and this delightful, scholarly failing gives the book even greater appeal to the general and reference reader.

Dr. Purseglove stoutly remains a Columbian; before the voyage there was "no movement of crops *by man* between the Old and the New World," although there is strong evidence that the floating coconut established itself "on open coasts without the aid of man." If coconuts were found on the Pacific coast of Panama before Columbus, and the case is not sure, they were ocean-borne. They drifted against the course of Thor Heyerdahl's famous raft, and northward of it. Perhaps long ago, before the rise of man, their ancient ancestors, very unlike the present delectable species, had left the South American lands where "the most numerous relatives of the coconut" flourish today, traveling afloat or via a then warm Antarctic continent.

A few more morsels of what one can learn are worth serving. The dicots jute and hemp provide the biggest part of natural fibers not used in clothing, but the "bulk of the hard fibers of commerce" come from the century-plant family: sisal and its relatives. Jute for sacking, hemp for twine, sisal for the coarse twines essential for the combine and baler, and the strongest, manila hemp (the fiber of the leaf stalk of a species of banana) for marine ropes—these comprise an industry that continues the tradition of the oldest human artifacts. Synthetic fibers, particularly polypropylene, threaten them all.

The tropical American pineapple, "scaly like a pine cone... but soft like a melon, surpassing every garden fruit and flavour," came by Columbus' own ships. Peter Martyr wrote the description cited, but he never himself tasted a pineapple. Only one survived the passage, and the King ate it! Now the fruit is plantation-grown from Australia to Kenya, mostly from a variety called Cayenne, taken from there to France around 1820. Once a single genotype, it is now a set of clones that differ by somatic mutations. The plant sets fruit without fertilization and commercial propagation is therefore always vegetative, although improvement breeding through cross-pollination is carried on. The normal pollinating agent is the hummingbird. In Hawaii there is no natural agent, and the introduction of hummingbirds is forbidden by law.

The fourth most important cereal crop in the world—after wheat, rice and maize—is sorghum. It is the staple in the drier parts of Africa, India and China. The wild plant is Ethiopian; domestication may have begun there some 5,000 years ago at the hands of farmers who could move to drier lands with this wheat-field weed. Its flour is too low in gluten to make good bread; the white varieties (with names such as milo, hegari and Guinea corn) are used for porridges and boiled dumplings; the darker, bitter ones make beer and those with large, juicy stems yield sugar syrup.

Airborne trade in cut orchid blossoms is a growing industry out of Hawaii and Singapore; the colorful orchids, of several species, often hybrid, are those that do well in the humid Tropics. They are seeded on sterilized artificial media and grown up poles in beds or in pots. The ice-cream eaters of the U.S. (with some help, mainly from France) support the Indian Ocean islanders who now maintain the orchid *Vanilla fragrans* among low shade trees.

For the sake of completeness, the volumes include sketchy treatments of bamboo and wheat, but crops such as sugar, millet, oil palm, banana, rice and maize are given the most serious attention. "There is a widespread belief that bamboos felled shortly after the full moon will suffer less damage from the bamboo-borer beetles" and that ripe bamboos are attacked less than unripe stalks. "Alas! it really is not so..../The lunar myth is utter tripe/And borers *like* their bamboo ripe." So rhymed two Trinidad experimenters 15 years ago.

The valuable index and list of species are found only in the second of these attractive volumes printed in England; the work really consists of two parts bound as four books. The price per page has gone up from less than 2.4 cents to almost four cents, an increase of about 11 percent per year, compounded continuously.

MATHESON GAS DATA BOOK, by William Braker and Allen L. Mossman. Matheson Gas Products, East Rutherford, N.J. ($17.50). EFFECTS OF EXPOSURE TO TOXIC GASES—FIRST AID AND MEDICAL TREATMENT, by William Braker and Allen L. Mossman. Matheson Gas Products ($5). GENERAL CATALOG 28. Matheson Gas Products (free). What these three proprietary references do is admit the general reader to an industry of surprising interest. Everyone knows the nature of the petroleum industry: from crude oil it refines a few major products such as gasoline on an enormous scale. But it also produces fine chemicals, the myriad delicately refined organics that trace their origin back to the coal seam or the wellhead. Crude air is similarly a major raw material for the heavy chemical industry. Its major refined products, made by the hundreds of millions of tons each year worldwide, include liquid oxygen for steelmaking and nitrogen for fertilizer. We who use natural gas or see the welder at work with his steel cylinders of oxygen and fuel and maybe notice the helium bottle when the kids buy balloons in the park do not, however, normally realize how thoroughly elaborated the fine-chemical industry of gases is. These are costly reagents, kept not in glass bottles with stoppers but in strong forged-steel cylinders under pressure in order to increase their density to practical values, thereby storing mechanical as well as chemical energy.

This specialized New Jersey firm with plants at a number of places in the U.S. and Canada and one in Europe is one of the leading enterprises in gaseous fine chemicals. Air is a crude source for them, but the entire domain of organic and inorganic chemistry yields up gaseous compounds of scientific and technical utility. Matheson bottles air itself and offers it for sale in five grades. Plain dry air (it will not dew above −75 degrees F.) is the cheapest, only $7 per standard cylinder. Use it as a pressure source or where moisture makes trouble. Something better? They offer synthesized air, made from oxygen and nitrogen, without the argon and the other variable natural components. You can buy it in four grades, one freer than the other of hydrocarbons or carbon dioxide; these gases are meant for analysis of organics by combustion or for testing engine emission. Purity is not cheap: the best "ultra zero air" costs about $60 per standard cylinder, but it is guaranteed to hold less than .1 part per million of hydrocarbons. A cylinder of "Tennessee natural gas as it is received from the pipeline" (no guarantees, but it usually runs 93 percent methane and about 5 percent ethane and propane) is priced at about double the cost of dry air. But xenon, rarest of the rare elements of the air (about one part in 11 million by volume), is available at modest purity (99.9 percent) at nearly $1,000 a pound.

Gases with odor? Methyl mercaptan ("colorless, highly flammable, toxic gas of an extremely disagreeable odor") is used by your gas company to mark odorless methane, a built-in leak detector for the home; it comes liquefied under its own vapor pressure. The fishy horror trimethylamine (it smells more like ammonia in high concentrations) comes the same way, for use in syntheses. Powerful oxidants? Try ozone, the standard oxygen-atom triangle. It is as perishable as fresh milk; it is shipped dissolved in freon, in a small stainless-steel cylinder packed in dry ice. Its half-life is about three days at room temperature. Of course, if you use a good deal (for germicides, bleaching, sterilization and so on), you will want to make it on the spot, by silent corona discharge in air, and the generator you need is described in the catalogue.

Fluorine is the best oxidant of all. "The most reactive element," it comes in steel cylinders like the rest, at $100 a pound, but order it only when you know how to handle it. The brass or iron piping and fittings you use must be degreased, nitrogen-purged and passivated in a dilute stream of the fluorine, allowed to enter slowly enough to form a metal-fluoride film on the surface. "Continue... until a piece of filter paper or swatch of cotton held close to the exit by means of long metal forceps begins to burn."

The electronics industry uses pure gases to add the soupçon of ionic impurities that gives the semiconductor lattice its subtle transport properties. "Doping" with streams of arsine (AsH_3) or phosphine (PH_3) or the unstable diborane ("no more than .22 lb... per cylinder is shipped") is commonplace; ultrapure materials are deposited from the gas phase out of the delightfully named germane or from silane, both methane analogues.

The *Data Book* is a big volume with a few pages of practical information and the key chemical background for each of more than 130 gases. It is too repetitive for reading through, since it gives complete handling and equipment tips for each entry, but it steadily repays browsing and is a permanent reference source. The *Catalog* adds an essential economic comment by way of a price table for every gas and for all sorts of equipment. The first-aid and medical book, a thin hard-bound guide, collects and extends the safety remarks included in the big data book in compact, explicit and accessible form.

BLEPHARISMA: THE BIOLOGY OF A LIGHT-SENSITIVE PROTOZOAN, by Arthur C. Giese, with the collaboration of Shōichirō Suzuki, Robert A. Jenkins, Henry I. Hirshfield, Irwin R. Isquith and Ann M. DiLorenzo. Stanford University Press ($17.50). Swimming into the microscopic field of view, the big, slow pink spindle (a giant specimen, visible to the unaided eye, spans a full three-quarters of a millimeter) seems to flutter at you a complex set of the eyelashlike fused cilia that surround the mouth opening. (The Greeks called the eyelash *blepharidos.*) The senior author of this monograph of work-in-progress collected his first specimen 44 years ago on the Stanford campus. By now we know a great deal about the self-stained and showy organism, which is distributed worldwide in salt and fresh ponds but is by no means common. There is much more still to learn. The volume, whose contributors spread from New York to Japan, summarizes the known biology of the genus with an eye to allowing students to do more than admire the beast. The free-form *Amoeba,* the slipper *Paramecium* and two or three more kinds of protist have been studied to a rewarding level of detail. This book is intended to place the eyelash-bearer in that group, although it may still fall a little short of the mark.

In most localities one must search many a pool to find the animals. (One hit in 10 is good.) In India after the monsoon they are much more common, and in Japan Professor Suzuki once saw the runoff from a rice paddy stained red with *Blepharisma.* (Any good scientific supplier of living materials will sell you a herd for a couple of dollars.) The animals are bacterial feeders and cannot yet be raised in a defined medium. They survive cold well, although they withstand freezing only in the form of cysts, and they multiply fastest at just under 30 degrees Celsius, dividing in some 10 to 20 hours. Their size and their complex organelles, particularly around the mouth opening, have made them the subjects of an elegant microsurgery with simple instruments (fittingly enough, a "long, rigid eyelash... from a hog... serves very well") to elucidate the processes of regeneration and differentiation.

The results are neither simple nor complete, but they make it plain that three distinct pools of DNA are to be found in the organism, one in a large macronucleus that typically threads its way down the center of the cell; one in the variable micronuclei, a few to a few dozen tiny dense granules that at times reveal elegant chromosomal patterns; the third in the small granules in which the moving cilia are rooted. All of these are self-duplicating, and all play a part in *Blepharisma*'s drama of division, conjugation, regeneration and entry into the state of the resting, dehydrated cyst.

The book begins with a rewarding color plate in which one can see the pinkness of the beast. Here is a leopard that can change its spots. The color is actually carried as a kind of network of tiny pigment granules outside the cell wall; shock the microleopard, say with a dose of novocaine, and the entire membrane bearing the pigment grains is shed and the animal swims off, bleached but apparently normal, minus its capsule of spots. It will form the capsule again a few times, under some protest; albino mutants live happily enough. The pigment is both an indicator and a photosensitive substance. It appears to be a compound related to a plant pigment found in the St.-John's-worts, which induces a damaging photosensitivity in range animals that eat the weeds. Any pink blepharisma seems to have a kindred disease: when oxygen is present, bright light or near-ultraviolet will kill it by inducing a lethal increase in oxygen consumption.

The redder the cell, the more sensitive the organism to light. Albinos are only slightly susceptible; a red cell can be bleached to blue by a few hours' exposure to weak light; the blue cells are not killed even by bright light. It is not easy to see why all this mechanism has evolved. The proposal made by Professor Giese is that the cell needs its red spots to screen out small amounts of the far-ultraviolet and is content to pay for that protection with higher sensitivity to strong visible light.

"We could increase our knowledge . . . if individuals who found pink protozoans anywhere in the world would ship a culture" to Dr. Hirshfield at New York University. This is an unfinished story well told, with a splendid atlas of the microanatomy of the form at magnifications of many tens of thousands, and an invitation to discovery. Is *Blepharisma* an ancient obligate symbiosis of several living forms? What is its formal genetics? Can its regeneration build our ideas of development? Are there in natural pools any of the pink cannibal giants that soon dominate laboratory cultures once any size differences arise? The questions, and how to ask them, are clear in this book; the answers are still partial.

THE DARK FIELDS OF VENUS: FROM A DOCTOR'S LOGBOOK, by Basile Yanovsky, M.D. Harcourt Brace Jovanovich, Inc. ($7.95). A physician since 1937, the author is a novelist and a poet. He belonged to the pre-World War II literary group of young Russian emigrés in Paris but has been writer and doctor in New York since 1942. In 1970, unwilling to continue as an anesthesiologist in a New York hospital he feels "had turned into an abortion mill," he accepted an offer to work for New York City as a physician in venereal-disease clinics. "The most dramatic test for syphilis, exposing the agent of the disease against a dark background under the microscope," gave him the title for this book.

It is a compact, witty, half-loving, half-cynical, opinionated, sardonic and philosophical work. It consists of 327 anecdotes, all but a few in the form of a logbook entry on the encounter and exchanges between one doctor and his intricately diverse patients in their intimate, hasty and ambiguous relationship across the syringe of healing penicillin. It is medically frank, sharply New-York-in-the-1970's, a view of commune and call girl, of the abject and the arrogant, the addict and the executive.

"Most of our 'intellectual' patients are actors, directors,... connected with the theater, films, photography. They are all unemployed 'for the moment,' young, and mostly classified S 90, that is, reported as contacts by acutely infected persons. Incidentally, the majority are homosexuals.

"This seems to be a closed social stratum, polluted with V.D. (like the waitresses of Old Vienna)."

"He teaches math.... Wife's name Priscilla. 'Is it strictly confidential?... It may be quite embarrassing otherwise.' 'Yes, especially with the computers. They collect all the data....' 'If I left right now, would you destroy my chart?' 'Yes,' I said, 'I'll do it.' (I'll do it for Priscilla.)"

"The Dark Field exposed one pale corkscrewlike *Treponema* epileptically twisting under the cruel beam of light. Everybody came to look.... It is always a big catch, and people admire it the way fishermen do when a huge bass or salmon is brought in on the line."

"He was sixty-five years old. We seldom see patients of this age in our clinic. In Paris, middle-aged and elderly patients used to be predominant."

There is a wonderful diagram: a 13-step flow chart "through which the poor V.D. customer is forced to move... which might have been worked out by the Scholastics.... I once had the temerity to propose... to initiate step 11 directly after step 7. The reaction was as if I had suggested a Russian revolution."

Dr. Yanovsky is learned, aristocratic, devout, a man who sees strength in tradition, a worldly admirer of the life of the mind. Surely his tales are not so naturalistic and artless as they seem, but they add up to a polished kind of ethnography of distress in our time, a true culture taken from the bloodstream of the city, incubated in a rich medium of thought, an engrossing narrative lying between art and sociology.

INDEX OF ADVERTISERS

JUNE 1973

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

ENFORCING THE CLEAN AIR ACT OF 1970

AIR POLLUTION: A COMPREHENSIVE TREATISE. Edited by Arthur C. Stern. Academic Press, 1968.

THE CLEAN AIR ACT. Environmental Protection Agency, Washington, D.C., 1970.

THE ANCHOVY CRISIS

THE CLUPEOID RESOURCES OF TROPICAL SEAS. Alan R. Longhurst in *Oceanography and Marine Biology—an Annual Review,* edited by Harold Barnes, Vol. 9, pages 349–385; 1971.

THE FISH RESOURCES OF THE OCEAN. Edited by J. A. Gulland, pages 136–145. Fishing News Ltd., Surrey, England, 1971.

MEN, BIRDS AND ANCHOVIES IN THE PERU CURRENT—DYNAMIC INTERACTIONS. Milner B. Schaefer in *Transactions of the American Fisheries Society,* Vol. 99, No. 3, pages 461–467; 1970.

THE DYNAMICS OF THE ANDROMEDA NEBULA

THE ROTATION AND RADIAL VELOCITY OF THE CENTRAL PART OF THE ANDROMEDA NEBULA. F. G. Pease in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 4, No. 1, pages 21–24; January 15, 1918.

ROTATION OF THE ANDROMEDA NEBULA FROM A SPECTROSCOPIC SURVEY OF EMISSION REGIONS. Vera C. Rubin and W. Kent Ford, Jr., in *The Astrophysical Journal,* Vol. 159, No. 2, Part 1, pages 379–403; February, 1970.

STELLAR MOTION NEAR THE NUCLEUS OF M31. Vera C. Rubin, W. Kent Ford, Jr., and C. Krishna Kumar in *The Astrophysical Journal,* Vol. 181, No. 1, Part 1, pages 61–77; April, 1973.

ULTRAFAST PHENOMENA IN LIQUIDS AND SOLIDS

PICOSECOND LASER PULSES. N. Bloembergen in *Comments on Solid State Physics,* Vol. 1, No. 2, pages 37–42; June–July, 1968.

AN ULTRAFAST LIGHT GATE. M. A. Duguay and J. W. Hansen in *Applied Physics Letters,* Vol. 15, No. 6, pages 192–194; September 15, 1969.

OPTICAL PHONON LIFETIME MEASURED DIRECTLY WITH PICOSECOND PULSES. R. R. Alfano and S. L. Shapiro in *Physical Review Letters,* Vol. 26, No. 20, pages 1247–1251; May 17, 1971.

DIRECT MEASUREMENT OF THE VIBRATIONAL LIFETIMES OF MOLECULES IN LIQUIDS. A. Laubereau, D. von der Linde and W. Kaiser in *Physical Review Letters,* Vol. 28, No. 18, pages 1162–1165; May 1, 1972.

ELECTRONIC NUMBERS

INFORMATION DISPLAYS: INNOVATION IS THE RULE. Ronald K. Jurgen in *IEEE Spectrum,* Vol. 9, No. 11, pages 18–24; November, 1972.

LIQUID CRYSTAL DISPLAYS. Richard W. Gurtler and Craig Maze in *IEEE Spectrum,* Vol. 9, No. 11, pages 25–29; November, 1972.

DIGITAL READOUTS COURSE. Edited by Stephen A. Thompson in *The Electronic Engineer,* Vol. 31, July–December, 1972; Vol. 32, January, 1973.

MATCHING DRIVER CIRCUITRY TO MULTIDIGIT NUMERIC DISPLAYS. Alan Sobel in *Electronics,* Vol. 46, No. 9, pages 95–99; April 26, 1973.

LIFE IN TALL TREES

PHYTOSOCIOLOGY AND ECOLOGY OF CRYPTOGAMIC EPIPHYTES. J. J. Barkman. M. V. Koninklijke Van Gorcum Comp., Assen, the Netherlands, 1958.

THE ADAPTIVE GEOMETRY OF TREES. Henry S. Horn. Princeton University Press, 1971.

DIRECT, NONDESTRUCTIVE MEASUREMENT OF BIOMASS AND STRUCTURE IN LIVING, OLD-GROWTH DOUGLAS-FIR. W. C. Denison, D. M. Tracy, F. M. Rhoades and M. Sherwood in *Proceedings—Research on Coniferous Forest Ecosystems—a Symposium,* edited by J. F. Franklin, L. J. Dempster and R. H. Waring. Pacific Northwest Forest & Range Experimental Station, 1972.

THE HUMAN LYMPHOCYTE AS AN EXPERIMENTAL ANIMAL

IMMUNOGLOBULIN PRODUCTION: METHOD FOR QUANTITATIVELY DETECTING VARIANT MYELOMA CELLS. Philip Coffino, Reuven Laskov and Matthew D. Scharff in *Science,* Vol. 167, No. 3915, pages 186–188; January 9, 1970.

IMMUNOGLOBULIN PRODUCTION BY HUMAN LYMPHOCYTOID LINES AND

We can give you a good address anywhere in the world.

Your business may take you to all kinds of places. In some of them you'll know your way around.
In others you won't. In any case we'd like to put our address at your disposal. To have your mail sent to or leave a message at.
If possible, to arrange a business introduction or some sight-seeing.
We maintain offices all over the world, even in places we don't fly to. Like Belem or Katmandu, for instance. And in all of them, we're ready to help you. Certainly, we'll have a friendly smile when you get there.
Even if you didn't fly with us.

Lufthansa
the more you fly

CLONES: ABSENCE OF GENIC EXCLUSION. Arthur D. Bloom, Kyoo W. Choi and Barbara J. Lamb in *Science,* Vol. 172, No. 3981, pages 382–383; April 23, 1971.

SYNTHESIS OF PLASMA MEMBRANE-ASSOCIATED AND SECRETORY IMMUNOGLOBULIN IN DIPLOID LYMPHOCYTES. Richard A. Lerner, Patricia J. McConahey, Inga Jansen and Frank J. Dixon in *Journal of Experimental Medicine,* Vol. 135, No. 1, pages 136–149; 1972.

AN ADVICE-TAKING CHESS COMPUTER

COMPUTING MACHINERY AND INTELLIGENCE. A. M. Turing in *Mind,* Vol. 59, No. 236, pages 433–460; October, 1950.

PERCEPTION AND MEMORY VS. THOUGHT. Adriaan De Groot in *Problem Solving: Research, Method and Theory,* edited by B. Kleinmuntz. John Wiley & Sons, Inc., 1966.

THOUGHT AND CHOICE IN CHESS. Adriaan De Groot, translated by G. W. Baylor. Mouton, 1966.

INFORMATION-PROCESSING ANALYSIS OF PERCEPTUAL PROCESSES IN PROBLEM SOLVING. Herbert A. Simon and Michael Barenfeld in *Psychological Review,* Vol. 76, No. 5, pages 473–483; September, 1969.

COMPUTERS, CHESS AND LONG-RANGE PLANNING. M. M. Botvinnik, translated by A. Brown. Springer-Verlag Inc., 1970.

MATHEMATICAL GAMES

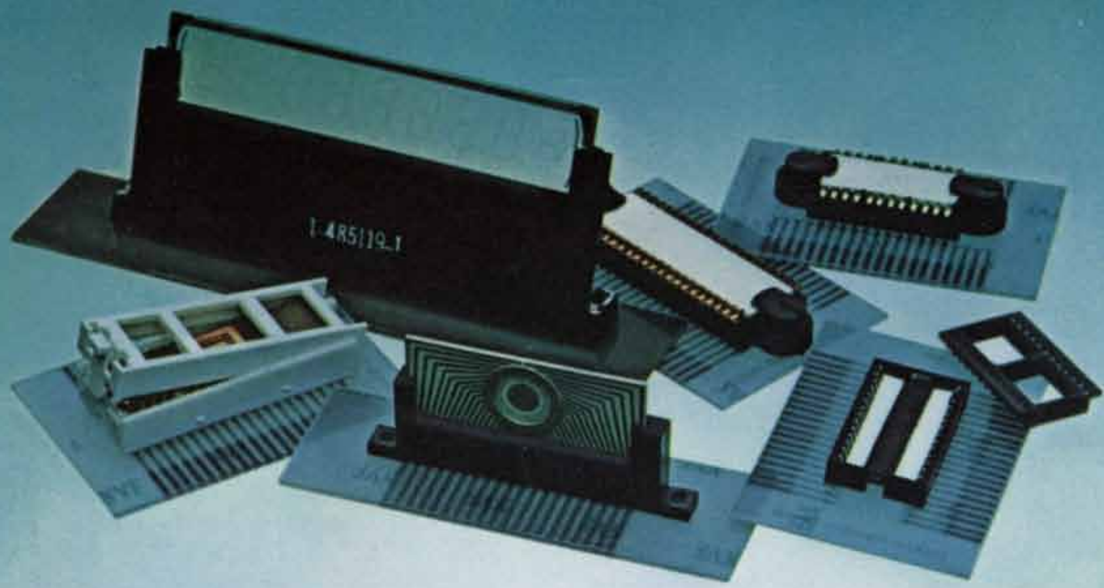
TOWARD A THEORY OF CROSSING NUMBERS. W. T. Tutte in *Journal of Combinatorial Theory,* Vol. 8, No. 1, pages 45–53; January, 1970.

LATEST RESULTS ON CROSSING NUMBERS. R. K. Guy in *Recent Trends in Graph Theory: Proceedings of the 1st New York City Graph Theory Conference,* edited by M. Capobianco, J. B. Frechen and M. Krolik. Springer-Verlag Inc., 1971.

THE AMATEUR SCIENTIST

THE EFFECT OF LIGHT ON THE SETTLING OF SUSPENSIONS. C. G. T. Morison in *Proceedings of the Royal Society of London,* Series A, Vol. 108, No. A746, pages 280–284; June 2, 1925.

TURBIDIMETRY AND NEPHELOMETRY in *Encyclopedia of Chemical Technology: Vol. XX,* edited by R. E. Kirk and D. F. Othmer. Wiley-Interscience, 1969.



AMP is making things smaller,

easier and denser for interconnecting MOS.

Take the 42-position connector in the foreground, for example. It's less than .250-inch high, with lid installed. And only .050-inch lead centers! Simple, easy to use. With zero entry force.

Then, there's our 40-position connector (far right, foreground). It's only .235-inch high, .100-inch lead centers and—with its snap-on adapter—converts in seconds to accept leaded as well as leadless packages. Also available for 24 and 28 package configurations.

We have others, too. Bottom metallized arrays, liquid crystal connectors, etc. But that's another story. To get the AMP microelectronics story, write AMP Industrial Division, Harrisburg, Pa. 17105.

AMP microelectronics: doing small things in a big way.

AMP
INCORPORATED

AMP is a trademark of AMP Incorporated.