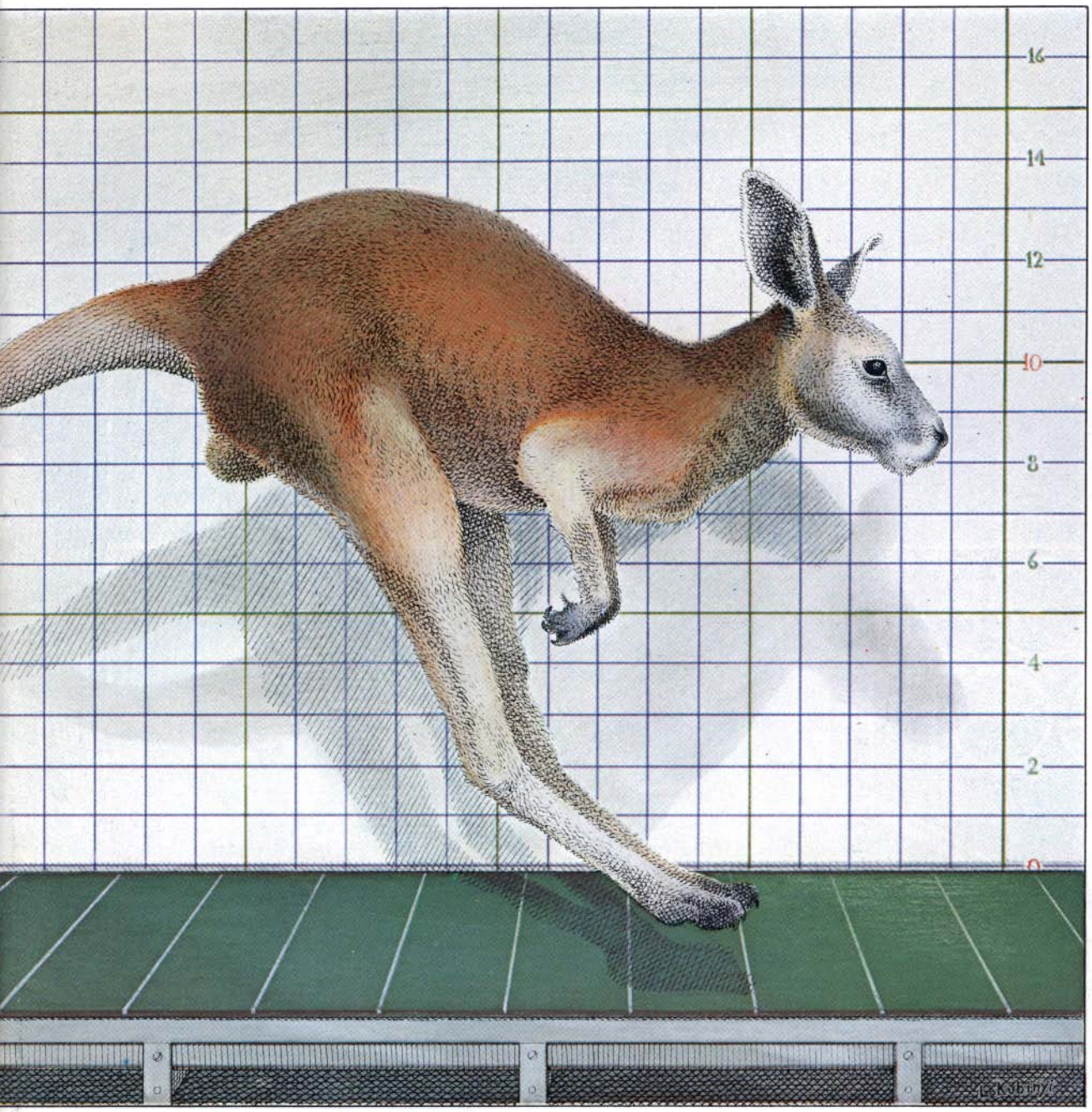
SCIENTIFIC AMERICAN



KANGAROOS

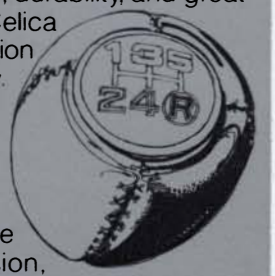
$1.50

August 1977

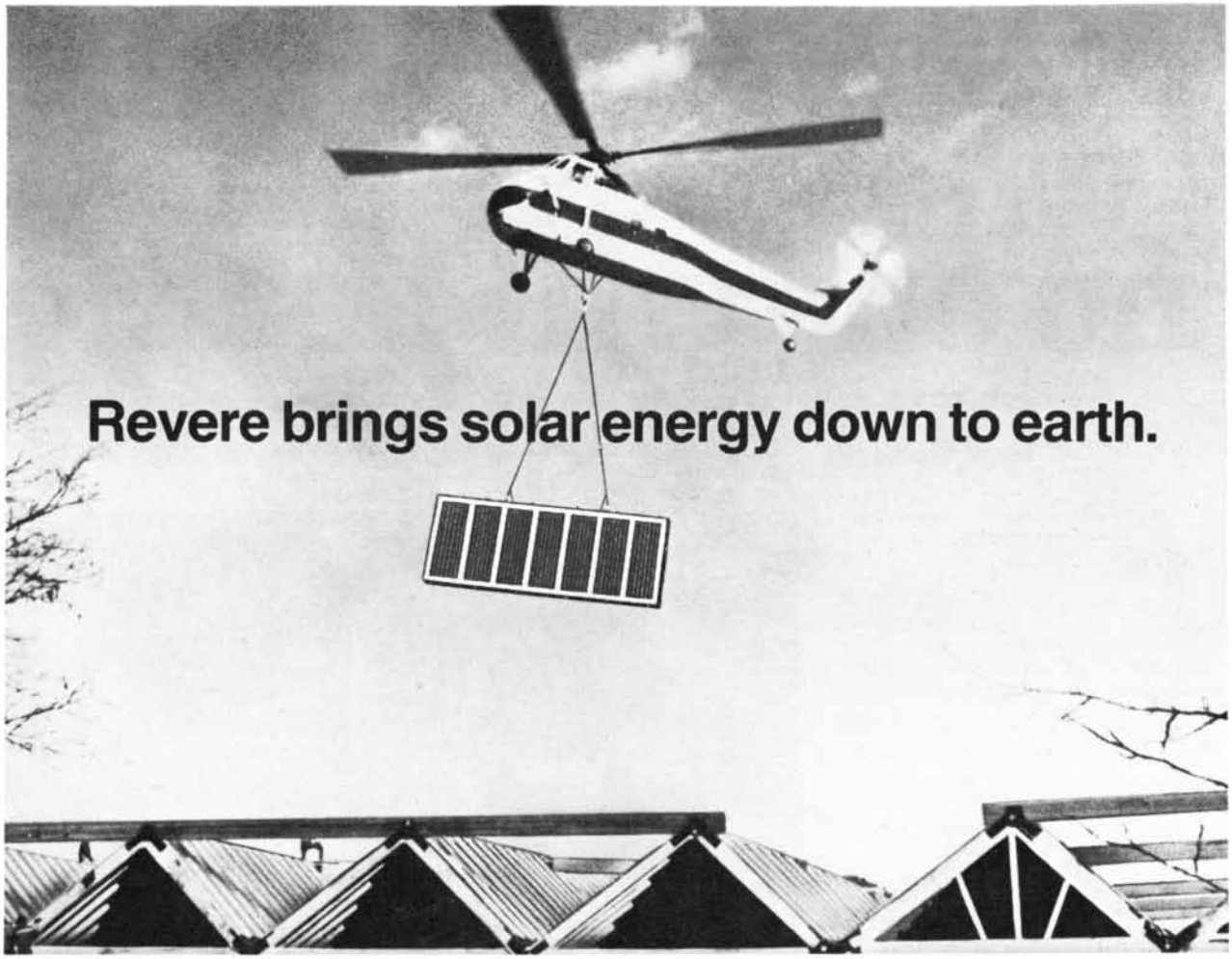
THE HOT ONE.

That's the 1977 Toyota Celica. Hot because of the way it looks, and the way it moves. Hot because there are three models, including the racy '77 GT Liftback. Hot because the Celicas are built with Toyota's famous toughness and durability. Their welded unitized-body construction eliminates body nuts and bolts to make them three of the most durable cars on the road.

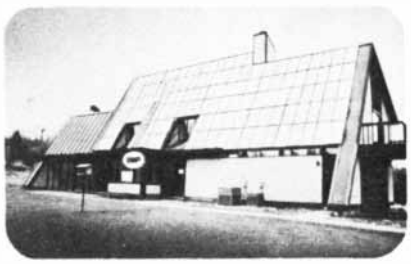
Hot Performer. The '77 Celicas are powered by the revolutionary 20R engine—a 2.2 liter overhead cam design. Built from the ground up to give power, durability, and great gas mileage. In 1977 EPA tests the Celica GT with 5-speed overdrive transmission got 37 mpg highway, 22 mpg city. These mileage figures are estimates. The actual mileage you get will vary depending on your driving habits and your car's condition and equipment. California and EPA designated high altitude ratings will be lower.



Hot Items. A lot of hot features come standard on the 1977 Celicas. Like MacPherson strut front suspension, steel-belted radials, power front disc brakes, electric clock, tachometer, reclining bucket seats, tinted glass, and much more. We're proud of the '77 Celicas. In fact, we're proud enough to say, if you can find a better built small car than Toyota...buy it.



The 'copter is lifting Revere copper solar collectors atop the Shenandoah Community Center in Georgia, the largest solar energy project in the world. The sun will provide 90% of the heat and 60% of the air conditioning for this 55,000 sq. ft. center. It will save $77,000 a year in fuel costs!



Revere Sun-Roof collectors are an integral part of the roof of this N.Y. State residence.

McDonalds restaurant uses Revere Sun-Aid collectors as a natural addition to the roof.

Revere modular solar collectors help reduce the fuel bills for this HUD demo house.

Today, Revere collectors are slashing fuel bills for a growing number of homes and commercial buildings in all types of climates. Revere produces modular solar collectors for existing buildings...a combination roof and solar collector for new construction... and also supplies the Sun-Pride™ solar system for domestic or service hot water.

The collectors as well as the components in the Sun-Pride system are backed by Revere, a company with a reputation for reliability.

Revere is a trailblazer in solar energy and has helped make it practical today. For information write: Solar Energy Department, Revere Copper and Brass Incorporated, P.O. Box 151, Rome, N.Y. 13440.

REVERE

After people learn what we've done, no one will heckle our speakers.

We're as close to the impossible as possible.

Our new speakers color sound.

Anybody's do.

Should they tell you otherwise, they speak with forked frequency response.

We at Sony developed our new speaker line with this grim reality in mind.

Our goal was to create speakers with a minimum of coloration. With a frequency response flat and wide. With low distortion. And with repeatability. Which means that each speaker we turn out will sound like the one before and the one after.

Searching and researching.

Before you can make a good speaker you have to make a lot of bad ones.

We turned out dozens of prototypes that were made with the same specs, but sound like they weren't.

That's because your ear is more sophisticated than our measurements.

You can hear how pure water is, for example.

The purity of the water in which the pulp for the speaker cone is pressed influences the sound.

But it would hardly change the frequency response—or any other measurements.

Now there are dozens and dozens of elements that interact this way.

Changing one changes the other and almost changed our minds about going into the speaker business.

But we stuck it out. Applying the age-old technique of trial and error.

That's why we labored for three years to bring you our speakers. While other manufacturers rushed frantically to market with theirs.

We keep the whole world in our hands.

Understanding how to control the sound of our speakers, we realized we had to control what went into them.

So we did the only logical thing.

We built a plant. Which does nothing but produce—under outrageously close control—the components for our speakers.

Few companies watch what you hear so carefully.

Improvements that are heard and not seen.

As you can see, a lot goes into producing a speaker that's not easily seen. (One beautiful exception—the handsome finish on our cabinets.)

That includes the carbon fiber we mix into the speaker cone paper.

Carbon fiber is light and strong. Light, so our speaker needs less power to operate it, and is therefore more efficient.

Light, so our speaker cone reacts quickly to stops and starts in the signal. The result: improved transient response.

Strong, so the cone won't bend out of shape in the high frequency range.

Moreover, carbon fiber won't resonate much. It has a low Q, and it took someone with a high IQ to realize it would absorb the unwanted vibration.

Unwanted vibration is also reduced by our cast aluminum basket, which is superior to a cheap stamped metal one.

We're confident that the results of our three year effort will be clear after three minutes of listening.

At which point, far from heckling our speakers, you'll be tempted to give them a standing ovation.

The SSU-3000 and SSU-4000.
Great speakers like these deserve an audience.

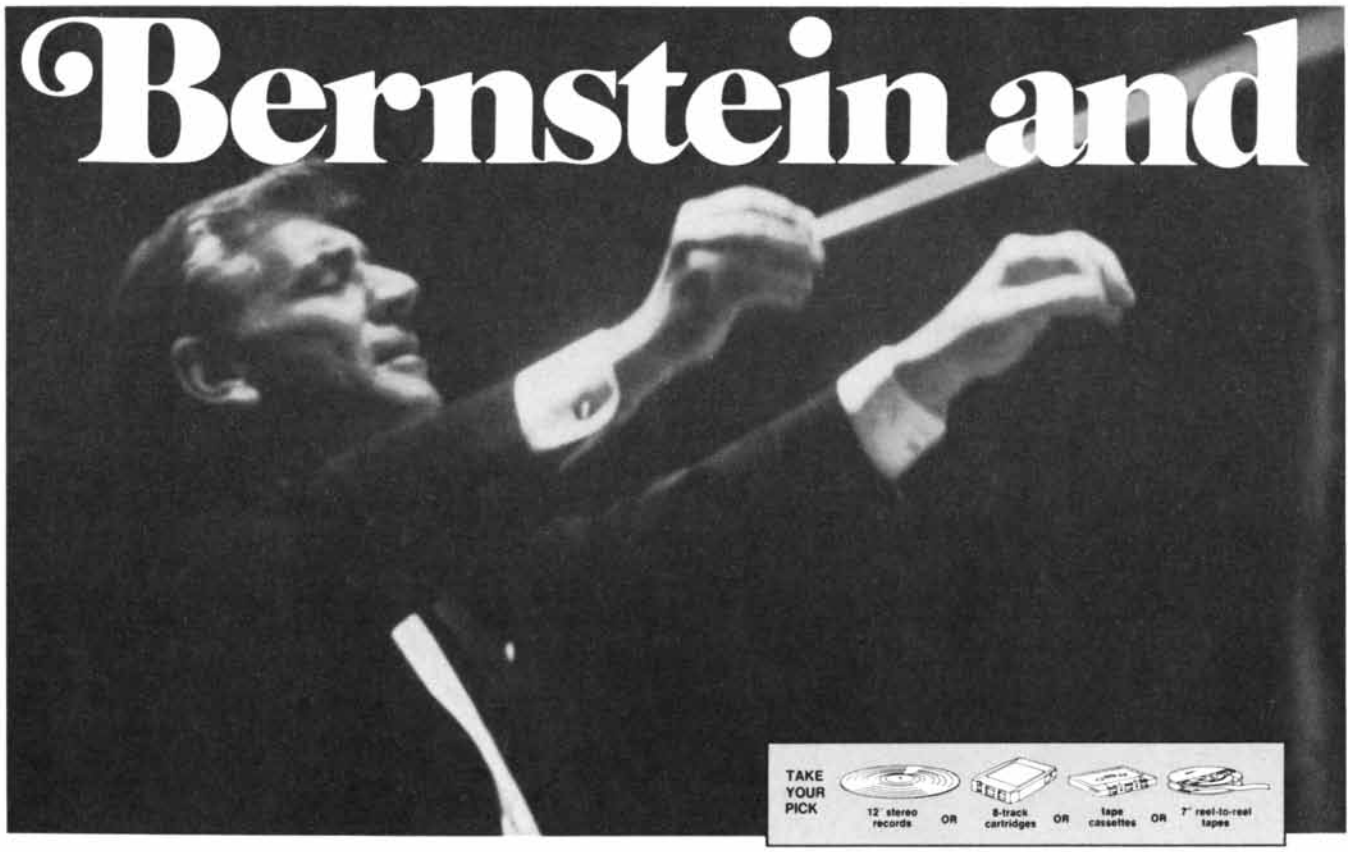
SONY®

Established 1845 **SCIENTIFIC AMERICAN** August 1977 Volume 237 Number 2

ARTICLES

DEPARTMENTS

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



262493-262494. Bach: Suites 1-4 and more lute music played by John Williams on the guitar (Counts as 2—Columbia)

173641. Bach: Greatest Hits—*Air On a G String; Sleepers Awake;* many others (Columbia)■

223800-223801. Bach: 6 Brandenburg Concertos—A. Newman & Friends (Counts as 2—Columbia)☆

273334. Beethoven: Symphony No. 3 (Eroica)—Solti and the Chicago Symphony (London)

273409. Beethoven: Three Favorite Piano Sonatas (Moonlight, Pathetique, Appassionata)—Vladimir Horowitz (Columbia)

270702. Beethoven: Symphony No. 7 (Dance)—Stokowski and the New Philharmonia (London)

154856. Beethoven: Symphony No. 9 (Chorale)—Mormon Tabernacle Choir; Ormandy and the Phila. Orch. (Columbia)†

223016-223017. Beethoven: Greatest Hits—*Fur Elise; Minuet in G;* first movement *"Moonlight" Sonata; Minuet in G;* excerpt from *"Moonlight" Sonata;* etc. (Counts as 2—Columbia)

269142. Beethoven: Piano Sonatas No. 18 and 23 (Appassionata)—Lazar Berman (Columbia)■

205120. Berlioz: Symphonie Fantastique—Boulez and London Sym. (Columbia)†

271965. Bizet: Carmen and **L'Arlesienne Suites**—Stokowski, cond. (Columbia)

263293. Claude Bolling: Suite for Flute & Jazz Piano—Jean-Pierre Rampal, flute; Bolling, piano (Columbia)††

223388-223389. Brahms: Piano Concertos No. 1 & 2—Rudolf Serkin; George Szell, Cleveland Orch. (Counts as 2—Columbia)

258673-258674. Brahms: Sonatas for Violin and Piano (complete)—Stern, violin; Zakin, piano (Counts as 2—Columbia)†

267963. Brahms: Piano Concerto No. 1—Artur Rubinstein; Zubin Mehta, Israel Phil. (London)

216036. Chopin: Piano Concerto No. 2; DeFalla: Nights In the Gardens Of Spain—Alicia de Larrocha, pianist (London)

246843. Chopin: Mazurkas, Etudes, and more—new recordings by Vladimir Horowitz (Columbia)

271882. Copland: Appalachian Spring; Gershwin: American In Paris—Zubin Mehta, L.A. Philharmonic (London)

262352. Dvorak: Piano Concerto—Justus Frantz, piano; Bernstein and N.Y. Phil. (Columbia)

269688. Dvorak: Symphony No. 9 (New World) and **Carnival Overture**—Zubin Mehta, L.A. Phil. (London)

161414. The Glory of Gabrieli—music for multiple choirs, brass and organ. E. Power Biggs; Gregg Smith Singers; and others (Columbia)†

187112. Gershwin: Greatest Hits. Bernstein et al: *Rhapsody in Blue; An American in Paris; Three Preludes;* etc. (Columbia)

177428. Gershwin: Porgy & Bess (Symphonic Picture) and **An American In Paris**—Ormandy, Philadelphia Orchestra (Columbia)†

252155-252156. Gilbert & Sullivan: Pirates Of Penzance (complete)—D'Oyle Carte Opera Co. (Counts as 2—London)†

230391. Grieg: Piano Concerto; Rachmaninoff: Rhapsody on a Theme of Paganini—P. Entremont, piano; Eugene Ormandy, Phila. Orch. (Columbia)■

257956. Haydn: Symphonies 101 (Clock) **and 103** (Drum Roll)—Bernstein and the N.Y. Phil. (Columbia)

264655. Music of Victor Herbert—Beverly Sills, A. Kostelanetz (Angel)

236448. Scott Joplin—*Entertainer,* many more. E. Power Biggs, pedal harpsichord (Columbia)

250720. Mendelssohn: Piano Concertos—Murray Perahia, piano; Marriner cond. (Columbia)

242750. Mendelssohn and Tchaikovsky Violin Concertos—Stern, violin; Ormandy cond. (Columbia)

269555. Mozart: Piano Concertos No. 14 and 24—M. Perahia, English Chamber Orch. (Columbia)☆

260588. Mozart: Piano Sonatas 14, 16, 18—Glenn Gould (Columbia)†

213330. Offenbach: Gaite Parisienne—Bernstein, N.Y. Phil. (Columbia)☆

252551. Orff: Carmina Burana—Michael Tilson Thomas, Cleveland Orch. and Chorus (Columbia)

227009. Prokofiev: Peter and the Wolf; Tchaikovsky: Nutcracker Suite—Bernstein, N.Y. Phil. (Columbia)■

243600. Ravel: La Valse; Mother Goose Suite—Bernstein and New York Philharmonic (Columbia)■

188920. Respighi: Pines and Fountains of Rome—Eugene Ormandy, Philadelphia Orch. (Columbia)■

230409. Rimsky-Korsakov: Scheherazade—Bernstein, N.Y. Phil. (Columbia)

260596-260597. Schoenberg: Gurre-Lieder—Boulez and the BBC Symphony Orch. (Counts as 2—Columbia)†

176354. Schubert: Symphony No. 9 in C (The Great)—Ormandy and the Philadelphia Orchestra (Columbia)†

192690. Sibelius: Greatest Hits—*Swan of Tuonela; Finlandia;* etc. (Columbia)†

250795. Sibelius: Symphony No. 2—Tauno Hannikainen, The Sinfonia of London (Columbia Musical Treas.)

173658. Johann Strauss: Blue Danube; Tales From Vienna Woods; Emperor Waltz; etc. Ormandy, Phila. Orch. (Columbia)††

268482. Johann Strauss: Tales From Vienna Woods; Die Fledermaus Overture; others. Bernstein and N.Y. Phil. (Columbia)

202796. Richard Strauss: Also Sprach Zarathustra—Bernstein and New York Philharmonic (Columbia)

155309. Stravinsky: Firebird and **Petrushka Ballet Suites**—Columbia Symphony conducted by Stravinsky (Columbia)†

224733. Stravinsky: Rite Of Spring—Bernstein, London Sym. (Columbia)††

250845. Tchaikovsky: Piano Concerto No. 1—Watts, piano; Bernstein, N.Y. Phil. (Columbia)

137778. Tchaikovsky: Romeo and Juliet; Sleeping Beauty Ballet Suite—Ormandy and Philadelphia Orchestra (Columbia)†

220947-220948. Tchaikovsky: Greatest Hits—*Capriccio Italien; 1812 Overture;* etc. (Counts as 2—Columbia)†

253872. Villa-Lobos: Bachianas Brasileiras No. 2, 5, 6, 9—Mady Mespie, soprano; Capolongo, Orch. de Paris (Angel)■

255620. Wagner: Organ Orgy—opera excerpts transcribed for organ! Anthony Newman (Columbia)

233684. Wagner: Tristan und Isolde (Prelude and Love-Death, etc.)—Boulez, N.Y. Phil. (Columbia)■

INSTRUMENTAL AND VOCAL COLLECTIONS

265926. Age Of Gold—orchestral showpieces by Prokofiev, Shostakovich, Borodin, etc. Bernstein, N.Y. Phil. (Columbia)

174607. Bernstein's Greatest Hits, Vol. 2—*Danse Macabre, Polovtsian Dances;* etc. (Columbia)†

176602. E. Power Biggs' Greatest Hits—Bach: *Sleepers Awake!*; Ives: *Variations on "America";* many more (Columbia)†

267922. Country Fiddle Band—Gunther Schuller conducts 100 years of authentic country dance music (Columbia)■

259168. Footlifters: A Century of American Marches—G. Schuller, conductor (Columbia)

246645. Music From Walt Disney's "Fantasia"—*Night on Bald Mountain; Sorcerer's Apprentice;* many others (London)

262378-262379. Jascha Heifetz in Concert—he plays Franck, Bach, etc. (Counts as 2—Columbia)

268318. Heifetz-Piatigorsky Concerts—chamber music by Dvorak, Stravinsky, Gliere, others (Columbia)†

269571. Robert Merrill & Mormon Tabernacle Choir—Yankee Doodle Dandies! *Over There; Till We Meet Again;* etc. (Columbia)

261891-261892. Mormon Tabernacle Choir—Rock Of Ages. 31 great hymns (Counts as 2—Columbia)

Just two of the great performers whose latest recordings are now offered on stereo records, cassettes, 8-track cartridges and reel tapes

THE CLASSICAL DIVISION OF COLUMBIA MUSIC CLUB INVITES YOU TO TAKE

any 8 for only $1.00 plus shipping and handling

if you join now and agree to buy only 8 selections (at regular Club prices) during the coming two years

YES, Leonard Bernstein and Sir Georg Solti are but two of the great conductors whose thrilling performances are now offered on these pages. You'll also find superb recordings by Boulez, Horowitz, Mehta, Ormandy, Serkin, Stern, Stokowski (to name but a few)...in widely-acclaimed performances of the greatest symphonies, concertos, orchestral masterpieces of all time! And as a new member, you may take ANY 8 of these selections—in your choice of stereo records, cartridges, cassettes or reel tapes—all for only $1.00.

Your only membership obligation is to buy just eight more selections in the coming two years, at the regular Club prices (which currently are $6.98 or $7.98, plus shipping and handling. Multi-unit sets and Double Selections may be somewhat higher).

Think of it...you need to buy as few as four records or tapes a year...and you won't have any difficulty in selecting them, because you'll receive our Classical Bulletin regularly every four weeks (13 times a year). And we'll also send you the regular Club Magazine, which contains additional classical recordings, as well as hundreds of selections from every field of music. In addition, up to six times a year we may send some special issue of the Bulletin, offering extra-special savings.

This Bulletin will tell you about many of the newest classical releases and we'll even select one as "The Selection of the Month"—but there is no obligation to accept it...order only those you want *when you want them!*

A special response card will be enclosed with each Bulletin...and you may use it to order or reject any selections offered simply by mailing it by the date specified. And if you want *only* The Selection of the Month, you need do nothing—it will be shipped to you automatically. What's more, you'll have at least ten full days in which to make your decision—if you ever have less time than that, just return The Selection of the Month at our expense, for full credit.

You may cancel membership at any time after you've purchased eight selections—or continue under our special bonus plan, which offers you substantial savings on all future purchases. Here, indeed, is the most convenient way possible of acquiring the classical library you want—*at the greatest savings possible!* So do act now.

Note: selections with two numbers are 2-record sets or double-length tapes. Each of these "double selections" counts as 2—so please write in both numbers.

Each selection is available as a record, 8-track cartridge, cassette or reel-to-reel tape—unless it is followed by one of these symbols:

† Selection is available as a record *only*
☆ Available on records and cassettes *only*
†† Available on records and cartridges *only*
■ Available on records, cartridges, cassettes

236828-236829. Ormandy/Philadelphia Orch. March Album—*Stars & Stripes, American Salute, Marche Militaire,* many more (Counts as 2—Columbia)††

257311. Renata Scotto—arias by Puccini, Cilea, Mascagni, etc. (Columbia)†

264663. Beverly Sills and Andre Kostelanetz—Plaisir d'Amour. Songs by Delibes, Poulenc, Bizet, Martini, etc. (Columbia)

274126. Solti/Chicago Symphony—Ravel: *Bolero;* Debussy: *Afternoon of a Faun* and *La Mer* (London)

271973. Stokowski Encores—Chopin: *Mazurka;* Byrd: *Pavan;* Dvorak: *Slavonic Dance;* more (London)

262923. Barbra Streisand—Classical Barbra. Songs by Debussy, Canteloube, Handel, etc. (Columbia)■

259564. Joan Sutherland & Luciano Pavarotti—duets from *Lucia Di Lammermoor; Rigoletto;* etc. (London)†

191841. Szell's Greatest Hits—Smetana: *Moldau;* Mozart: *Marriage of Figaro Overture;* etc. (Columbia)†

272567. Thijs Van Leer/Introspection 2—famous flutist plays his own music; also Granados, Bach, etc. (Columbia)☆

Columbia House

All applications are subject to review and Columbia House reserves the right to reject any application

Columbia Music Club/CLASSICAL DIVISION
Terre Haute, Indiana 47811

Please accept my membership application in the Club under the terms outlined in this advertisement. Send me the 8 classical selections listed below for only $1.00, plus shipping and handling. I agree to buy eight more selections (at regular Club prices) in the coming two years—and may cancel membership at any time after doing so.

SEND MY SELECTIONS IN THIS TYPE OF RECORDING (be sure to check one): **6RJ/DM**

☐ 8-Track Cartridges ☐ Reel Tapes
☐ Tape Cassettes ☐ Records

Write in numbers of 8 selections		

☐ Mr.
☐ Mrs.
☐ Miss

(Please Print) First Name Initial Last Name

Address Apt. No.

City

State Zip Code

Do You Have A Telephone? (Check one) ☐ YES ☐ NO

APO, FPO, Alaska, Hawaii, Puerto Rico: write for special offer

361/F77

THE COVER

The illustration on the cover re-creates a laboratory experiment in animal energetics conducted at Harvard University. The purpose of the experiment was to compare the energy cost of two-legged hopping gaits with that of four-legged gaits over a range of speeds. The only large mammals that move by hopping are the kangaroos of Australia, and so four red kangaroos were flown from New South Wales to Massachusetts with the aim of inducing them to hop on a treadmill. Two of the four could be trained to hop as is shown in the illustration. For the experiment the kangaroos on the treadmill wore masks as they hopped, making it possible to measure their consumption of oxygen and the energetic cost of hopping (see "Kangaroos," by T. J. Dawson, page 78).

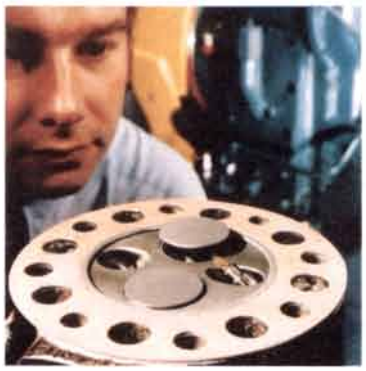
THE ILLUSTRATIONS

Cover illustration by Laszlo Kubinyi

Page	Source
25	U.S. Air Force (*top*), *Jane's All the World's Aircraft* (*bottom*)
26–31	Gabor Kiss
33	Wise Observatory, Tel-Aviv University (*top left and top center*); Haute-Provence Observatory (*top right*); National Geographic Society–Palomar Observatory Sky Survey (*bottom*)
34–39	Adolph E. Brotman
41	Ralph Morse
42–48	Allen Beechel
61	Wendell A. Duffield, U.S. Geological Survey
62	Henry N. Pollack and David S. Chapman
63–76	Andrew Tomko
79	Lorelle M. Raboni
80	R. McBlain
81–89	Lorelle M. Raboni
91	Daniel Branton, Harvard University
92	A. E. Walsby, University College of North Wales
93–96	Ilil Arbel
97	Daniel Branton, Harvard University
99	Stephen Shore
100	Dan Todd
101	Chrysler Corporation
102	Dan Todd
104	Dan Todd (*top*), Stephen Shore (*bottom*)
106	*The New York Times*
108	Jerome Kuhl
109	Noel T. Florendo, Baptist Memorial Hospital, Memphis, Tenn.
110–112	Jerome Kuhl
113	Alan D. Iselin
114	Jerome Kuhl
115	Bruce K. Krueger, Washington University
116–118	Jerome Kuhl
120–124	Ilil Arbel
127	Ben Rose
128–130	Michael Goodman
133	*Mission to Earth: Landsat Views the World.* © 1976, National Aeronautics and Space Administration

Agitate an air-fuel mixture—make it turbulent—and it will burn faster. Every combustion expert knows this qualitative relationship. And every automobile engine relies on it for efficient operation.

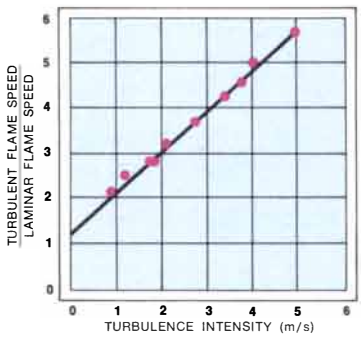
Would more turbulence improve engine operation? Perhaps, up to a point. But where is that point? Until now, no one has been able to quantify how turbulence affects combustion in an engine.



In attempting to do so, engineers at the General Motors Research Laboratories installed a triaxial hot-wire anemometer in the combustion chamber of a single-cylinder engine. While the engine was being driven by a dynamometer, a computer scanned and analyzed the deluge of data generated.

Next the researchers ran the engine in a firing mode to get instantaneous chamber pressures. Then they cranked these pressure readings into a combustion heat-release model that calculated turbulent flame speed and its ratio to laminar flame speed.

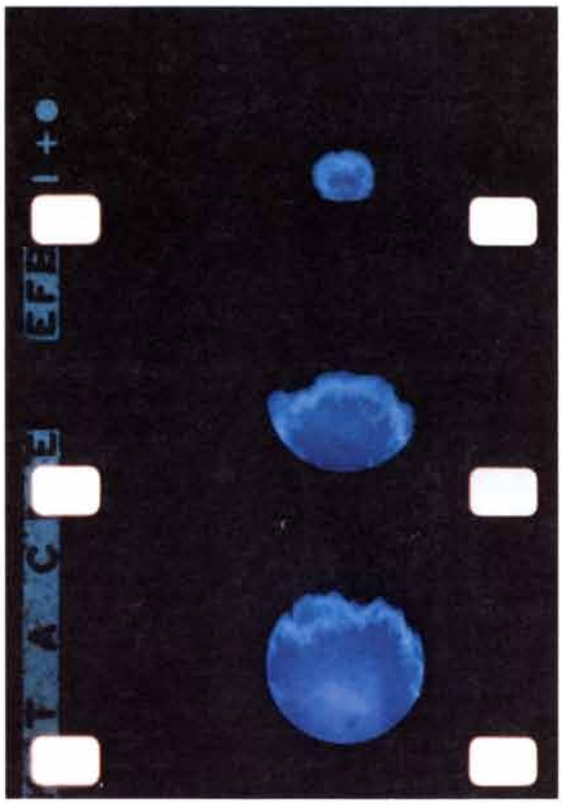
What did they learn? First, the relationship between flame speed ratio and turbulence intensity is linear. Second, turbulence intensity depends linearly on volumetric flow rate through the intake valve.



And the payoff? Powerful insight for assessing the effects of new engine designs on fuel economy and exhaust emissions. In fact, Lab engineers are currently using this insight to pursue optimum combustion chamber configurations.

Lord Kelvin was so right. ". . . when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind."

We're putting numbers on combustion.



GM

General Motors Research Laboratories
Warren, Michigan 48090

LETTERS

Sirs:

In the article "Cruise Missiles" [SCIENTIFIC AMERICAN, February] Kosta Tsipis and *Scientific American* perform a useful service in describing this important development in military technology. However, the article's conclusions that the development and deployment of strategic (that is, long-range) cruise missiles are undesirable and that it is possible to distinguish between them and tactical (that is, short-range) cruise missiles using a satellite monitoring system are based on judgments with which we have serious disagreement.

Tsipis asserts that strategic cruise missiles are vulnerable to enemy air defenses and that they "appear to be operationally inferior to existing strategic weapons." We question whether defenses will be as effective as he implies, considering the small radar cross section of the missiles and the low altitude at which they can penetrate. Equipping B-52's with long-range cruise missiles would make it unnecessary for the aircraft to penetrate some of an opponent's air defense and would thereby add to the effectiveness of the bomber component of the strategic triad.

Tsipis suggests that if the U.S.S.R. were to deploy strategic cruise missiles, the U.S. would have to deploy air defenses that up to that time had not been necessary because there had not been any credible threat from the U.S.S.R. In actuality the existing Russian bomber force can quite credibly deliver nuclear weapons with a total of thousands of times the destructive power of the Hiroshima bomb. The main reason for not defending against the Russian bomber force and the existing older generation of Russian submarine-launched, nuclear-armed cruise missiles is that even if we did, we would still be vulnerable to attack by ballistic missiles. This will continue to be true in the event that the U.S.S.R. deploys a new generation of strategic cruise missiles.

Tsipis argues that cruise missiles should not be deployed on attack submarines because this "could reduce the security of the U.S. deterrent fleet of Polaris/Poseidon submarines not only by forcing a rapid growth of Russian antisubmarine-warfare capabilities but also by eliminating the distinction between tactical and strategic submarines and thereby removing the current tacit inhibition against attacks on strategic submarines." It is not clear that the Russians would increase their antisubmarine-warfare capabilities more rapidly than they otherwise would or that, if they did, the effect would be an increase in the vulnerability of our Polaris/Poseidon force. (We assume that some of the increased capability would be allocated to defend against attack submarines equipped with cruise missiles.) We are not aware of any "tacit inhibitions" on attacking Polaris/Poseidon forces that do not apply to attack submarines, with or without cruise missiles.

Tsipis states that "the U.S. has opposed the inclusion of cruise missiles in the numerical quota for strategic delivery vehicles, because—given the visual indistinguishability of the different types of cruise missile—such a provision would prevent the deployment of tactical cruise missiles as well as strategic ones. The U.S.S.R., on the other hand, insists on including all cruise missiles potentially capable of long-range missions in the quota for strategic weapons, precisely because there would be no way to determine whether a given cruise missile deployed by the U.S. is a tactical weapon or a strategic one." We are not aware of any instance where the U.S.S.R. has opposed a limitation provision on the grounds of verification difficulties. We believe the U.S. opposes limits and the U.S.S.R. favors them because cruise missiles are useful weapons and the U.S. is ahead in the technology. (It is not, however, ahead by "at least 10 years," as Tsipis states.)

Although we have other differences on points made by Tsipis, we shall limit the rest of our comments to our most serious disagreement with his article: the contention that it is technically pos-

Scientific American, August, 1977; Vol. 237, No. 2. 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 C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 35 cents.

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

Name

New Address

Old Address

"Any talk about one exclusive form of energy is just irresponsible nonsense. This country will need all the energy it can get from all sources to meet the challenge that confronts us."

"We're not running out of gas. We're running out of cheap, readily available gas. That's an important distinction."

Are We Running Out of the Fuel of the Future?

No, say the experts. Despite recent shortages, clean, efficient natural gas will play a significant role well into the 21st century. The truth is, we simply cannot do without our most modern form of energy.

Hugh Downs interviews Robert E. Seymour, Chairman of the Board, Consolidated Natural Gas Co., and Chairman, American Gas Association.

HUGH DOWNS:
Last winter the U.S. experienced severe and disruptive shortages of natural gas. We are now officially in the middle of an energy crisis that President Carter has said requires waging the moral equivalent of war. Yet you have stated that you're optimistic about the future of gas energy.

ROBERT SEYMOUR:
I'm indeed optimistic. Let's look at the facts. Geological experts estimate that we have at least 35 and possibly as much as 60 years of supply from conventional sources, and that figure rises steeply when you add the gas from supplemental and non-conventional sources. So there is cause for optimism as far as the future of gas energy is concerned. Gas is certain to play a major role in the U.S. energy picture well into the 21st century. I consider that good news, because there is simply no other form of energy that could take over the load gas is carrying today.

HUGH DOWNS: **Just what is that load?**

ROBERT SEYMOUR:
Gas supplies over half of the energy users in the residential-commercial sector, and the largest share of energy consumed by industry. You get an even better perspective by looking at it this way: the gas industry supplies almost three times as much energy per year as the electric industry.

HUGH DOWNS:
That's something most people aren't aware of, I suspect.

ROBERT SEYMOUR:
I'm afraid you're right. President Carter has called natural gas our most precious fuel, and I suspect that in addition to its high-energy content his assessment was due to the fact that natural gas just doesn't pollute air, land, or water. Far from being *old-fashioned,* natural gas is really our most modern, most timely fuel.

HUGH DOWNS:
The critical question, it seems to me, is how does the cost of our "most precious," "most modern" fuel compare with other energy forms, let's say electricity?

ROBERT SEYMOUR:
In terms of energy bought by residential users, electricity

IMPORTANCE OF NATURAL GAS TO U.S.

Types of Users	Gas Customers Number	Gas Customers % of All Users by Market	Gas Consumption Amount	Gas Consumption % All Energy Consumed (by Market)
Residential/ Commercial	44,753,000	54.6	7.6 quads	42.7
Industrial	180,000	44.5	11.7 quads*	57.3*
Gas Transmission & Distribution	—	—	0.6 quads	—
Total	44,933,000	—	19.9 quads	—

*Includes gas used for electric generation.
Excluding transportation, gas provides 35.9% of all energy consumed in the U.S.

"Conservation is vital but conservation alone is not enough."

today costs from three to five times as much as gas. It depends on where you live.

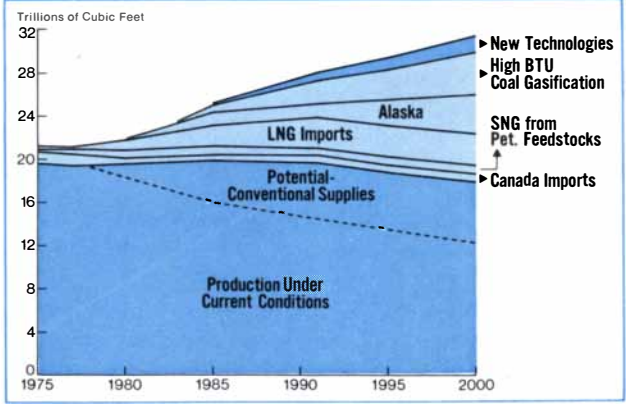
HUGH DOWNS:
Why then all the talk about an all-electric economy?

ROBERT SEYMOUR:
Well, you know that talk about one exclusive form of energy is just irresponsible nonsense. This country will need all the energy it can get from all sources to meet the challenge that confronts us.

HUGH DOWNS:
That's what I've heard. Nevertheless, even if gas continues to play a vital role in our total energy picture, isn't it likely to be a diminishing role over the next 20 or 30 years?

ROBERT SEYMOUR:
Not necessarily. There are knowledgeable people within the gas industry who feel that in the year 2000, gas will

FUTURE TOTAL POTENTIAL SUPPLIES OF GAS



contribute the same percentage to our total energy demand as today. Now that would mean increasing today's output by more than 50 percent.

HUGH DOWNS: **Do you think that's likely?**

ROBERT SEYMOUR:
I think it's possible, provided a lot of things go right. But that's talking about 30-32 quads of energy per year. I'm counting on 25 quads—versus 20 today. I'm a cautious optimist.

HUGH DOWNS:
You'd better explain quads before we go on.

ROBERT SEYMOUR:
One quad is a quadrillion British thermal units. That's the same as a trillion cubic feet of natural gas. Or 182 million barrels of oil. Or 42 million tons of coal. Or 293 billion kilowatt-hours of electricity. It's a way of making astronomical figures manageable.

HUGH DOWNS:
Where will these astronomical amounts of natural gas—25 or 30 quads—come from?

ROBERT SEYMOUR:
The largest portion will continue to come from conventional sources, including Alaska and the Outer Continental Shelf. Approximately 18-21 quads, I would say. Then you

U.S. INVESTMENT FOR ELECTRIC VS. GAS

Present (1975 Data)	Invested Plant ($ Bil.)	Delivered Energy (Quads/Yr)	Plant Investmt. Per Delivered Quad ($ Bil.)
Electric Utilities	161[1]	7.0	23.0
Gas Utilities	51	20.2	2.5
Future Increments			
Electric Utility			
—Coal Electric (1000 Mw)	0.9	.021	42.9
—Nuclear (1000 Mw)	1.1	.021	52.4
Gas Utility			
—Coal Gasification (250Mcf/d.)	1.2	.092	13.0
—LNG[2] (365 Bcf/yr)	1.6	.36	4.4
Incremental U.S. Investment for 10 Additional Quads			
—Electric (½ Coal + ½ Nuclear)	476	10	47.6
—Gas (½ LNG + ½ Coal Gas)	87	10	8.7

[1]Edison Electric Institute [2]Includes only U.S. Investment

have the supplemental supplies—liquefied natural gas, synthetic natural gas from petroleum feed stocks, and coal gasification. These supplemental supplies will build up gradually until the end of the century, when they should account for at least 6 quads.

HUGH DOWNS: **Or maybe more?**

ROBERT SEYMOUR:
Yes maybe ten or twelve. It's very possible. Now further down the line we should have gas from unconventional sources: the geopressurized geothermal zones in the Gulf of Mexico and various tight rock formations in the Rocky Mountain and the Appalachian area, and even hydrogen from the sea. These are new technologies, and it's hard to be sure about the amounts or the time frame.

HUGH DOWNS:
How about the cost? All energy seems to be costing more. Will that be the case with these new sources?

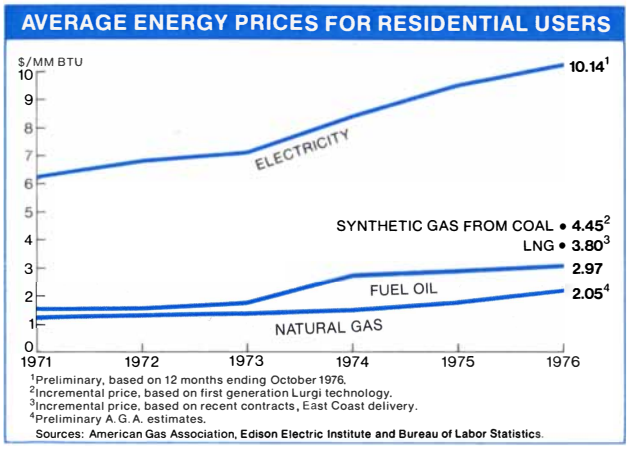
ROBERT SEYMOUR:
Well, there are a great many uncertainties about the unconventional sources. But we do know a lot about the supplementals, and even more about the cost of conventional sources, of course. Let me start with those.

Today it costs a producer—the people from whom we, the utilities and pipelines, buy our gas—several times as much to drill the same well as it did twenty years ago. But that's only part of the story. Today producing companies frequently have to drill much deeper—several miles down—or locate their rigs in much deeper water, or both. These costs are tremendous—but it's still the cheapest energy we can get. Now the problem is that there just isn't enough of it to

meet our country's growing needs. That's where the supplementals come in.

HUGH DOWNS:
Supplemental gas will be more expensive but still cheaper than other forms of energy?

ROBERT SEYMOUR:
Definitely. Take coal gasification, potentially the largest and also probably the costliest of the supplementals. First generation coal gas is estimated to come in at a little under $4.50 per thousand cubic feet. Now compare that with the cost of its only major alternative—coal-generated electricity, which is more than $10.00 for the same amount of energy. So coal gas costs less than half as much. Does that answer your question?



HUGH DOWNS:
Yes. I'd like to stay with coal gasification for a minute. It's a very important subject, it seems to me, because of the enormous coal reserves we have in the United States. What about the environmental aspect?

ROBERT SEYMOUR:
Air pollution associated with coal gasification plants is estimated to be about one-tenth that of scrubber-equipped electric power plants producing the same number of usable Btu's. Another environmental point: coal gasification uses only about one-eighth as much water.

HUGH DOWNS: **That's pretty persuasive.**

ROBERT SEYMOUR:
Yes. But remember, I'm not arguing against the diligent use of coal to meet the electric needs of our country. I'm simply saying that these vast coal supplies must be channeled in ways that will best provide our long-term base-load energy needs. Nor am I—or the gas industry—trying to belittle the urgent need to develop alternative future energy sources such as solar.

HUGH DOWNS:
I'm glad you said that. The National Space Institute, for instance, has been making a case for solar energy for some time—a very persuasive case, I think.

ROBERT SEYMOUR:
I agree. The gas industry has an active research program investigating the feasibility of using natural gas to assist solar in heating residential and commercial buildings.

HUGH DOWNS:
There's one more subject I'd like to cover, and that's conservation. It's certainly an important part of President Carter's energy plan. What's the gas industry's position on that?

ROBERT SEYMOUR:
As to the need to conserve we agree 100%. As an industry, our track record in conservation goes back almost ten years when we started our Energy Use Analysis program for builders, architects, and large volume users. We've added a number of other programs since then to further improve the efficiency of appliances and space heating systems. Now these are efficiency improvements on what is already the most efficient form of energy there is.

But conservation alone is not enough. Even at an economic growth rate of two percent, which the President has used in projecting our future energy requirements, we'll need to develop all the new sources we have discussed.

HUGH DOWNS:
Let me see if I can summarize, then. First, our potential reserves are sufficient to last us well into the 21st century. We're not running out of natural gas, just out of cheap, easily-available gas. That's important because of the leading role gas plays in the U.S. energy picture today.

Second, there is no realistic alternative to gas. Right now no other form of energy could take over the enormous load carried by gas today.

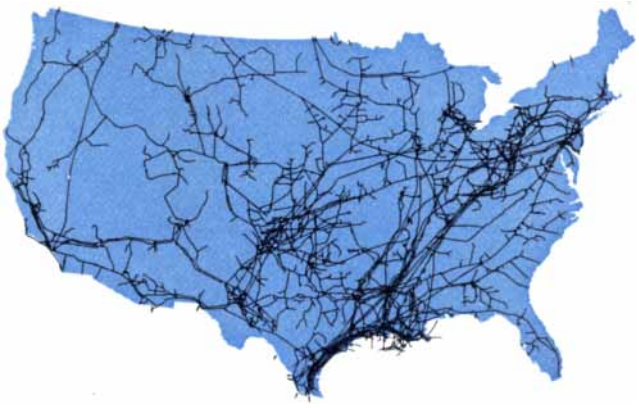
Third, although the cost of production will go up, gas will still be less expensive and more efficient than the only realistic alternative: electricity.

And fourth, gas is the cleanest, most environmentally acceptable fuel.

ROBERT SEYMOUR:
I would add one last, tremendous advantage: the gas industry's 1-million mile network of pipelines and gas mains. It's the most modern and the most efficient energy delivery system ever devised—and it's in place, waiting to be fully utilized.

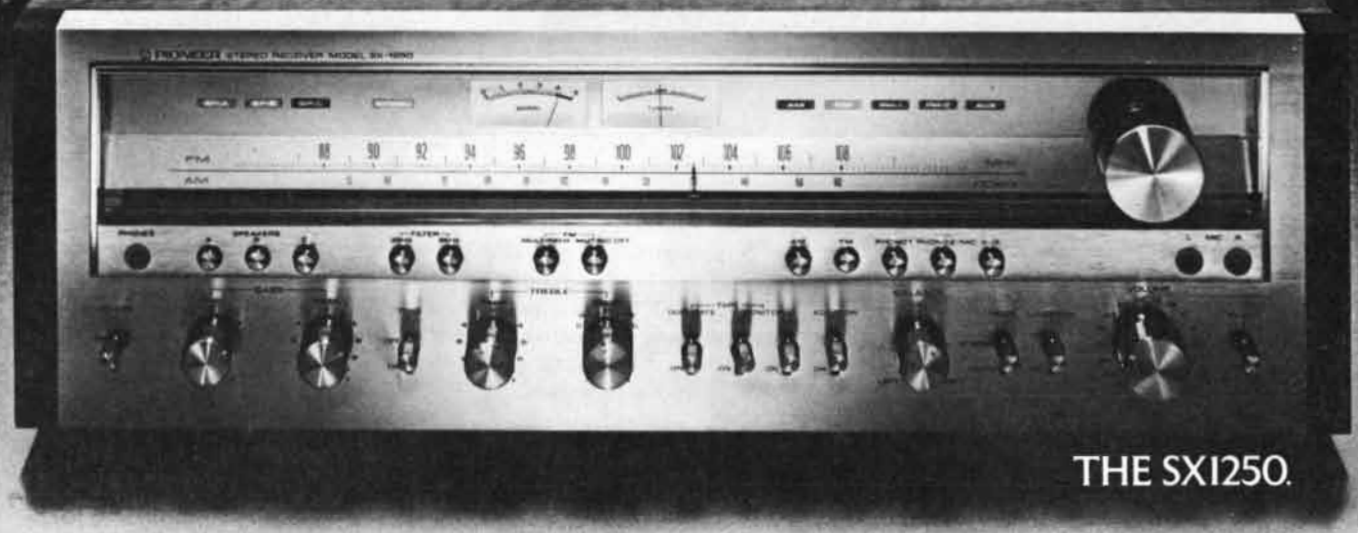
HUGH DOWNS:
All of which means gas will play a significant role well into the 21st century.

ROBERT SEYMOUR:
No doubt about it, when you look at the facts.



Built at a cost of $50 billion and irreplaceable today, the U.S. gas transmission and distribution network of 1 million miles is one of the most efficient methods to transmit energy, consuming less energy per unit shipped than, for instance, electricity.

WHY THE FIRST HIGH POWERED RECEIVER IS STILL THE BEST HIGH POWERED RECEIVER.



WHEN YOU'RE NOT IN A RUSH TO CATCH UP, YOU'VE GOT THE TIME TO BUILD THINGS RIGHT.

When Pioneer first introduced the 160 watt* SX 1250 last year, it prompted our competitors to hastily introduce a bevy of high powered receivers.

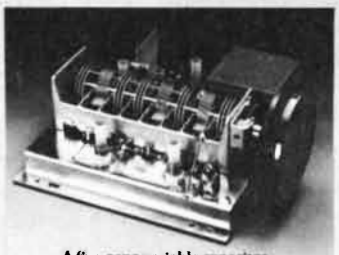
Unlike the others, however, the SX 1250 wasn't a rush job. And the time and care that went into it can both be seen and heard.

Inside the SX 1250, for example, you'll find that we took the time to shield every critical section. So spurious signals from one section can't leak into another. And dirt and dust can't get in to affect performance. So the SX 1250 not only produces crisp, interference-free sound when it's new, but still sounds great as it grows old.


An extraordinary power supply, for an extraordinary amount of power.

In our power supply, instead of finding a conventional transformer, you'll find a heavier, more advanced toroidal-core transformer. It's less susceptible to voltage variations. And less likely to leak noise. Which means you get a cleaner, clearer sound.

And where most high powered receivers come with a three, or four gang variable capacitor for FM tuning, the SX 1250 features a *five* gang zinc plated variable capacitor that cleans up FM reception much better. And helps to pull in stations that some three or four gang capacitors can't even touch.


A five gang variable capacitor for FM reception that never varies.

Obviously, these are only a few of the refinements that went into the SX 1250. But given just these few things, it should come as no surprise that the SX 1250 even weighs more than most of our competitors' high powered offerings.

So before you run out and buy just any high powered receiver, consider all the time and engineering that went into the SX 1250. And weigh your decision carefully.

High Fidelity Components
PIONEER
WE BRING IT BACK ALIVE.

*160 watts per channel minimum RMS continuous power output at 8 ohms, from 20 to 20,000 Hz, with no more than 0.1% total harmonic distortion.

sible to distinguish reliably between tactical and strategic cruise missiles with remote sensors. Tsipis argues that in order to get adequate range strategic cruise missiles will use a turbofan engine, that tactical missiles will use the cheaper turbojet engine and that with satellite sensors it will be possible to distinguish between the infrared signatures of the two. Actually it is extremely unlikely that even in the absence of clouds it would be possible to verify that missiles with turbofan engines were not being tested, particularly given the fact that the observations would have to be made against a background of thousands of aircraft flights, most of them by aircraft with turbofan engines larger than those used for cruise missiles. The case for using a turbojet engine with tactical missiles is less strong than Tsipis suggests. Indeed, the tactical version of the Navy's sea-launched missile has the same turbofan engine as the strategic version.

Tsipis suggests that the size of cruise missiles would be helpful and that limiting the size of the missile to a volume of half a cubic meter would allow effective tactical missiles and rule out strategic missiles. We question whether such a limit would do either. The volume of tactical missiles with heavy high-explosive warheads can be expected to overlap that of strategic missiles with much lighter nuclear warheads.

Finally, Tsipis states that "the new technology that has made cruise missiles possible can also be applied to the development of [suitable] monitoring systems." We believe the technologies needed for the two purposes have almost nothing in common.

We conclude that the military and economic case for strategic cruise missiles is a stronger one than Tsipis presents and that verification of compliance with an agreement that proscribes them while allowing tactical cruise missiles will be much more difficult than he suggests.

JOHN DEUTCH

Professor of Chemistry

GEORGE RATHJENS

Professor of Political Science

JACK P. RUINA

Professor of Electrical Engineering

Massachusetts Institute of Technology
Cambridge

Sirs:

The letter from Deutch, Rathjens and Ruina raises 12 objections to my article. Six are in effect forcefully stated personal opinions. Three are questions of technical judgment. Three are matters of fact.

The first of the personal opinions addresses the question of the expected effectiveness of air defenses against cruise missiles. Although these subsonic vehicles have a small radar cross section when they are seen head on, they offer a considerably larger cross section when they are seen sideways. Terminal defenses consisting of radar-guided fast-firing cannon and/or surface-to-air missiles could be effective against cruise missiles that are not protected by active or passive countermeasures. A future look-down-shoot-down Russian capability could also be effective against them. If strategic cruise missiles are deployed, they could remain in the U.S. arsenal until the end of the century. It seems to me somewhat optimistic to imply that they will remain largely invulnerable throughout their life cycle even to future countermeasures.

The second objection refers to the statement in the article that strategic cruise missiles are operationally inferior to the existing strategic triad. A cruise missile's subsonic speed and the extended period of time it may spend flying over hostile territory certainly makes it inferior to a warhead reentering ballistically. Its range and launching characteristics, described in the article, make it inferior to a submarine-launched ballistic missile. Finally, its unrecallability, its inflexible targeting properties and the fact that it does not carry either passive or active countermeasures make it operationally inferior to a manned strategic bomber. Cruise missiles are, because of their low acquisition and life-cycle costs, considerably less expensive than any of the other three strategic weapons, but they are not necessarily as effective.

The third objection is that the cruise missile would enhance the survivability of the B-52. This the article explicitly states. That fact, however, does not make the cruise missile an effective replacement for any of the existing members of our strategic triad.

The fourth objection is that Russian deployment of strategic cruise missiles will not encourage the deployment of air defenses in the U.S. First, the article states that such a deployment would be induced by the acquisition of long-range cruise missiles by third countries. Second, I believe the deployment of strategic cruise missiles by the U.S.S.R. would undoubtedly support arguments in the U.S. in favor of building such air defenses.

The fifth objection is that the deployment of cruise missiles by the U.S. will not accelerate the growth of Russian antisubmarine-warfare capabilities. The traditional Russian response to an emerging offensive threat is to strengthen defenses against it. Therefore U.S.

Stretch your cramping your

If you think an economy car would put you in a compromising position, look again.

With a Volkswagen you could actually wind up with more room and comfort than you have now.

Our Dasher, for example, comes in a beautifully appointed 2-door hatchback or 4-door sedan, with fully reclining bucket seats, plush carpeting, and enough room for five passengers.

In fact, the Dasher beats just about every car in its class in combined interior room and trunk space. And in a wagon, it has more cargo area than any car in its class.*

Then there's our Rabbit, which has been catching everyone by surprise. Even Detroit.

To start with, it takes off like a shot. With acceleration from 0 to 50 mph in an amazing 7.7 seconds. (That's faster than a Triumph Spitfire.) And with front-wheel drive that literally pulls you around corners.

It also has more room for

*Based on EPA Interior Volume Index measurement, published in the official 1977 EPA Guide 2nd Edition.

dollars without style.

people than 25 other cars you could buy (including Monza, Mustang II, Pinto, Toyota Celica, and Datsun B-210). And more room for luggage than 53 other cars. (Would you believe, with the rear seat folded down it can hold 21 bags of groceries. And with the seat folded up it still has more trunk space than a Cadillac Seville.)

Finally, we come to Scirocco, our true sports car that gives you extravagance and plain good sense at the same time.

Last year Scirocco was the Trans Am Champ for cars under 2 liters. And speaking of style, the Scirocco was just named one of the "25 best-designed factory-made products available in America today" by Fortune Magazine.**

It also gives you something few sports cars can offer: room for four and more trunk space than a Ford LTD II.

Yet (amazingly!) a Scirocco has the same fuel economy as a Rabbit.

Both Scirocco and Rabbit get 37 mpg on the highway and 24 mpg in the city. Dasher gets 36 mpg highway, 24 mpg city. (Of course, these are EPA estimates with standard transmissions. Your mileage may vary depending on how and where you drive, optional equipment and your car's condition.)

With all this going for our Volkswagens, we suggest you stop in to your Volkswagen dealer for a test sit.

It's one way to get through the energy crisis without getting bent out of shape.

Economy without sacrifice.

**Fortune, May 1977.

sea-launched cruise missiles will at least provide added credibility to arguments in favor of accelerated antisubmarine-warfare capabilities in the U.S.S.R.

The sixth objection is that in antisubmarine warfare there is no distinction between hunter-killer submarines and missile-carrying submarines. I would tend to agree with the convincing arguments offered by Professor Harvey Brooks in his book *The Future of the Sea-based Deterrent,* in which he shows that such a distinction may very well apply in Russian antisubmarine-warfare operations.

The first of the three questions of technical judgment raised by Deutch, Rathjens and Ruina is that the U.S. is not 10 years ahead of the U.S.S.R. in technology related to cruise missiles. However, both Malcolm R. Currie, the former Department of Defense director of research and engineering, and William P. Clements, the former Deputy Secretary of Defense, used this 10-year figure on occasion while they were in office. At any rate, what is important here is not the precise time gap but whether U.S. security will diminish if, as my article proposes, we agree on a moratorium on further development of the strategic sea-launched and air-launched cruise missiles.

The second question of technical judgment is that it is unlikely it "would be possible to verify that missiles with turbofan engines were not being tested." This objection suggests a misreading of the article. I clearly stated that it would be possible to detect turbofan-powered cruise missiles during testing but that it would be impossible to ascertain that no such tests had occurred. A monitoring infrared-satellite system similar in function, although not as sensitive as the system that would be required to detect cruise missiles, is already under development in the U.S. There is no compelling reason to believe this is the ultimate technological achievement in satellite-borne infrared telescopes. A monitoring system that could detect cruise-missile tests thus appears feasible. On the other hand, it is impossible to construct a system that could sense that strategic missile tests are *not* taking place. Since the proposed satellite could detect the presence of cruise-missile testing but could not verify the absence of such testing, the satellite could be used only for early warning of cruise-missile development in the U.S.S.R. and could not be used for verification.

The third question of technical judgment refers to the size of future tactical cruise missiles. Here Deutch, Rathjens and Ruina seem to have misunderstood the major point of the article, which is that the price each side must pay for being able to distinguish between the other side's tactical and strategic cruise missiles is to agree to limit the size of the tactical cruise missiles and to use only turbojet engines in them. Hence the question of whether or not a tactical cruise missile can be built that will be larger than half a cubic meter, introduced by Deutch, Rathjens and Ruina, is unrelated to the central proposal made in the article.

The first factual point raised by Deutch, Rathjens and Ruina, namely that the Navy tactical sea-launched cruise missile is equipped with a turbofan, is irrelevant for the same reason: the article does not say that such an engine cannot be used on a tactical cruise missile. The article states that the nonuse of a turbofan engine is technically and economically feasible and proposes that such an engine *should not* be used in order to facilitate the distinction between tactical and strategic versions of this weapon system.

The second factual point is that the U.S.S.R. has never "opposed a limitation provision on the grounds of verification difficulties." The article does not state that it has. Instead it points out that the verification issues raised by the widespread introduction of "gray area" weapons such as cruise missiles are sufficiently different from those in previous arms-control negotiations to induce a new Russian attitude. To expect on the basis of past experience that the Russian attitude toward verification will remain steadfast in spite of the qualitative discontinuity in the arms-limitation negotiating environment caused by gray-area weapons seems unwarranted.

The third factual point is whether the technology used in U.S. strategic cruise missiles is applicable to monitoring satellites. The basic technology of large-scale integrated electronics and sensors has equally crucial applications in strategic-cruise-missile guidance systems and in sophisticated satellite-borne monitoring systems. Neither of the two would be possible in their present form without this essentially new technology.

As the first version of the article was being written the Navy decided to change the engine of the tactical sea-launched cruise missile from a turbojet to a turbofan. Neither I nor the 10 reviewers of the first draft of the article, nor other readers of that version and later ones, either were aware of the changeover or noticed the discrepancy. Thus the article contains the dated information that the tactical sea-launched cruise missile is powered by a turbojet. Although this misinformation does not affect the central issues raised by the article, I regret my inattentiveness and thank my colleagues for setting the matter straight.

KOSTA TSIPIS

Massachusetts Institute of Technology
Cambridge

CHOOSE ANY ONE OF THESE VALUABLE COMPLETE WORKS

as a trial member of the BOOK-OF-THE-MONTH CLUB®

The Suggested Trial: You simply agree to buy four Club choices within a year at substantial savings on most books you choose

CARL SANDBURG's *Pulitzer Prize-Winning* ABRAHAM LINCOLN

The entire six-volume set

More than a portrait of a great American, this monumental 6-volume biography is a tapestry of America in crisis and resolution. Winner of a Pulitzer Prize for history... a work of timeless beauty and humanity.

for only $18.50

PUBLISHER'S LIST PRICE: $120

THE STORY OF CIVILIZATION

by WILL and ARIEL DURANT

Eleven matched volumes

This multi-volume history of all mankind — ancient and modern, Occidental and Oriental, is one of the great literary achievements of the twentieth century.

for only $22.50

PUBLISHER'S LIST PRICE: $244.45

THE NEW COLUMBIA ENCYCLOPEDIA

New fourth edition . . . completely revised after a lapse of twelve years

Cited in earlier editions as the best single-volume reference work, this latest illustrated edition has over 7000 new articles from world affairs to sports, many useful features: list of Nobel Prize winners...tables of rulers of many countries...Third World persons, events, art forms and places.

for only $15

PUBLISHER'S LIST PRICE: $79.50

THE COMPACT EDITION of the OXFORD ENGLISH DICTIONARY

Contains the complete contents of the 13-volume *Oxford English Dictionary*, generally considered the final arbiter of the meaning, origin, history and usage of words in the English language. Bausch & Lomb magnifying glass included.

for only $17.50

PUBLISHER'S LIST PRICE: $95

IF YOU are not aware of the many clear advantages, both economic and intellectual, of shopping in America's Bookstore®, the dramatic opportunity outlined here is well worth your attention. Until now, high retail prices have put the magnificent works shown here beyond the reach of many reading families. However, as a trial member of the Book-of-the-Month Club, you can choose any one of these works at an extraordinary saving — as much as $221.95.

This bounty for booklovers is not a one-time affair but is a demonstration of a continuing benefit. Not only will you be able to keep abreast of the important new books you are eager to read, but if you continue your membership you will enjoy similar savings on expensive library sets and high-priced reference works. It's a wonderful way to build a well-rounded library and save money, too.

BOOK-OF-THE-MONTH CLUB, INC.
Camp Hill, Pennsylvania 17012

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1927: "Colonel Charles A. Lindbergh left New York on Friday morning, May 20, at 7:52 A.M. Eastern Standard Time and reached Paris on Saturday, May 21, at 5:24 P.M. E.S.T., after flying for approximately 33½ hours and covering about 3,650 miles. The average speed of the flight was 109 miles per hour. Lindbergh had in the air-cooled Wright J engine, or 'Whirlwind,' a well-tried mechanism of remarkable lightness, efficiency and endurance. The design of this engine was begun on February 28, 1920. It has passed through a long development in successive models: the J-1, J-2, J-3 and J-4. The J-5 successfully met the spectacular endurance test of April 12–14, when a Bellanca plane stayed aloft for 51 hours 11 minutes 25 seconds, with a stock engine that had already had some 179 hours of flight. Moreover, its fuel economy is equal to that of the very best engines ever built. On the flight to Paris the fuel consumption averaged less than 13 gallons per hour, or more than 8½ miles to the gallon. The *Spirit of St. Louis* is a stock monoplane built by the Ryan Airlines of San Diego. The only alterations made were in the addition of 10 feet to the wing span, increasing it to 46 feet, in installing the huge gasoline tank of 425 gallons capacity in lieu of the passenger compartment and in totally enclosing the pilot's cockpit. The enclosing of the cockpit was intended to increase the speed and fuel economy, since the slightest break in the fuselage means additional air resistance. Speed and economy were purchased, however, at the expense of visibility. Lindbergh could see only through a small aperture in front that he called his periscope and through the window at his left. The plane fully loaded for the trip weighed 5,200 pounds. This meant a loading of 26 pounds per horse-power and, with a wing area of about 320 square feet, a loading per square foot of wing of 16.2 pounds. The plane had difficulty in getting off, touching ground twice before the final get-away, and showing the slowest possible rate of climb at start. The get-away probably offered the most hazardous moment of the flight. Had the sacrifice of visibility not been made to get the last degree of aerodynamic refinement, Lindbergh might have failed to get clear. It is also a tribute to the sterling qualities of the Ryan monoplane that the huge overload was sustained without structural modification

other than the increase in wing span, and that the gross weight of 5,200 pounds is the largest sustained by a 200-horsepower engine. The actual cost of the Ryan monoplane was $6,000. It was built and ready to fly in 60 days. The instruments and the engine cost $6,900, bringing the total equipment cost to some $13,000. This can hardly be considered expensive for the result that was achieved."

SCIENTIFIC AMERICAN

AUGUST, 1877: "Richard S. Floyd, one of the trustees of the Lick Trust, has just returned from an extended European tour. He talked with many scientific men of this country and others but almost invariably found them afraid to commit themselves to an opinion about whether a greater success can be obtained with a reflecting telescope or a refracting one. Professor Newcomb of the U.S. Naval Observatory in Washington, after his visit to Europe to look into the matter for Mr. Lick, reported warmly in favor of a gigantic refractor and forwarded estimates as to its cost. He has since changed his opinion, however, and now recommends a silver-on-glass reflector of about seven feet diameter. Dr. Huggins, the well-known astronomer, also favors a large reflector, as it is better for spectroscopic analysis, his favorite study."

"Experiments have recently been conducted in France with the view of testing the illuminating power of the so-called electric candle devised by M. Paul Jablochkoff. This simple means of producing a steady electric light consists in placing two carbon pencils side by side, separated by a bar of the composition known as kaolin. On the passage of the current the carbons slowly burn down and the kaolin is consumed by the heat at exactly the same rate. The carbons are thus kept always the same distance apart, and the light playing between them is rendered constant without the aid of complex regulators. In the experiments the current was produced by a magneto-electric machine worked by a small steam-engine, and the results are described as having been eminently satisfactory. The electricity might be generated in some central establishment and laid on through wires to the several centers of illumination, just as freely as gas is at present distributed through pipes to any number of burners."

"Among the many ways in which electricity is called on to give assistance in various physical investigations, one of the most elegant and interesting is the application of the electric spark to render momentarily visible a body that is rapidly moving or changing its form. The duration of the electric spark is so short—probably not more than 1/24,000th of a second—that a body such as a rotating wheel or an oscillating rod, moving in a dark room with extreme rapidity, will, if illumined by an electric spark, seem stationary, since the wheel or rod has not time to change its position appreciably during the short instant for which it is visible."

"Sir William Thomson has endeavored to prove the age of the earth by three well-known arguments of a purely physical nature. The first is based on the age of the sun's heat, the second on tidal retardation of the earth's rotation and the third on the secular cooling of the earth. 'We must,' to quote Sir William's own words on the subject, 'allow very wide limits on such estimate as I have attempted to make, but I think we may, with much probability, say that the consolidation of the earth cannot have taken place less than 20,000,000 or more than 400,000,000 years ago.'"

"It is well known that the electrical conductivity of the metalloids selenium and tellurium increases if they are exposed to the action of luminous rays, an effect the opposite of that produced when the substances are raised in temperature. Dr. Börnstein has shown that the same phenomenon occurs with platinum, gold and silver, and his experiments lead to the probability that sensitiveness to light is a general property of all metals."

"The strike on the Baltimore & Ohio Railroad that commenced on the 16th inst. assumed such character and proportions that the state authorities of Virginia were unable to cope with it, and the interference of United States authority has been called to quell the disturbance. A strike had already commenced on the Pennsylvania Central Road, and apprehensions of similar trouble are feared in Michigan. At Cincinnati developments indicate a strike on the several roads centering at that place. The employees of the Great Western Railroad of Canada are protesting against a reduction of wages but have made no other move. Employees of the western division of the Erie road have struck, demanding the pay received before the reduction. The trouble of the Pittsburgh and Fort Wayne road is assuming a serious aspect. Meetings are being held by the employees of the Union Pacific, and an outbreak is feared. The central council of the Labor League of the United States held a meeting at Washington and have recommended moderation and to avoid strikes, and resolve that moral agitation is the strength and power by which labor can acquire tangible reformation and that mob violence and riot lead only to anarchy and the final destruction of human liberty."

THE AUTHORS

HERBERT SCOVILLE, JR. ("The SALT Negotiations"), is secretary of the Arms Control Association, a nongovernmental research organization in Washington. Following his graduation from Yale University in 1937, he did graduate work in physical chemistry at the University of Cambridge and the University of Rochester, obtaining his Ph.D. from Rochester in 1942. He has served as deputy director for research of the Central Intelligence Agency, as assistant director for science and technology of the U.S. Arms Control and Disarmament Agency and as director of the arms-control program of the Carnegie Endowment for International Peace. In addition to Scoville's present work with the Arms Control Association he is chairman of the Council on Foreign Relations SALT Working Group.

MICHAEL J. DISNEY and PHILIPPE VÉRON ("BL Lacertae Objects") are both observational astronomers. Disney received his Ph.D. in theoretical astronomy from the University of London in 1968. After two years at the University of Arizona, however, he took up the nomadic life of the observational astronomer, spending four years at the Mount Stromlo Observatory in Australia and two years at the University of Groningen in the Netherlands. He is currently working at the Royal Greenwich Observatory in Sussex, where he spends much of his free time "trying to catch one of the monster carp in the moat of the observatory." Véron, a native of France, attended the École Polytechnique in Paris and then spent two years as an army officer in Algeria. In 1962 he began studying radio astronomy at the Paris Observatory and, after working for a year with Allan R. Sandage at the Hale Observatories in Pasadena, obtained his Ph.D. from the University of Paris in 1967. Now on a two-year leave of absence at the European Southern Observatory in Geneva, Véron devotes his spare time to the history, astrology and other lore of comets.

W. S. BOYLE ("Light-Wave Communications") is executive director of the Research and Communications Sciences Division of Bell Laboratories. A native of Nova Scotia, he served in the Royal Navy flying Spitfires during World War II. After the war he attended McGill University, receiving his Ph.D. in physics in 1950. He joined Bell Laboratories in 1953 and contributed to the development of solid-state lasers. Later he and Donald F. Nelson of Bell Laboratories developed the first continuously operating ruby laser. Boyle is coinventor of the charge-coupled device, a microelectronic technology with expanding applications. He spends his free time cruising on Long Island Sound in his 33-foot ketch.

HENRY N. POLLACK and DAVID S. CHAPMAN ("The Flow of Heat from the Earth's Interior") are respectively professor of geophysics at the University of Michigan and assistant professor of geophysics at the University of Utah. Pollack studied geology at Cornell University and the University of Nebraska and obtained his Ph.D. in geophysics from the University of Michigan in 1963. He then did postdoctoral research with Francis Birch's geothermal group at Harvard University. In 1970 Pollack met Chapman while he was making geothermal measurements high on the Central African Plateau. Chapman, a native of Canada, was educated at the University of British Columbia. For six years thereafter he was a volunteer teacher in Zambia, first in a secondary school and then as a lecturer in physics at the University of Zambia. His collaboration with Pollack began in Africa and continued at the University of Michigan, where he received his Ph.D. in geophysics in 1975.

T. J. DAWSON ("Kangaroos") is professor and head of the School of Zoology at the University of New South Wales in Australia. "My interest in the biology of Australian animals," he writes, "stems from my boyhood wanderings in the bush around the small town of Nyngan in the dry inland sheep country of Australia." After obtaining his doctorate from the University of New England in Australia in 1964, he spent three years in the U.S., initially with Knut Schmidt-Nielsen at Duke University and then with Theodore Hammel at Yale University. Since returning to Australia he has been working on aspects of the environmental physiology and ecology of marsupials and monotremes, with emphasis on the kangaroo. "The field aspect of the work I enjoy greatly," Dawson writes, "as does my family. Some disgruntled colleagues even suggest that my work in the Australian outback is merely an excuse to wander in and observe this beautiful country and its animals."

A. E. WALSBY ("The Gas Vacuoles of Blue-Green Algae") is Reader in the department of marine biology at University College of North Wales. After obtaining his bachelor's degree in botany from the University of Birmingham he did research on the blue-green algae with G. E. Fogg at the University of London, receiving his Ph.D. in 1965. From 1971 to 1973 he was a research fellow at the University of California at Berkeley, where he continued his work on gas vacuoles in collaboration with Daniel Branton. Now back in the United Kingdom, he spends his free time playing the cello in chamber-music ensembles with friends and, when the Welsh weather allows, tending the heather and vegetables in his garden.

HOWARD S. IRWIN ("The History of the Airflow Car") is president of the New York Botanical Garden and the Cary Arboretum. He joined the Botanical Garden in 1960, after receiving his Ph.D. in botany from the University of Texas. The Airflow car has been a lifelong avocation of his, both as a hobby and as a source of enlightenment about the automobile's role "as a servile object and as an expression of our culture." He owned his first Airflow as a teenager and recently acquired a third.

JAMES A. NATHANSON and PAUL GREENGARD (" 'Second Messengers' in the Brain") are both at the Yale University School of Medicine. Nathanson has been interested in science since elementary school, when he performed experiments on learning and memory in a home basement laboratory. He attended Trinity College and then studied medicine at Yale. While at Yale he became interested in the biochemical approach to brain function and went on to obtain his Ph.D. in neurobiology in Greengard's laboratory. After graduation and a year of medical internship he spent two years at the National Institutes of Health, returning to Yale in 1976 as a resident in neurology. In Nathanson's spare time he is an amateur astronomer, and he has published a paper on the classification of galaxies. Greengard is professor of pharmacology at Yale. He first became interested in science when, as a 17-year-old in the Navy during World War II, he was sent to the Massachusetts Institute of Technology to take part in the development of an early-warning system to protect Navy ships from Japanese kamikaze planes. After the war he attended Hamilton College and Johns Hopkins University, where his interests evolved from mathematics and physics to neurobiology. After receiving his Ph.D. from Johns Hopkins in 1953 he spent the next five years doing postdoctoral work in England. Greengard became director of the biochemistry department of the Geigy Research Laboratories in 1959 and moved to Yale in 1968. "The chemicals in my brain," he writes, "devote much of their time to thinking about possible mechanisms by which the chemicals of the brain make thoughts."

SEARS ROADHANDLER:

It had to outperform a legend before it could be called Sears best steel-belted radial.

Sears original steel-belted radial tire proved its toughness and excellence by racking class wins in the famous Baja event.

It was the beginning of a legend... and a challenge to develop an even better tire. A tire that was better than the tire that beat the Baja. To outperform

its famous predecessor, the original Sears steel-belted radial, they couldn't make a tougher tire, so they made it better. Better in the areas of handling, braking and ride.

In the vitally important area of wet cornering, Sears RoadHandler demonstrated its superiority under controlled laboratory conditions. RoadHandler's bigger "footprint" allowed more water to travel along the grooves instead of between the tire surface and the wet road. This helps prevent hydroplaning (skidding on a thin film of water). In fact, laboratory tests demonstrated Sears RoadHandler to be superior in just about every area of traction.

To demonstrate the performance of this newer tire, Sears put the RoadHandler through the mill.

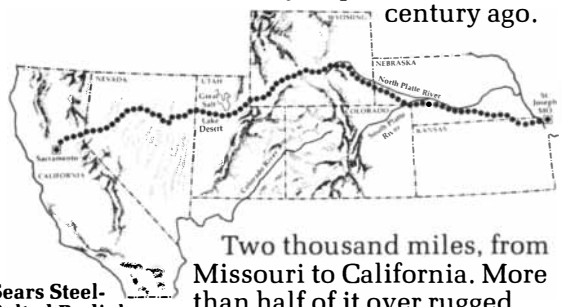
First, five gruelling weeks of day and night driving – 40,000 miles over every kind of road.

Sears Steel-Belted Radial. The tire that beat the Baja. Sears original.

Sears RoadHandler, 13% wider tread. 10% deeper tread (at center). Extra tread row.

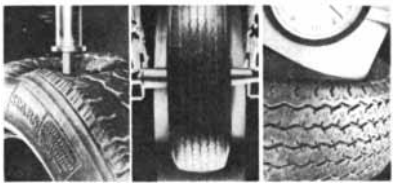
Sears ★ Best

Next, the same tires tackled the Pony Express route of a century ago.



Two thousand miles, from Missouri to California. More than half of it over rugged, trackless country.

After 42,000 miles without any tire failure, Sears RoadHandlers faced the most critical test of all. Federal Motor Vehicle Safety Standard 109. A test for new tires.



These Sears RoadHandlers not only passed, they exceeded every government requirement for strength, heat resistance and bead unseating (keeping the tire on the rim in hard turns.)

Add it all up... and it's hardly surprising that Sears RoadHandler is Sears best steel-belted radial.

Sears

Tire and Auto Centers and Catalog.

"STRAIGHT TALK, GOOD VALUES AND SATISFACTION."

Opel makes news.

The results of the Buick Opel 5-Car Showdown are in!

Opel finishes...uh...2nd.

A lot of car makers compare their car to other cars.

We compare our car *with* other cars.

In a daring, fender-to-fender competition called The Buick Opel 5-Car Showdown. In it, we pitted our Opel against four better known competitors in point-by-point, side-by-side, independently supervised tests of acceleration, cornering flatness, parking-lot maneuverability, pulling power, gradability, and a few other areas. In short, some of the things you'd like to know when you go out to shop for a car.

It was a bold move. After all, what if we didn't win?

Well, to make a long story short—we didn't. When all the tests were completed and all the figures tallied up, Opel finished second, right behind VW Rabbit. You can imagine how thrilled we were. But look at it this way: in order to finish second overall, we had to beat Toyota Corolla, Datsun B-210, and Subaru DL in a number of instances. (And in some areas, we beat VW, too, as you'll see.)

Which is a victory. And not just a moral one.

You see, we know Opel is a little dynamo of a car. But apparently no one else did. Because it seemed that whenever anyone went out looking for a practical little import, they looked real hard at the other cars in the Showdown. But hardly ever at ours.

So we wanted to show that Opel could hold its own against its famous competitors. That it should definitely be considered when you wander out to shop.

In other words, we were confident Opel was good enough to take on this competition.

Opel announces 4-doors.

Then we went one step further. And made Opel good enough to take on your family. By offering you our new, just-introduced-in-this-country Opel 4-door Sedan.

After all, if you've got a family (or maybe some friends you like taking along on spirited spins around the countryside), you need to carefully examine a car's ins and outs. So we gave you two more.

Easy come. Easy go.

Final Test Results (Showdown Point Summary)	VW Rabbit	Buick Opel	Toyota Corolla	Subaru DL	Datsun B-210
Interior Noise	5	5	5	5	5
EPA Interior Room Estimates	5	3	4	2	1
EPA Trunk Capacity Estimates	5	3	2	4	1
Pulling Power (30mph 3rd gear) (55mph 4th gear)	3	5	4	2	1
Parking Lot Maneuverability	3	5	4	2	1
EPA Mileage Estimates	3	2	4	4	5
Estimated Range	2	3	5	5	4
Maintenance Stops* (Recommended)	5	3	4	2	4
Acceleration (0-55mph, 20-35mph) (30-70mph)	5	4	3	2	1
Gradability** (30mph 3rd gear) (55mph 4th gear)	5	4	3	3	2
Cornering Flatness**	4	5	3	4	5
Steering Quickness	5	5	5	4	5
Grand Total	50	47	46	39	35

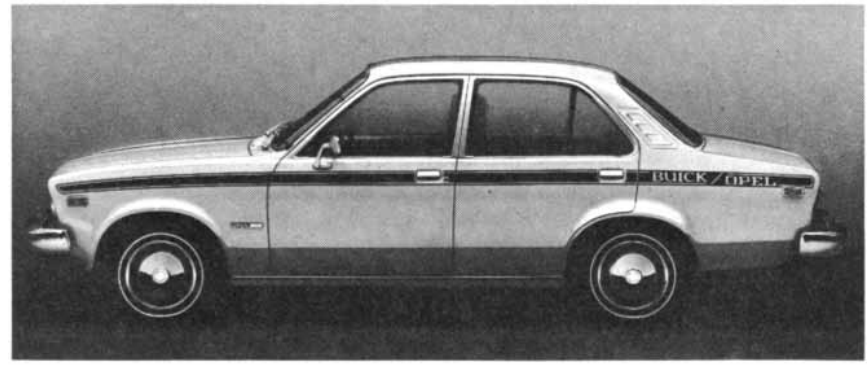
*The less frequent the number of visits over 37,500 miles of normal driving, the higher the score. The number and type of inspections, adjustments and replacements would vary by visit.

**Based upon opinion that less lean is preferable to more lean.

***Includes vehicle weight.

In California, Opel placed second on an overall basis. Individual test results vary from those shown above.

See your Buick Opel dealer for complete details.



Opel makes sense.

You can learn a lot from a test drive.

Or even a test-sit.

That will let you experience our body-contoured vinyl seats. Seats tailored for comfort and good lateral support. With seatbacks that adjust.

It'll show you—in one quick glance—how easy it is to read Opel's well thought-out instrument panel, and reach Opel's well-placed controls.

And it will reveal (most abundantly) Opel's thick, standard pile carpeting (one more way this small car lives up to its Buick name). Perhaps even startle you with an amazing sight in small cars these days: a real glove compartment . . . that locks.

What you will most certainly notice first, though—if you happen to be sitting in the driver's seat of an Opel Deluxe Coupe—is that you're not just looking at a bunch of warning lights. But gauges. A tachometer right next to the speedometer. An oil-pressure gauge. An ammeter. An electric clock with a second hand. All standard. To let the enthusiast in you keep a close eye on the kind of response you're getting from that gutsy 1.8 litre dynamo under the hood.

But to really feel how Opel goes on the road, you should really feel it. On the road.

And, oh, what a feeling.

Opel performs.

Indeed it does. As the 5-Car Showdown will testify. First of all, in tests of cornering flatness and steering quickness, none of the competitors out-ran Opel. And in tests of acceleration and gradability, only the VW Rabbit (a car much-ballyhooed for its fuel-injected engine) was able to nose out our Opel for top honors.

Opel's strong showing in these areas isn't surprising when you consider its own accoutrements. Like the dynamic overhead-cam, 4-cylinder hemi engine. Four-coil-spring-suspension. Front stabilizer bar. Rear track bar. Floor-mounted, full-synchronized, short-throw, 4-speed manual transmission (with 5-speed and automatic available). Rack-and-pinion steering. And lots more.

Nevertheless, it still may surprise you. But then, up to this point, maybe you haven't considered Opel.

Maybe you should.

More than just mileage: range.

When you do consider Opel, you'll find it quite considerate of you. With EPA mileage estimates of 36 mpg on the highway and 23 mpg in the city. (The actual mileage you get will vary depending on the type of driving you do, your driving habits, your car's condition and available equipment. Estimates lower in California.) But don't stop there.

You won't have to. Because with Opel's 13.7-gallon gas tank (largest of any of the cars in the Showdown, by the way), it's estimated range is about 350 miles.

(By the way, this range estimate is a function of both gas tank capacity and miles per gallon, and is based on multiplying the EPA combined estimate, which is 27 mpg, by gas tank capacity Actual range may vary.)

The best news of all.

If by now our intrepid little Opel is beginning to sound like an intriguing proposition, let us suggest you hurry to your Buick Opel dealer for further consultation.

Not only will he give you a brochure with complete details on the 5-Car Showdown, he'll be happy to provide you with one of those revealing test-drives we mentioned.

And, of course, he'll show you how very affordable Opel is to buy.

Come on, now. Any car with this much going for it is at least worth considering, isn't it?

BUICK OPEL

Established 1845 SCIENTIFIC AMERICAN August 1977 Volume 237 Number 2

The SALT Negotiations

Current strategic-arms-limitation talks can put a real ceiling on the quantitative arms race and restrain the qualitative race. A failure to do so may destabilize the present strategic balance

by Herbert Scoville, Jr.

The Carter Administration is in the process of reviewing national security policies and assembling its first defense budget. A central question in the deliberations is the part arms control can and should play, and the strategic-arms-limitation talks (SALT) with the U.S.S.R. are the most critical negotiations currently under way. Both the U.S.S.R., with its new family of large-payload intercontinental ballistic missiles (ICBM's) and the U.S., with its programs for replacing Minuteman ICBM's with more accurate ones, are contemplating new deployments that could greatly reduce the stability of the strategic balance and would also cost billions of dollars. The interim agreement on offensive arms that was reached in 1972 expires on October 3, and unless some new agreement is reached by that date the pressures for stepping up the arms race will be very great. Success in this round of SALT is therefore a matter of grave consequence for U.S. security and also for the size of the armament bill Americans have to pay.

The U.S. and the U.S.S.R. achieved their first broad arms-control agreements in 1972, when the anti-ballistic-missile treaty and the interim agreement were signed at the conclusion of the first round of strategic-arms-limitation talks (SALT I). The second phase of the talks (SALT II) has been occupying the two superpowers on and off since then. Now a deadline looms, because the interim agreement ran for only five years. A step toward meeting the October deadline was taken almost three years ago by the Vladivostok accords, a set of guidelines for SALT II that set numerical ceilings on the major strategic weapons: 2,400 strategic delivery vehicles (bombers and missiles), with no more than 1,320 of them to be armed with multiple independently targetable reentry vehicles (MIRV's). The good news then was that the accords for the first time put a ceiling on bombers as well as missiles, and also on the number of missiles that could be fitted with MIRV's (although not on the total number of MIRV's). The bad news was that the agreement established ceilings so high as to be in effect nonlimiting; that it was imprecise about the definition of what types of weapons—in particular cruise missiles and medium-range bombers—should be included under the ceilings, and that by failing to deal with advances in weapons technology it institutionalized a new kind of arms race, a qualitative race instead of a primarily quantitative one. These have since turned out to be significant issues. What is complicating the current phase of the SALT negotiations is that each side is attempting to deal—in different ways—with those shortcomings of the Vladivostok guidelines.

The new Carter Administration took the initiative in March, when Secretary of State Cyrus R. Vance went to Moscow with two alternative proposals. One was "comprehensive." It sought to reverse the quantitative arms race by actually reducing existing arsenals and to curb the qualitative race by restricting the modernization or replacement of existing weapons systems. The other was a fallback "deferral" proposal. It would simply have confirmed the Vladivostok ceilings, leaving for future consideration the two issues that had blocked a final SALT II agreement during the Ford Administration: the control of the American cruise missile and the Russian "Backfire" medium bomber. The Russians turned down both propositions. The deferral proposal's overall ceilings were apparently satisfactory to them, but they could not accept its deferral of the cruise-missile issue. As for the comprehensive proposal, the Russians said its particular limitations and prohibitions were calculated to freeze U.S. strategic superiority.

After that initial rejection the negotiations resumed within a three-part framework established in Geneva in May. There would be, first, a five-year treaty confirming the Vladivostok overall ceiling or a somewhat lower one; second, a protocol, to run for three years, constraining the development or deployment of cruise missiles, mobile missiles, heavy ICBM's and the Backfire, and third, a statement of principles to govern continuing negotiations on long-term, substantial reductions in strategic arms and on controlling the qualitative race. This summer's negotiations have presumably dealt with the details to be fitted into that framework.

Here I shall mainly describe and analyze the various provisions of the original comprehensive proposal put forward by the U.S. in Moscow, attempting thereby to demonstrate the nature of the major current issues in arms control, the complex interrelations of numerical and qualitative controls and the difficulty of establishing meaningful controls in an era when technological advances are becoming far more significant than such conventional measures of strategic supremacy as the number and megatonnage of each side's missiles. I shall try to relate the provisions of the comprehensive proposal to the three-part framework now under discussion.

The American comprehensive proposal went far beyond any controls that had been formally put on the table by either country in eight years of SALT negotiations. It represented in particular an effort to limit those weapons systems

most calculated to endanger the stable mutual-deterrence balance between the U.S. and the U.S.S.R., that is, systems that can give either side a "counterforce" capability threatening the land-based intercontinental ballistic missiles of the other side. Any weapon, such as an ICBM with several MIRV's, each having a combination of accuracy and yield that gives it a high probability of destroying a "hardened" intercontinental-missile launcher, is a counterforce weapon. It threatens an important component of the other side's retaliatory arsenal and thus of its deterrent strength.

The nine basic elements of the U.S. proposal (as they have been revealed so far) fell into two categories. The first four provisions imposed limits—lower than those now in effect—on the number of strategic delivery vehicles. The last five provisions addressed the problem of qualitative control by imposing various restrictions on the development, testing and deployment of new weapons.

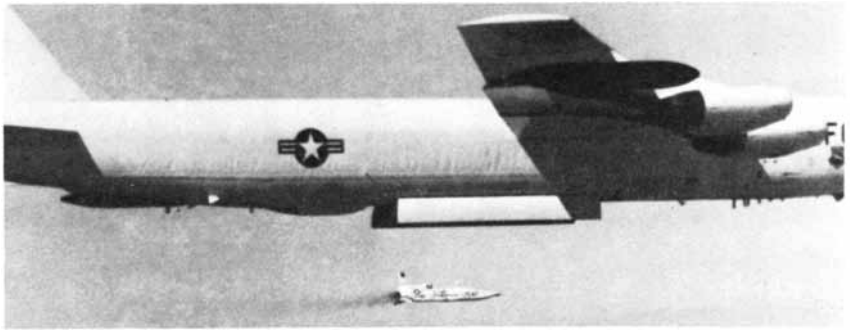
The quantitative provisions would have reduced the total number of delivery vehicles, or of ICBM's, submarine-launched ballistic missiles (SLBM's) and strategic bombers, from the Vladivostok level of 2,400 to between 1,800 and 2,000. Only 1,100 to 1,200 of the missiles could be MIRVed (compared with 1,320 under the Vladivostok terms), and only 550 of those MIRVed missiles could be land-based ICBM's (a subtotal that was not specified at Vladivostok). Finally, the number of very heavy Russian ICBM's would be reduced from the 308 established by the interim agreement to 150.

By making some assumptions about how each side might decide to allocate its forces among ICBM's, SLBM's and bombers, and also about bomb loads and the number of MIRV's that would be fitted on various missiles, it is possible to assess the impact of the lowered limits on the two sides' arsenals [*see illustration on pages 28 and 29*]. The actual allocations might be different, but that would probably not alter the strategic balance significantly. (The major uncertainties in the allocations have to do with the Russian program for installing MIRV's on their SLBM's, a program that has had low priority until now but that would presumably become more urgent, as I shall explain, if the MIRVing of land-based missiles were to be curtailed.)

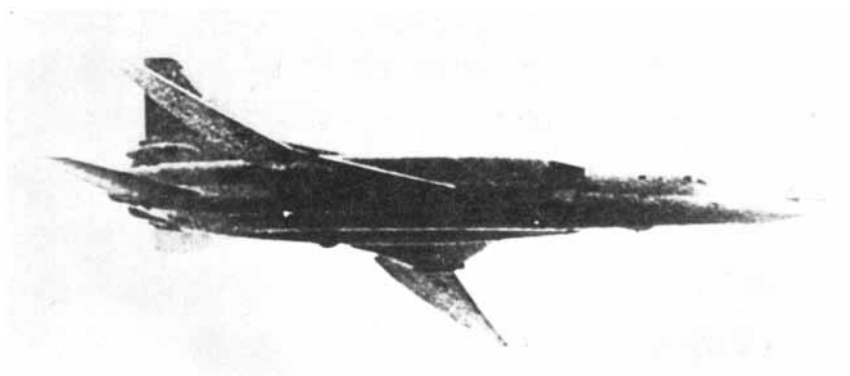
The U.S.S.R. has concentrated on larger ICBM's that can carry bigger payloads, which can be translated into either more explosive yield per warhead or more warheads per missile, or some of each. I have assumed on the basis of Russian tests that the heavy Russian ICBM's would carry eight warheads and that the light ICBM's would carry six. The U.S. land-based Minuteman III missiles carry three warheads. As for bombers, I assume that the American B-52 could carry up to 10 weapons, either bombs, long-range cruise missiles or short-range air-to-surface missiles. There is uncertainty about the bomb load of Russian strategic bombers, but the U.S. Department of Defense generally estimates two or three weapons per aircraft; the Russians have a very small strategic bomber force and so the uncertainty is of little consequence. I have excluded the Backfire medium bomber from the strategic-force calculations because there is a real question as to its value for intercontinental missions; it falls in a category similar to that of the U.S. bombers stationed in Europe within range of the U.S.S.R., what is called the forward-based system.

Reaching the lower levels contemplated by the comprehensive proposal would present different kinds of problems, involving both real and perceived inequities, for the U.S. and the U.S.S.R. The proposal would require the Russians to cut back their heavy ICBM's from 308 to 150, all of which could be fitted with multiple warheads. That requirement provoked a strong reaction from the Russians, since the U.S. has no heavy ICBM's and nothing in the comprehensive proposal would require the U.S. to eliminate any specified existing weapons. This particular controversy could have been avoided if the provision had been phrased differently, simply setting a ceiling of 150 on the number of heavy missiles the Russians could MIRV or even on the number of older Russian heavy SS-9 missiles that could be replaced by the newer SS-18. The SS-9 has never been tested with MIRV's, and even most of the SS-18's have only single warheads, so that the Russians now have only a small number of MIRVed heavy ICBM's. A limit on MIRVing, or at least on the number of new SS-18's, would be verifiable by satellite because the SS-18 requires a larger silo than the SS-9. Such a ceiling would have about the same strategic effect as a cutback because the single-warhead SS-9 ICBM's do not pose a significant counterforce threat.

Limiting the total number of MIRVed ICBM's to 550 might be particularly difficult for the U.S.S.R., which relies more than the U.S. on ICBM's. (The limit of 550 also has a "cosmetic" problem, since it was apparently picked to coincide with the present number of MIRVed U.S. Minuteman missiles.) One cannot be sure how many of the

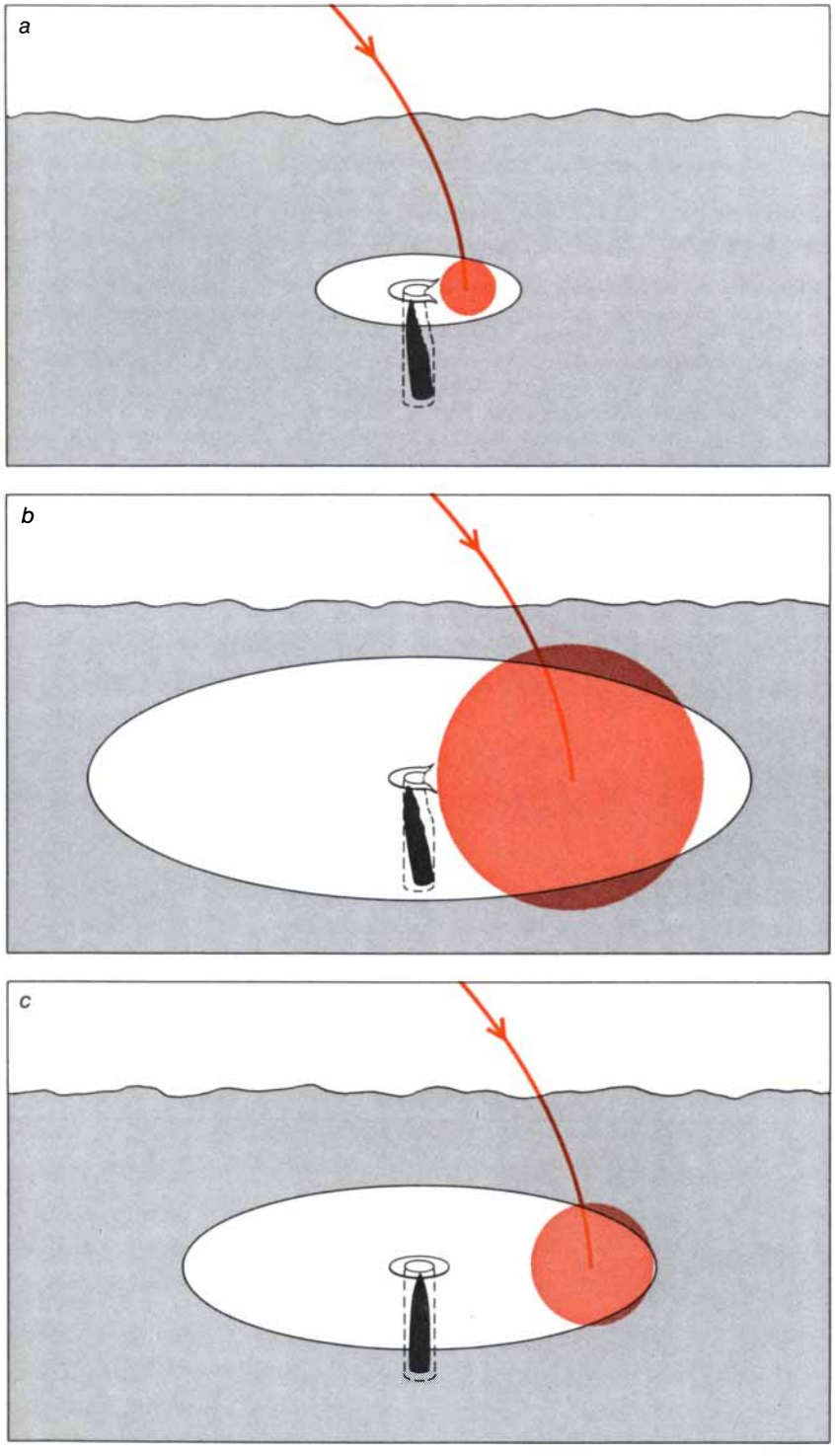


U.S. CRUISE MISSILE is a subsonic, jet-powered drone similar to the World War II "buzz bomb" but with greater range and accuracy. It can be launched from land, from a surface vessel or submarine or from the air—from a B-52 in this photograph. A major SALT issue is Russian insistence that long-range cruise missiles be included under a strategic-weapons ceiling.



RUSSIAN "BACKFIRE" is a supersonic medium bomber able to attain Mach 2 at high altitude. Its combat radius is 3,500 miles, which means it could reach the U.S. only by flying subsonically from arctic bases in the U.S.S.R. on a one-way "suicide" mission, by landing in a "third country" or refueling in flight. One issue is whether or not Backfire is a strategic weapon.

Vladivostok accord's 1,320 MIRVed delivery vehicles the Russians would have allocated to ICBM's, but in view of the low priority given to SLBM's in the past the number might well have been far more than 550. On the other hand, the Russians now have fewer than 200 MIRVed ICBM's, so that they would not be forced to scrap any. And one can argue that if the limit on MIRVed ICBM's were set much above 550, a counterforce capability would be within sight, given reasonable improvements in accuracy.



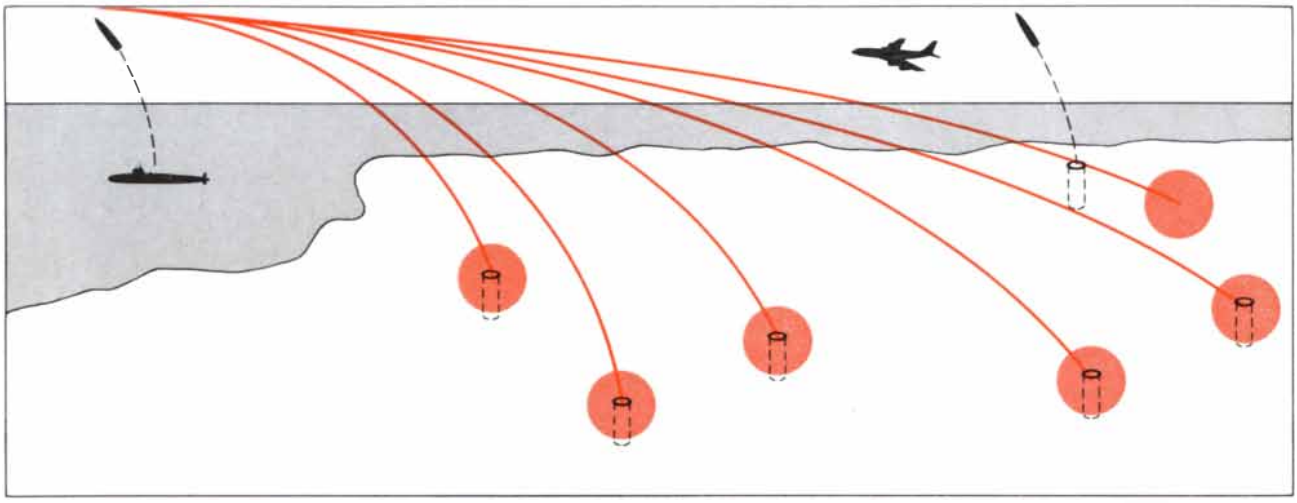
ACCURACY AND EXPLOSIVE YIELD combine to determine the probability of a missile's destroying a "hardened" target. A warhead of modest yield but with a small circular error probability (CEP) can destroy a missile silo (*a*), as can a warhead with a larger yield but a larger CEP, or less accuracy (*b*); either warhead is an effective counterforce weapon. A weapon with insufficient yield or accuracy to knock out a silo is not an effective counterforce weapon (*c*).

Merely meeting the ceiling of 1,800 to 2,000 total delivery vehicles would probably force the Russians to scrap ICBM's. They would have the alternative of reducing their SLBM force, but that stood at 840 in the spring and appears to be on the way to the 950 allowed by the 1972 interim agreement. Scrapping ICBM's would presumably be preferable to scrapping expensive submarines, which constitute the more reliable deterrent. If the Russians wanted (as they presumably would) to attain the proposal's ceiling of 1,100 to 1,200 MIRVed missiles, the low limit on multiple warheads for ICBM's would require MIRVing from 550 to 650 SLBM's. The Russians have only recently begun, however, to test a missile with two MIRV warheads for the newest of their two classes of missile-launching nuclear submarines. Such a missile will be very inferior to the 10-warhead U.S. Poseidon, which has been operational for five years, and will be even further behind the Trident I missile scheduled to be deployed in about two years. For various reasons, then, the Russians would almost surely prefer to MIRV a large number of ICBM's. The restrictions on the MIRVing of ICBM's contemplated by the U.S. proposal were therefore understandably viewed as one-sided, in that they curb the weapons the U.S.S.R. values most and the U.S. values least.

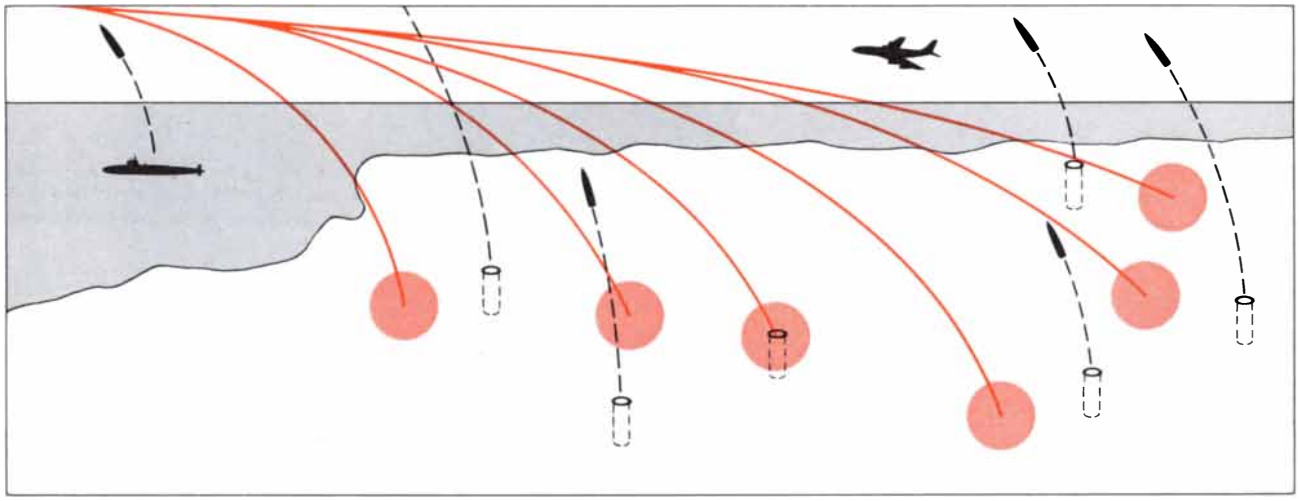
The U.S.S.R. is far behind the U.S. in intercontinental bombers, with about 135 aircraft dating from the mid-1950's, and it is apparently making little effort to catch up. The Russians would presumably reduce strategic-force levels, if necessary, by scrapping some old bombers. The U.S., on the other hand, would probably maintain a significant force of B-52's, whose large bomb loads would be a significant element in the total nuclear explosive power that could be unleashed against the U.S.S.R.

The new Russian Backfire, which has attracted so much attention, has very little intercontinental capability. It can reach the U.S. only from bases in the Russian Arctic on a one-way subsonic high-altitude mission. A two-way mission would require in-flight refueling, which has not been given high priority in Russian air tactics. The Backfire's capability as a strategic bomber—defined as a bomber that can reach the other country's territory—is certainly less significant than that of U.S. bombers based in Europe or on aircraft carriers, which can reach strategic targets within the U.S.S.R. The U.S. contends that the primary role of the forward-based aircraft is not strategic but rather to support its NATO allies in the event of a European conflict. Similarly, the Russians maintain that the Backfire is for missions against peripheral targets such as China and Western Europe.

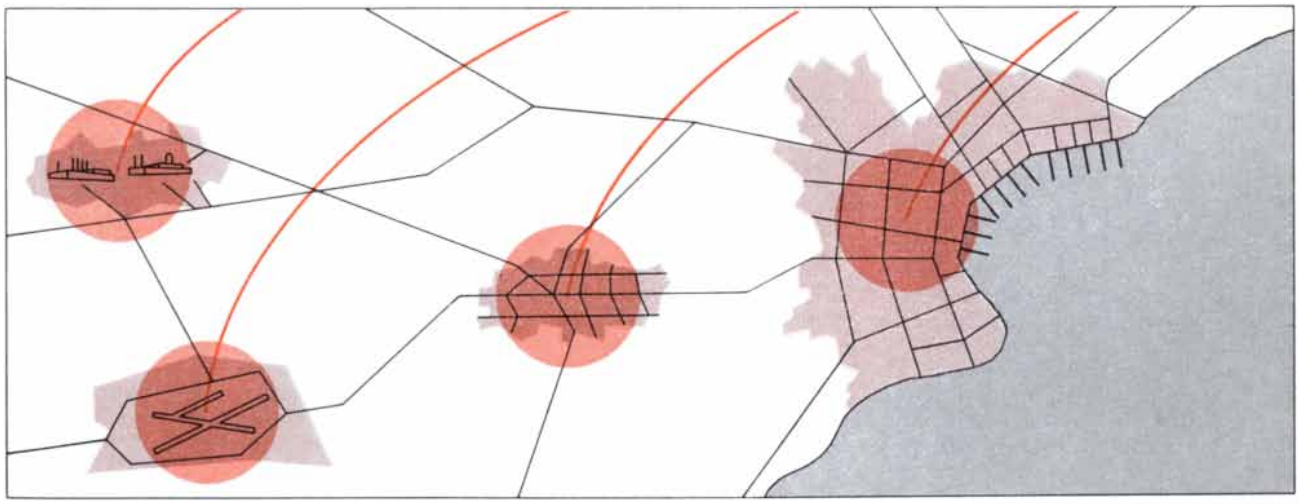
There are, then, various ways the U.S.



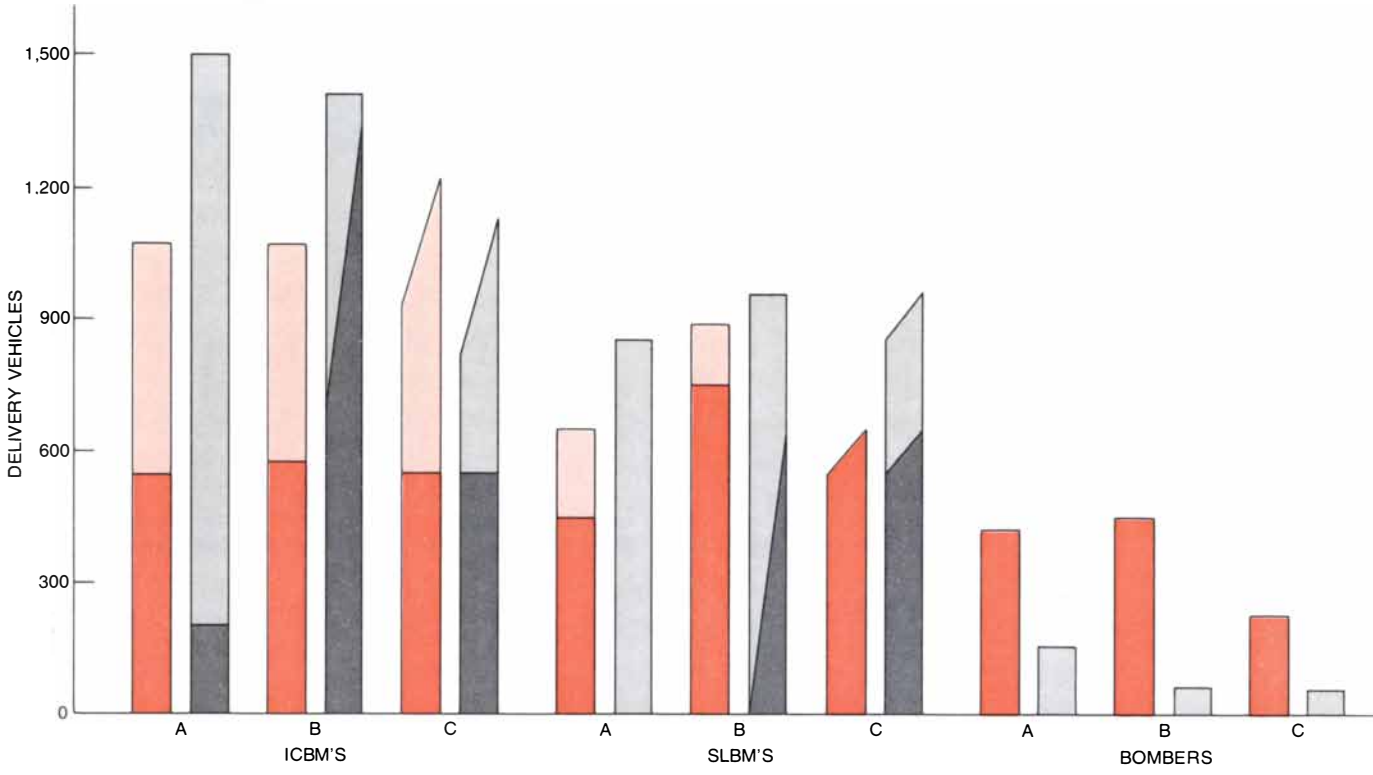
EFFECTIVE COUNTERFORCE ATTACK by missiles with multiple independently targetable reentry vehicles (MIRV's) knocks out most of the targeted intercontinental ballistic missiles (ICBM's). Bombers on alert and missile-launching submarines can still retaliate.



INEFFECTIVE COUNTERFORCE ATTACK is one in which the attacking missiles lack either the accuracy or the explosive yield to destroy many of the targeted ICBM's. Most of the ICBM's survive, and they can be launched to carry out a large-scale retaliatory attack.



EFFECTIVE RETALIATION by the targeted nation on the nation that launched a first strike is depicted. ICBM's, submarine-launched ballistic missiles (SLBM's) and bombers attack population centers, port facilities, industrial complexes and military targets other than missile silos. Far less accuracy is required for effective retaliation on such targets than is required for a successful counterforce attack.



STRATEGIC FORCES of the U.S. (*color*) and the U.S.S.R. (*gray*) are compared as of last spring (*A*) and as they might be in 1985 if the Vladivostok ceilings (*B*) or the lower ceilings set forth in the U.S. comprehensive proposal (*C*) were in force. (Actual numbers might be somewhat different, since assumptions must be made about how the two nations would choose to balance their forces.) The left-hand

and the U.S.S.R. might reduce their strategic forces to reach the ceilings contemplated in the comprehensive proposal, but the options are constrained by existing force structures, by the technology available in each country and by the two countries' respective bureaucracies. To reach the proposed overall ceiling of 2,000 to 1,800 delivery vehicles would be more difficult for the U.S.S.R., for which it would mean a cut of from 500 to 700 from present levels, as opposed to a cut for the U.S. of between 150 and 350. The U.S. could take its reduction easily in bombers or older ICBM's, whereas the U.S.S.R. would have to scrap newer ICBM's or SLBM's as well. The ceiling of 1,100 to 1,200 put on MIRVed delivery vehicles would not in itself require any reduction of existing forces by either country, but the subtotal of 550 ICBM's with multiple warheads would be hard for the Russians to take. The lowered ceiling would force the U.S. to discard some existing MIRVed missiles if it were to go ahead with proposed programs such as the Trident submarine.

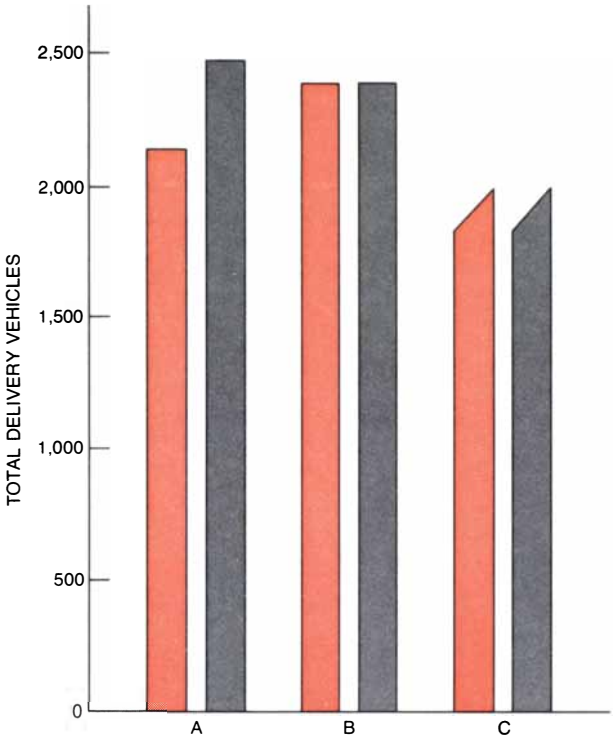
Under the comprehensive proposal the Russians might by 1985 have about a two-to-one lead in the number of ICBM warheads; the yield of their warheads would be about one megaton, compared with about 200 kilotons for the American warheads. The Russian numerical advantage in ICBM's would be somewhat smaller than it was under the Vladivostok accords but substantially larger than it is now, since the U.S.S.R. is only beginning its MIRV deployment. The Russian advantage in number and yield would be counterbalanced by U.S. superiority in accuracy and probably in reliability.

In SLBM warheads the U.S. would have a numerical lead of two or three to one. Moreover, the U.S. missiles and submarines are technologically superior, and the U.S. has a significant geographical advantage because of the ease with which its vessels can operate in the Atlantic Ocean. In bombers the U.S. would have an overwhelming lead even if the Backfire were counted as a Russian strategic bomber. The U.S. lead in bombers and SLBM warheads would more than balance the Russian lead in ICBM's. The U.S. would therefore be somewhat ahead in total missile warheads and very far ahead in total "force loadings," or total warheads and bomb loads, although the lead might not be quite as great as it would be under the simple Vladivostok formulas. In truth, however, such leads have no real military meaning and should have no political meaning, since each country has more than enough weapons to destroy the other in any conceivable nuclear conflict.

What counts is that the U.S. and the U.S.S.R. would, on reaching the comprehensive proposal's levels, be in rough equivalence. Each country would have a lead in the weapons it has concentrated on in the past. If technology were frozen (an important "if," which depends on the qualitative limits I shall discuss below), neither side would have a reliable counterforce capability, that is, the ability to destroy most of the other side's ICBM force. Both countries would have a secure deterrent, and the strategic balance would be stable, since there would be little incentive for initiating a nuclear first strike.

The Carter proposals dealing with qualitative advances in weapons were potentially more important than the restrictions on numbers of weapons. It has been clear for some years that the strategic-arms race is driven primarily by the continuous quest for technological improvement and advantage rather than by the mere desire to increase numbers. Yet neither the 1972 interim agreement nor the Vladivostok accords did anything to restrain such improvement; the replacement of existing weapons with new, improved models was either tacitly contemplated or in some cases explicitly allowed.

The comprehensive proposal would forbid the modernization or replacement of ICBM's and the development and deployment of mobile ICBM's. It would limit the flight-testing of ballistic missiles to six tests a year for ICBM's and six for SLBM's. It would ban strategic cruise missiles, defined as missiles



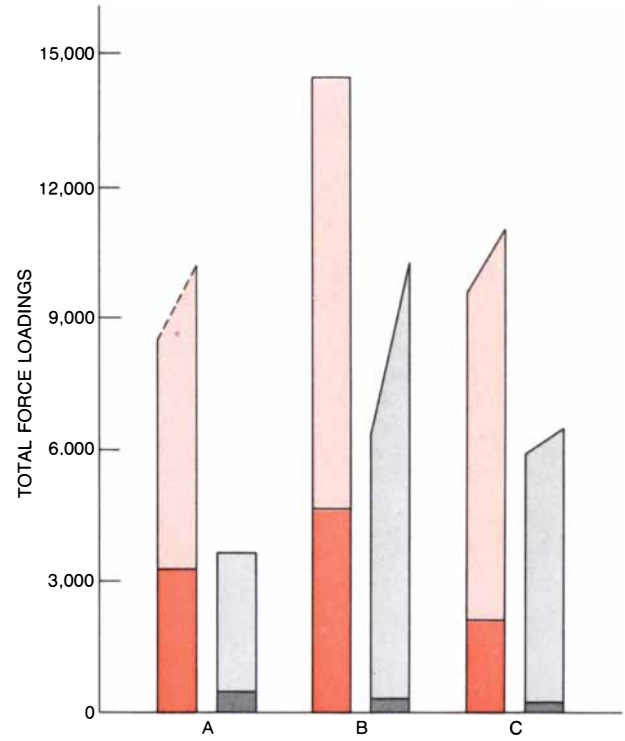


chart shows numbers of ICBM's, SLBM's and strategic bombers and indicates how many of the missiles might be MIRVed (*dark color and dark gray*). Slanted lines are used to indicate a range of possibilities. The center chart shows the total number of delivery vehicles. The right-hand chart shows total "force loadings": the sum of strategic-aircraft bomb loads (*dark tones*) and missile warheads (*light tones*).

with a range in excess of 2,500 kilometers (1,550 miles), but shorter-range cruise missiles would be unrestricted. Finally, it would require Russian assurances that the Backfire would not be given intercontinental capability or be deployed as a strategic weapon. The proposed bans on modernization and on mobile ICBM's and the restriction on testing would essentially prevent, or at least delay for a long time, any changes in existing missile forces. The stringent restrictions on ICBM's in particular would keep either nation from acquiring a counterforce capability that could threaten the other side's ICBM force. These provisions would markedly increase the stability of the strategic balance and would actually soon bring the arms race to a halt.

A six-test annual quota would effectively enforce the ban on modernization and replacement because the quotas would have to be used up largely by the "confidence" tests required to make sure weapons already in place are operational. Without additional tests it would become almost impossible to incorporate significant new improvements into existing missile systems; to procure totally new weapons would be out of the question. "National technical means of verification" could ensure that the quotas were honored. Ever since the first Russian ICBM test in 1957 the U.S. has been able to observe missile tests with a high degree of reliability from outside the U.S.S.R. The combination of satellite reconnaissance and electronic or infrared monitoring from the periphery of the U.S.S.R. has provided a wealth of information on both the launch and the reentry phases of the test firings. The ability to monitor could be improved and made more reliable by an agreement that all test firing be done on predesignated test ranges, following a precedent set when the anti-ballistic-missile treaty of 1972 was negotiated.

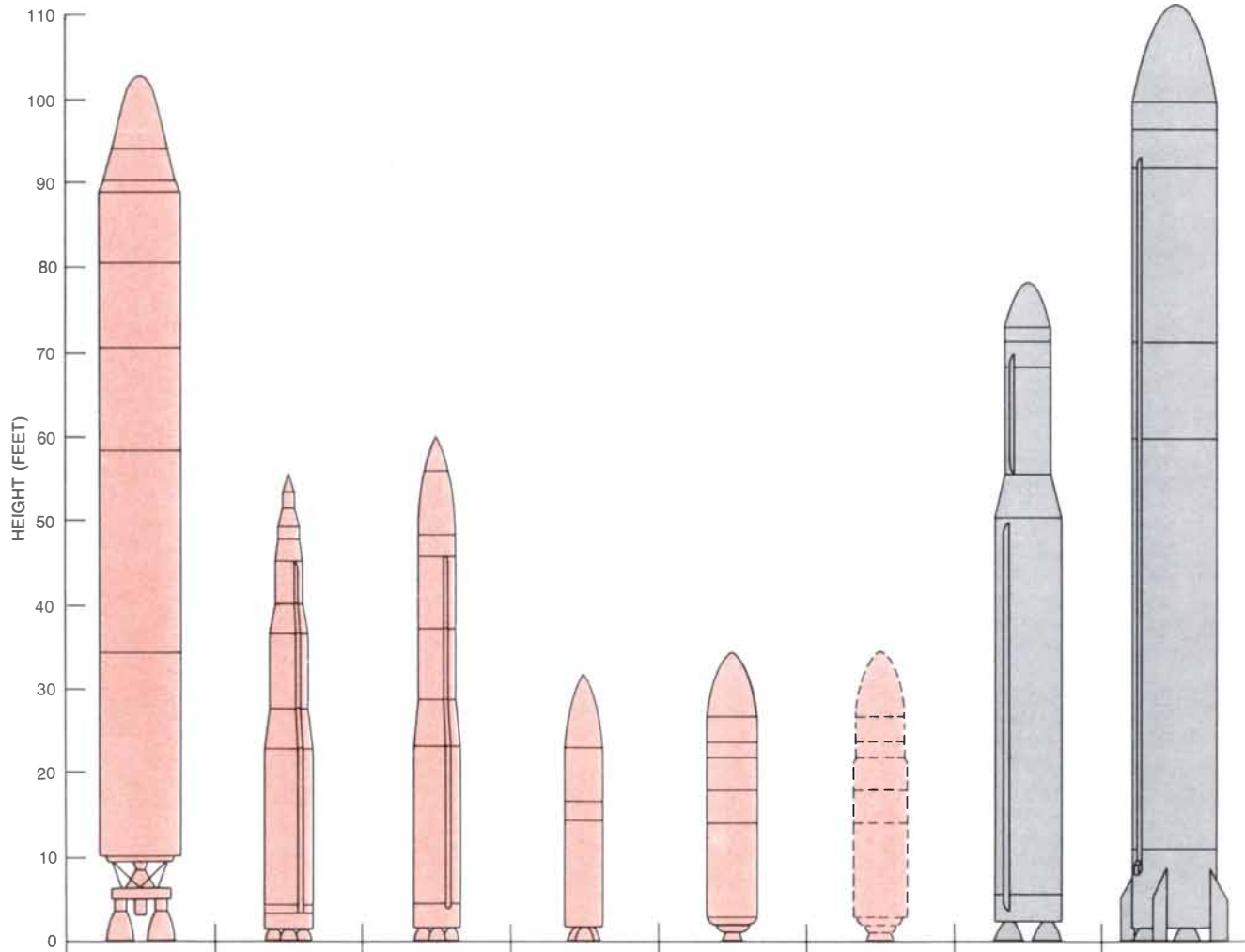
These qualitative proposals would effectively freeze the two nations at their technological levels as of whatever time the restrictions went into effect. Two American ICBM-development programs would be dramatically affected. The U.S. is now completing the development of the Mk. 12A warhead, with the NS-20 guidance system, to replace the MIRV's on the Minuteman III missile; deployment of the guidance system is scheduled to begin late this year. The new warhead will improve the accuracy of the Minuteman MIRV's and approximately double their explosive yield, giving each of them a high probability (more than 70 percent) of destroying a hardened missile silo. The Minuteman would thus for the first time attain what the Russians could view as a significant counterforce capability. The U.S. has also begun to design a totally new ICBM, the MX missile, as a replacement for the Minuteman. The MX is to have higher-yield warheads that can be terminally guided (steered during the later stages of their trajectory) to zero in on a target. This would be a true counterforce weapon and therefore an extremely destabilizing one. The MX is also to be mobile and hence extremely difficult to knock out; it will be able either to move back and forth in a long, hardened trench or to fire from any one of a number of prepared, hardened launch pads. This will increase its survivability but at the same time will make any ceiling on missiles or launchers virtually unverifiable. If the Russians were to emulate this approach, limits on ICBM's would be forever nonnegotiable. Both of these destabilizing U.S. programs would be halted by acceptance of the U.S. comprehensive proposal.

The U.S.S.R. would be similarly constrained. The new Russian ICBM's, the SS-17, the SS-18, the SS-19 and most likely the SS-16, have probably been tested enough to be deployed in their present configuration, but MIRVed versions of them are not now considered to have the combination of yield and accuracy that can give them a true counterforce capability. Improvements would be foreclosed by the U.S. proposals, thereby preventing the Russians from acquiring a counterforce threat to the U.S. Minuteman force. The ban on mobile ICBM's would keep the Russians from making the SS-16 mobile or making the already mobile intermedi-

ate-range SS-20 into an intercontinental missile, two possible developments that have caused great concern in Washington. In short, in the ICBM area the U.S. would be left with the higher-technology weapons, but that would be compensated for by the larger payload of existing Russian missiles. Hence there would be an approximate balance of overall capability.

In the SLBM area, however, things would be different. Here, as I have indicated, the U.S. is currently far ahead of the U.S.S.R. Unless some provisions were included allowing the Russians to catch up, or at least to complete the development of a satisfactory long-range MIRV system for their SLBM's, it is unlikely that they would accept a six-tests-per-year quota. If a test limit came into effect fairly soon, the American advantage would be somewhat counterbalanced by the fact that the quota would inhibit the deployment of the Trident I SLBM and prevent the development of the Trident II. The cancellation of the Trident II should indirectly bring the Trident submarine program to a halt, since such a vessel would be almost impossible to justify if it were limited to carrying the Trident I missiles that smaller submarines can launch. (A halt to the Trident submarine program would not be a security loss because smaller submarines would be better in any case.) Since SLBM's are basically deterrent weapons rather than counterforce ones, some arrangement for delaying the SLBM-test limit until the Trident I is developed and until the Russians have a MIRVed SLBM might not undercut arms control too severely, and it might even increase the stability of the deterrent balance.

An important element in the comprehensive proposal was its provision for dealing with the cruise missile, which has been under development in the U.S., but not in the U.S.S.R., since 1972. It is in effect a relatively small and cheap long-range World War II "buzz bomb" plus microelectronics, enabling it to strike a specified small target with remarkable accuracy [see "Cruise Missiles," by Kosta Tsipis; SCIENTIFIC AMERICAN, February]. The Russians have insisted that the Vladivostok accords implied the inclusion of cruise missiles with a range greater than 600 kilometers (375 miles) under the ceiling on strategic delivery vehicles. The U.S. comprehensive proposal, however, defined as strategic only cruise missiles

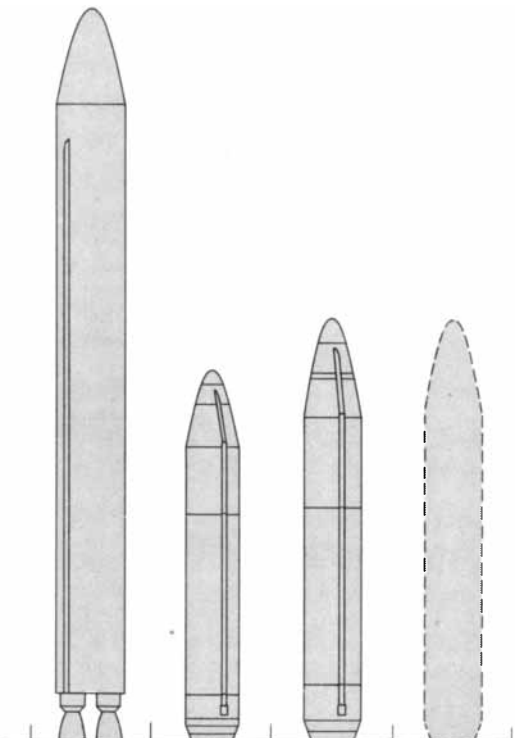


	TITAN	MINUTEMAN II	MINUTEMAN III	POLARIS	POSEIDON	TRIDENT I	SS-17	SS-18
NUMBER DEPLOYED	54	450	550	160	496	0	40	60
MIRV'S ON EACH	1	1	3	1	10-14	10-14	1-6	1-8
RANGE	INTERCONTINENTAL			2,500	2,500	4,000	INTERCONTINENTAL	

INTERCONTINENTAL BALLISTIC MISSILES of the U.S. (*color*) and the U.S.S.R. (*gray*) are depicted. For each missile the number known or estimated to be deployed, the number of MIRVed warheads and the range are given. For each country three ICBM's are shown, followed by three SLBM's. The U.S. has announced it will soon deploy the new Minuteman III Mk. 12A warhead with a high-

with a range greater than 2,500 kilometers (1,550 miles). That is more than enough to reach strategic targets in the U.S.S.R. west of the Urals from land bases in Europe, from submarines or other vessels or from aircraft flying beyond the Russian border. It was not surprising, therefore, that the Russians questioned the sincerity of the American offer to ban strategic cruise missiles so defined, particularly in view of the fact that the U.S. has itself been designating as strategic cruise missiles several weapons with ranges shorter than 2,500 kilometers.

Spokesmen for the Administration have maintained that cruise missiles with a range up to 2,500 kilometers are no different from other forward-based systems (the U.S. aircraft stationed in Europe) and like them should not be banned by a SALT treaty. The trouble with such an argument is that it reopens the entire question of the exclusion of forward-based systems from the treaty, an important concession the Russians made in the SALT I interim agreement and again at Vladivostok. Since the U.S. has about 500 forward-based aircraft and the U.S.S.R. has none (unless the Backfire is classified as being analogous), it is clearly to the American advantage not to include such aircraft as strategic delivery vehicles. It would be better to exclude the Backfire (unless it is given true intercontinental capability) from SALT and to include in SALT those cruise missiles whose range is not so short as to be demonstrably tactical, in return for continued Russian acquiescence to the exclusion of forward-based aircraft. The Backfire, tactical cruise missiles and forward-based aircraft could be dealt with in the context of the negotiations on "mutual balanced force reduction" that have been under way in Vienna for the past three years between the NATO and the Warsaw Pact countries.



SS-19	SS-N-6	SS-N-8	SS-NX-18
140	554	296	0
1-6	1	1	2-3
	1,300-1,600	4,200	4,200 (?)

er explosive yield and a more accurate guidance system. The long-range U.S. Trident I is scheduled to be deployed in about two years.

Having been summarily rejected by the Russians in Moscow, the comprehensive proposal was clearly too extensive, and involved too many new and sensitive concepts, to provide a basis for negotiating a SALT agreement before the October deadline. Since the U.S. had put forward the deferral, or Vladivostok, proposal as an alternative, it was logical to end up by combining the two approaches into the three-part framework I described above: a five-year treaty setting limits at the Vladivostok levels or a little lower, a three-year protocol somehow restraining cruise missiles, mobile missiles, heavy MIRVed ICBM's and the Backfire and a statement of principles for negotiating real cutbacks and qualitative restrictions in the future. Exactly how to constrain the cruise missile may be the most important question and the most sensitive issue between the two sides.

A possible approach to such an agreement is suggested by a proposal the Russians made during talks in Moscow in January, 1976. They proposed to ban sea-launched and land-launched cruise missiles with a range greater than 600 kilometers but to leave air-launched cruise missiles with a range of up to 2,500 kilometers unrestricted—provided that the aircraft able to launch those missiles were counted under the Vladivostok ceiling of 1,320 on MIRVed delivery vehicles. There are already 1,046 MIRVed Minuteman III and Poseidon missiles; each of the 10 or more programmed Trident submarines will carry 24 MIRVed missiles, for a total of at least 1,286 MIRVed delivery vehicles. That would leave only 34 U.S. aircraft that could be fitted to launch cruise missiles. Nevertheless, this could be the starting point for a cruise-missile agreement. The concept of counting aircraft equipped to launch cruise missiles as MIRVed delivery vehicles is sound because it inhibits unrestricted procurement of air-launched cruise missiles, makes the rather high ceilings on MIRVed delivery vehicles into a more significant arms-control measure and still allows some aircraft to serve as strategic delivery vehicles without having to penetrate Russian air defenses.

There would be problems of verification. The range of an operational cruise missile cannot be reliably determined by observing tests; it may also be difficult to ensure that a long-range missile announced as intended for air launching is not instead launched from the ground or from a ship, in violation of the shorter-range limit for such weapons. Actually, however, the Russians have never been much concerned about verification, since it is virtually impossible for the U.S. to keep a major weapons program secret. From the U.S. point of view, if the U.S.S.R. does begin to develop modern long-range cruise missiles, and if the missiles exceed the 600-kilometer range, they would still not be a significant strategic threat: even a 2,500-kilometer land-launched cruise missile could not reach the continental U.S. from the U.S.S.R. Putting such missiles on submarines would not be of much military value to the U.S.S.R. or present a security risk to the U.S. because the Russians already have 840 longer-range ballistic missiles against which the U.S. has no defense. Moreover, it would be a poor use of the Russian submarine force and hardly worth the violation of a treaty. If the Russians put air-launched cruise missiles on Backfires to give the bombers an intercontinental capability, those bombers would have to come under the MIRVed delivery-vehicle ceiling.

The Russians have indicated they will not accept the U.S. comprehensive proposal's deep cuts in the Vladivostok ceilings, but a more modest reduction, of perhaps 10 percent instead of 25 percent, might be possible. This would not have much significance for arms control, but it would be important as a signal to the world that the nuclear superpowers were prepared for the first time to slow the arms race. As for qualitative controls, it is probably too much to hope for restrictions on the modernization or replacement of existing weapons by October. Such provisions might be established as goals, however, in the statement of principles to guide future negotiations, which should proceed without delay after a SALT II treaty is signed. It is conceivable that both countries might declare they would independently exercise restraint in their weapons programs during further negotiations, thus establishing a positive climate in which to begin bringing the qualitative arms race under control.

BL Lacertae Objects

They are among the more remote and more luminous objects in the universe. The source of their radiation is a small nucleus (perhaps the size of the solar system) that may shine brighter than a galaxy

by Michael J. Disney and Philippe Véron

The most curious objects in the sky shine with a light quite unlike that of stars and the familiar aggregations of stars. A number of giant elliptical galaxies radiate prodigiously at radio wavelengths, where stars and normal galaxies are quiet. The spiral nebulas called Seyfert galaxies, some of which also are radio emitters, have core regions whose intense and spectrally distinctive radiation cannot be explained as starlight. Most extraordinary of all are the cosmic beacons called quasi-stellar objects, or quasars. The quasars may be extremely remote; light from some of them that reaches us today appears to have been in transit for a major part of the history of the universe. These great distances imply luminosities out of all proportion to those observed in our own galaxy and in nearby galaxies. In a quasar radiation many times more powerful than that of an entire galaxy seems to emanate from a compact core region that may be roughly the size of the solar system.

In the past few years still another class of objects has been added to this catalogue of astronomical prodigies: the BL Lacertae objects. The first of them was discovered less than 10 years ago, and about 30 more are known today. Like the quasars they are apparently distant and unaccountably bright, but they have spectral peculiarities all their own. As in quasars and Seyfert galaxies, their intense radiation seems to come from a comparatively small volume.

At first the discovery of the BL Lacertae objects can only have added to the prevailing confusion over the nature of the many new luminous sources, which for a time seemed to be proliferating without bound. Now it has begun to appear that there may be a continuity of properties among all these bizarre objects. We still do not know the ultimate source of their energy or how the energy is converted into radio-frequency and visible radiation. On the other hand, we can begin to classify the profusion and variety of objects in a common scheme. They may well be manifestations of a single process operating in different environments.

All the BL Lacertae objects have been recognized through their radio-frequency emissions. The first of them came to attention when J. M. Macleod and B. H. Andrew of the Canadian National Radio Observatory directed a high-frequency radio telescope toward a known radio source designated VRO 42.22.01. That source was of interest to them because observations at lower radio frequencies (a few hundred megahertz) had shown that its radio spectrum has an unusual shape. In most radio sources the intensity of radiation declines at higher frequencies (or equivalently at shorter wavelengths). The radio spectrum of a quasar is often nearly flat: intensity is almost constant at all radio wavelengths. In VRO 42.22.01 the usual relation was seen to be inverted: the signal strength increased at higher frequencies. A simple extrapolation of this trend suggested that at even higher frequencies VRO 42.22.01 might be a quite strong source. Macleod and Andrew set out to test this hypothesis at the frequency of their instrument, 10,600 megahertz (equivalent to a wavelength of 2.8 centimeters). At that frequency VRO 42.22.01 turned out to be one of the 30 or so brightest objects in the sky.

VRO 42.22.01 had not then been identified with any object discernible at optical wavelengths, in part because it happens to lie in the same direction as the Milky Way, which complicates the search for an optical counterpart. In that part of the sky stars are abundant in the foreground, and together with dense clouds of dust and gas they tend to confuse our view of the more distant universe. Nevertheless, Macleod and Andrew refined their radio-frequency observations in order to obtain a precise fix on the position of the source. When that position was checked in sky photographs, it was found to coincide with the image of a star of the 14th magnitude.

By the standards of the casual observer a 14th-magnitude star is hardly a conspicuous object; it is far too faint, for example, to be visible to the unaided eye. (Increasing optical magnitude corresponds to decreasing brightness. The scale is a logarithmic one, calibrated so that a change of five magnitudes represents a hundredfold change in apparent luminosity.) Macleod and Andrew, however, were surprised to find an object this bright. All known point sources of radio emissions were active galaxies or quasars; among those objects only one quasar was brighter than the 14th magnitude. Moreover, the chance location of VRO 42.22.01 in the Milky Way meant that much of its light must be absorbed by the "smoke" of our own galaxy, implying that its true luminosity is even higher.

Several clues to the nature of VRO 42.22.01 were available even in the initial observations. At radio wavelengths the source remained unresolved, or in other words pointlike, suggesting that it is either small or distant or both. Perhaps the most important clue was that the radio emissions are polarized. The intrinsic polarization of a source indicates that the radiation is not generated by a thermal mechanism, such as radiation from the heated surface of a star. Moreover, the interaction of polarized radiation with the interstellar medium of our galaxy, which rotates the plane of polarization by different amounts at various frequencies, provides a crude measure of distance: from the rotation it is possible to estimate what fraction of the galaxy the radiation has crossed. Observation of VRO 42.22.01 and its optical counterpart suggested that the source is very distant and probably outside our galaxy. Finally, that conclusion was also supported by the optical appearance of the object. In deep, long-exposure photographs a faint fringe surrounds the object's core. An obvious hypothesis is that the fringe consists of stars and that VRO 42.22.01 is a distant galaxy. From this information and from the inverted radio spectrum Macleod and Andrew proposed tentatively that VRO 42.22.01 might belong to a particular class of quasars and active galaxies. One of the more distinc-

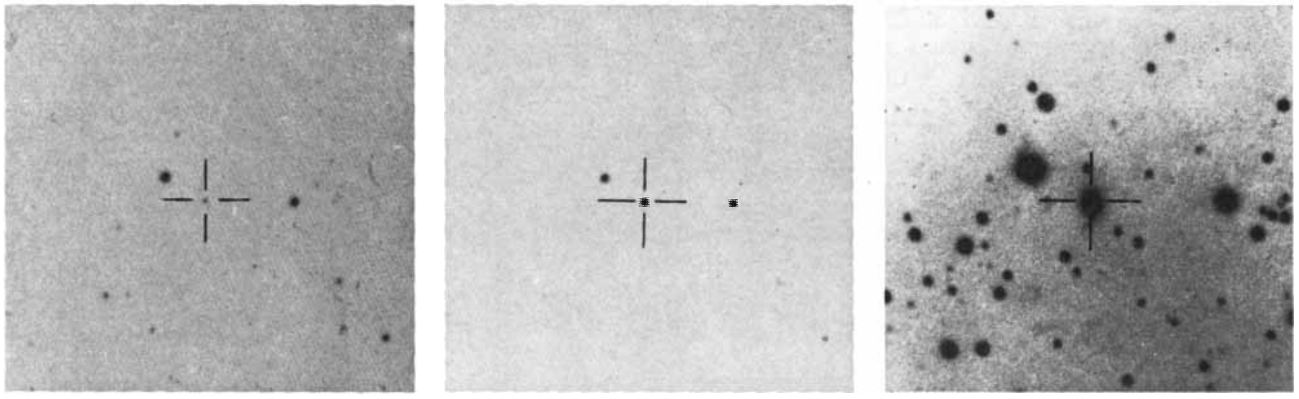
tive characteristics of the objects in that class is that their emissions are variable at both optical and radio frequencies.

Macleod and Andrew's suggestion was followed up by J. L. Schmitt of the David Dunlop Observatory of the University of Toronto, who undertook a search of the records of known variable stars. He discovered that Macleod and Andrew's object had already been identified as a variable star in the constellation Lacerta (the Lizard); it was designated BL Lacertae, often abbreviated as BL Lac. The variability had been discovered in 1929 by the German astronomer Cuno Hoffmeister, who showed that the star brightened and faded at irregular intervals. A 100 percent change in luminosity was observed in a week, and over the course of months the brightness could change by a factor of 15. At its brightest BL Lacertae outshone any quasar.

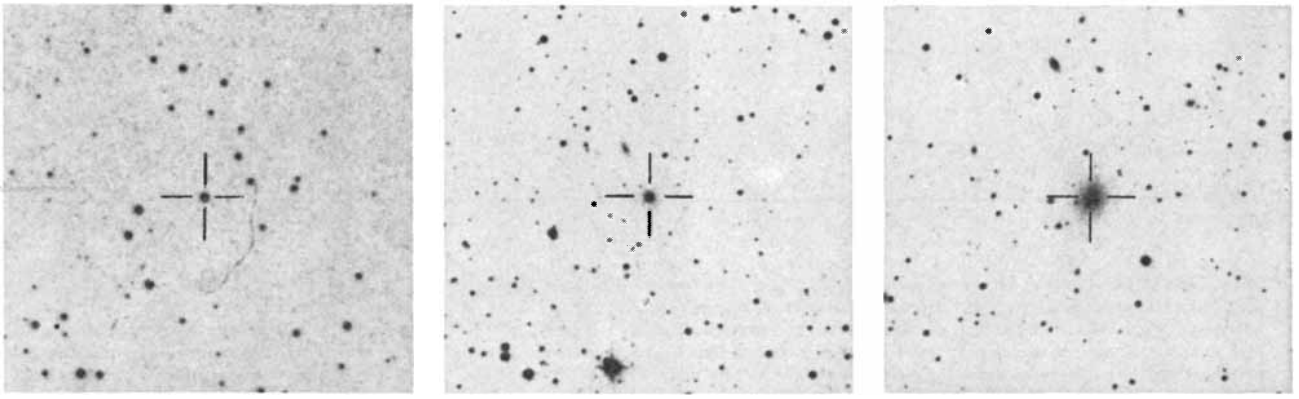
The most interesting and most difficult observational questions in astronomy often have to do with how far away an object is, how big it is and how much energy it radiates. Since we cannot measure these quantities directly the questions must be answered from the only source of information available: the feeble electromagnetic radiation that reaches the earth. Astronomy is the art of extracting information from these signals. The structural information embodied in a photographic image is only the beginning. Other information can be obtained, for example, from the spectrum, or distribution of radiant energy as a function of wavelength, from the state of polarization and from variations in luminosity.

The optical spectrum of BL Lacertae was a great surprise; in fact, it was mainly the peculiarities in the spectrum that led to the establishment of a new category of astronomical objects. The spectrum of a normal galaxy—one whose luminosity is simply the combined shining of many stars—is distinguished by a series of dark absorption lines. These are produced when the thermal radiation of the stars, which has a continuous spectrum, passes through the stellar atmosphere of comparatively cool gas. The atoms of a gas can absorb only specific wavelengths of radiation, namely those wavelengths whose energy corresponds to the difference in energy between quantum-mechanical states of the atoms. Absorption lines generated in a similar way can be detected in the spectra of some quasars, although the lines are faint. Much more prominent in quasar spectra are rather broad emission lines, which are brighter than the background rather than darker. The emission lines are thought to be produced when a cloud of gas is heated by an intense source of ultraviolet radiation. They are a defining characteristic of the quasars. Absorption lines and emission lines in a spectrum are also the mileposts employed in estimating the distances of galaxies and quasars.

When the optical spectrum of BL Lacertae was examined, neither absorption lines nor emission lines could be found. The spectrum was quite featureless, and hence inscrutable. In the absence of lines marking the known atomic transition frequencies the spectrum offered no clue to the distance of the object. Con-



BL LACERTAE, listed for almost 50 years as a variable star, is now recognized as the prototype of a new class of astronomical objects. The variability is evident in the two photographs at the left, made at the Wise Observatory of Tel-Aviv University. The photograph at the extreme left was made on September 25, 1973; in the center photograph, made two months later on November 17, BL Lacertae (marked by the reticle) is about four times as bright. In both of these pictures BL Lacertae appears to be completely starlike, but with longer exposures a fuzzy halo appears surrounding a bright core. In the photograph at the right, for example, made by G. Wlérick and D. Michet of the Haute-Provence Observatory in France, an indistinct halo can be detected. Its shape suggests the form of a giant elliptical galaxy.



THREE OBJECTS of varied appearance are all classified as objects of the BL Lacertae type. OJ287 (*left*) forms a completely stellar image, with no discernible nebula of stars. AP Librae (*center*) has both a distinct core and a galactic halo and closely resembles BL Lacertae itself. Markarian 501 (*right*) has a prominent nebula and a core so faint that on a long-exposure plate such as this one the starlike core region is obscured. All three of the photographs were made as part of the National Geographic Society–Palomar Observatory Sky Survey.

ceivably BL Lacertae might not lie at intergalactic distance at all but might be a star within our own galaxy.

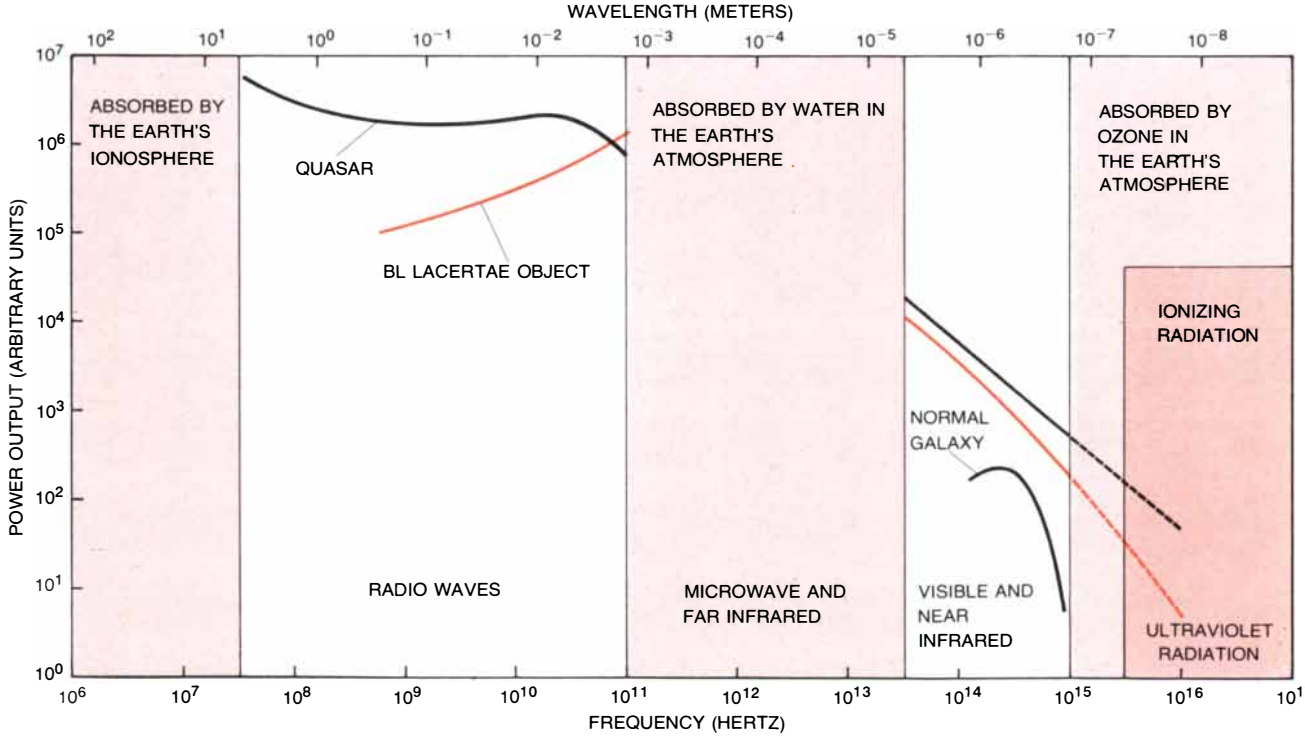
Another remarkable result was obtained when the variability of BL Lacertae was monitored with a photoelectric detector. The measurement has a strong bearing on the size of the emitting object. The shortest period in which a star or a galaxy can change its output of radiation is a measure of the maximum size of the object, or of the size of the emitting region that is subject to variation. The reason is that an overall change in brightness can be achieved only by propagating a signal of some kind throughout the region. The signal cannot travel faster than light and so the variable region cannot have dimensions larger than the distance light travels during the shortest period of variation. BL Lacertae was found to vary in brightness by as much as 30 percent during a single night. This rate implies that most of the light is emitted by a region with a diameter of no more than a few light-days. In size the object is closer in scale to a star than to a galaxy, which spans tens of thousands of light-years.

These properties of BL Lacertae amounted to a paradox. The radio emission and its polarization argued that the object must be an extragalactic one, such as a radio galaxy or a quasar. So did the halo of fuzz visible in some photographs, although that could be detected only with difficulty. On the other hand, the rapid variations in brightness suggested a small source, such as a star within our own galaxy. The location of the source within the belt of the Milky Way enhanced the plausibility of that notion. Neither of these hypotheses, however, could explain the featureless spectrum of BL Lacertae. The shape of the spectrum—or in other words the color of the object—was also difficult to account for. The object is too red for a quasar, even when allowance is made for the preferential scattering of blue light by galactic smoke, but neither is the spectrum a stellar one.

BL Lacertae could not be made to fit any of the existing astronomical categories, and so it became the prototype of a new category. The 30 or so similar objects that have since been identified were found essentially the same way: by searching for bright point sources of high-frequency radio energy. They had all escaped notice because most earlier radio astronomy was done at lower frequencies. The rather clumsy term "BL Lacertae objects" has been applied to all the members of the class, and we shall abbreviate it simply as "lacertae." Of course the objects generally have no connection with the constellation Lacerta.

The second of the lacertae to be found was a radio source designated OJ287; like many others it was identified in a sky survey made by the Ohio State University Radio Observatory at a frequency of 1,415 megahertz. OJ287 appears to be completely stellar; no surrounding nebula can be detected even in the most deeply exposed photographs. Indeed, it has turned out that only a minority of the lacertae have discernible halos, so that some of the objects are difficult to identify on a photographic plate. The identification of OJ287, however, was unambiguous; it has all the properties that make BL Lacertae so interesting and so conspicuous and that have become the defining characteristics of the class: an inverted radio spectrum, variability at all wavelengths, strong optical polarization and a featureless optical spectrum.

No understanding of the BL Lacertae objects is possible without an approximate idea of how far away they are. If they are starlike objects within our own galaxy, then their intrinsic lu-

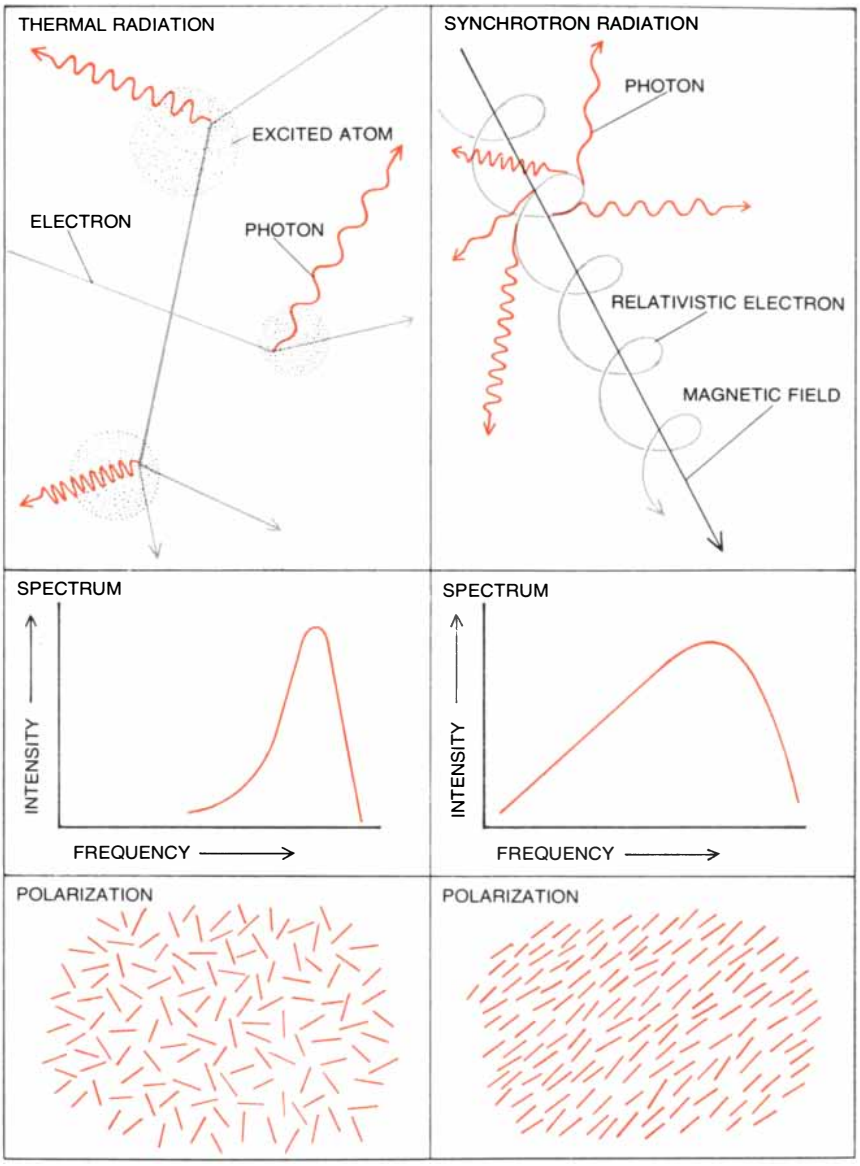


SPECTRUM OF RADIATION reveals large differences between the total energy emitted by a normal galaxy and that of a quasi-stellar object or a BL Lacertae object. Almost all the radiation of a normal galaxy is at optical wavelengths, and the spectrum has the characteristic shape of thermal radiation, such as starlight. The radiation of quasars and BL Lacertae objects covers a much broader range of wavelengths. In addition to intense optical emission both radiate strongly at radio wavelengths, the quasar spectrum being almost flat in this region and the BL Lacertae spectrum being "inverted" and sloping in a direction opposite to the usual one. Both quasars and BL Lacertae objects might be most powerful in the microwave and infrared parts of the spectrum, but those wavelengths are hidden from view by absorption in the earth's atmosphere. Terrestrial absorption also obscures ultraviolet emissions, including short-wavelength radiation capable of ionizing hydrogen atoms. Clouds of such ionized gas create emission lines in the spectra of quasars, but the lines are absent in BL Lacertae spectra. Extrapolation from the accessible part of the spectra suggests that lacertae may emit comparatively little ionizing radiation; alternatively, the lines may be missing because the elliptical galaxies associated with the lacertae are deficient in hydrogen gas.

minosity must be high, but by no means out of the ordinary. They would radiate perhaps 10 times as much energy as the sun. If they are at cosmological distances, so that their light must cross a substantial fraction of the visible universe to reach us, the quantity of energy being dissipated is of an entirely different order of magnitude. The lacertae would be among the most powerful sources of radiation in the universe, emitting perhaps 100 times as much energy as our entire galaxy. (The galaxy itself radiates about 10^{10} times as much energy as the sun.)

The debate over the distances of the lacertae, now settled in favor of the cosmological scale, has followed the pattern of a similar debate over the distances of quasars. The issues are much the same. All estimates of cosmological distance rest on the conjecture that the universe is uniformly expanding. We observe this expansion in the motion of distant galaxies, which all appear to be receding from us and from one another. As Edwin P. Hubble and Milton L. Humason demonstrated almost 50 years ago, the rate of recession seems to be proportional to the distance of the galaxies. Because of this motion, features in the spectrum of a galaxy, such as absorption or emission lines, are shifted to longer wavelengths, and thus wavelengths in the visible range are shifted toward the red end of the spectrum. From the red shift the velocity of recession can be determined and from the velocity the distance can be inferred.

Among the most distinctive properties of the quasars are their very large red shifts. If the red shifts are attributed to cosmological expansion, then some of the quasars are retreating from us at well over half the speed of light and they must lie at immense distances. That we can see them at all implies that they are hundreds of times more luminous than even the largest normal galaxies. Moreover, that luminosity must come from a rather small volume. Some quasars have been observed to vary in brightness on a time scale of a few days, indicating that the luminous region is no more than a few light-days in diameter. One hundred times the power of a galaxy must be extracted from a region with one millionth the diameter.



RADIATION IS EMITTED when an electron or other electric charge is accelerated; the nature of the acceleration determines the wavelength and other properties of the radiation. Thermal radiation (*left*) results from collisions and excitations of atoms and electrons within a hot, dense medium, such as the atmosphere of a star. The events that give rise to the radiation are random, but they obey statistical laws that give the spectrum a well-defined form and, for any given temperature, a comparatively narrow band of wavelengths. The polarization of the waves, or in other words their orientation if they could be seen head on, remains random. The radiation of quasars and BL Lacertae objects is thought to come from another process, the synchrotron mechanism (*right*). Synchrotron radiation is emitted when electrons are given the continuous accelerations entailed in following a circular orbit; in astronomical objects the orbits are actually helical, wound around the flux lines of a magnetic field. Electromagnetic waves are emitted tangent to the electron orbits with a broad range of frequencies. The synchrotron radiation is strongly polarized, the waves being lined up parallel to the plane of circular motion.

As might be expected, some astronomers have found these ideas hard to believe and have proposed alternatives. In these "local" quasar hypotheses the problem of the enormous energy densities is evaded by assuming that the quasars are much nearer and hence less luminous. In that case, however, the large red shifts of the quasars go unexplained.

This debate has continued inconclusively for a dozen years. The crucial evidence required to prove the cosmological hypothesis would be a demonstration that at least one quasar is embedded in a distant galaxy and that the surrounding stars have the same high red shift as the quasar itself. No quasar, however, has yet been observed with a definite nimbus of stars, nor is the failure to see one in any way surprising. The cosmological hypothesis presupposes such high luminosity for the core that it must overwhelm the contribution of the surrounding stars.

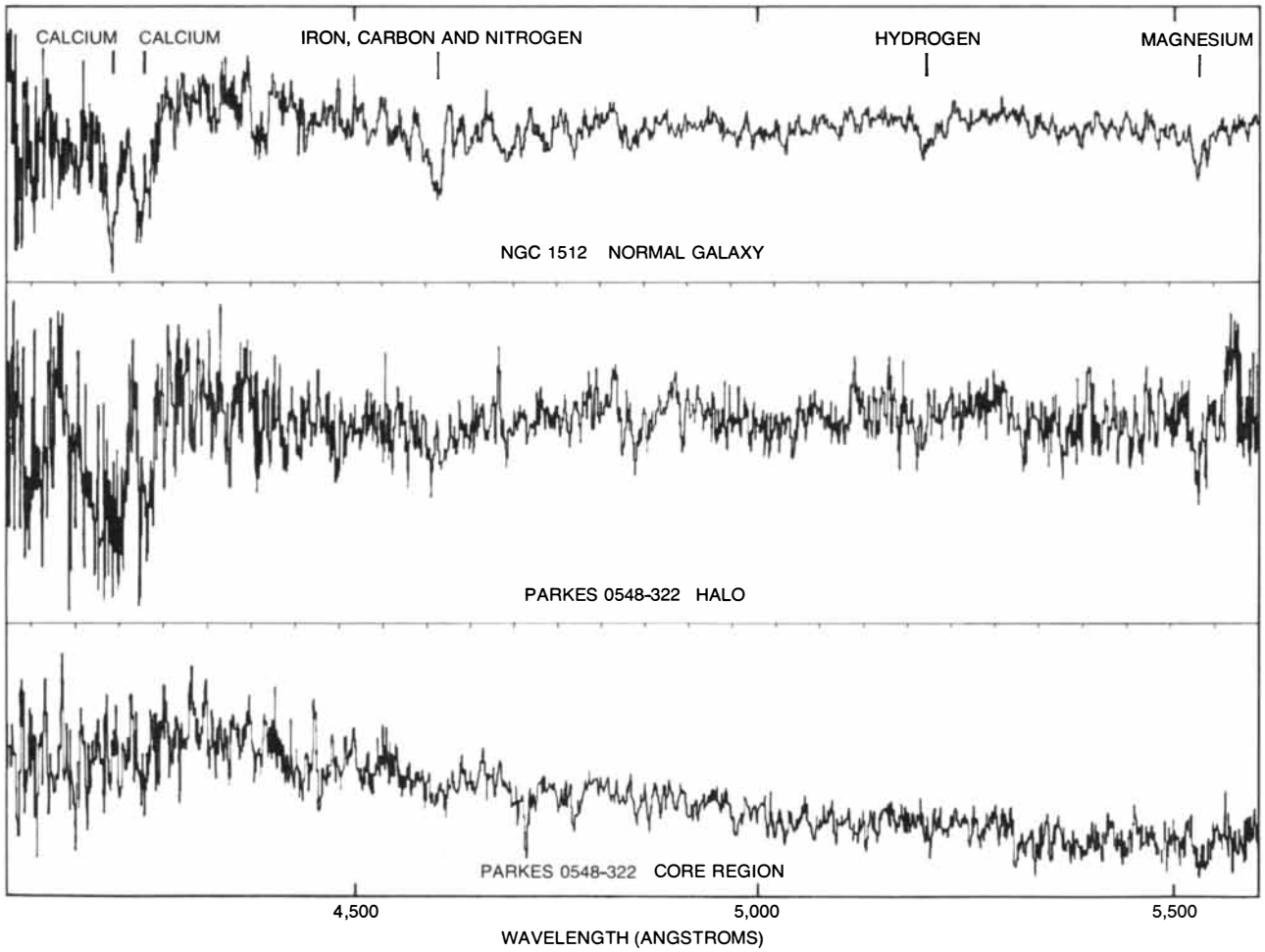
The need for a corroborative red-shift measurement lends obvious importance to the fuzzy borders seen around some of the BL Lacertae objects. If the outlying fuzz consists of stars, then the red shift of the entire object might be determined from the lines in their spectra. Ironically this method offers the only hope of gauging the distances of the lacertae, since the spectrum of the luminous core has no landmark emission or absorption lines. The challenge to the observer lies in recording the spectrum of a faint nebula in the immediate vicinity of a bright starlike core. The task de-

mands a large investment of telescope time, ideal atmospheric conditions and the good fortune to find the luminous core at minimum intensity.

In 1974 two astronomers at the Paris Observatory showed that the nebular region surrounding BL Lacertae has the apparent size and the brightness distribution of a giant elliptical galaxy at a distance corresponding to a recessional velocity of 21,000 kilometers per second, or about 7 percent of the speed of light. Soon afterward a group of astronomers from the Hale Observatories in California reported measuring a red shift for this nebular region. They obtained their result by mechanically occulting the core, allowing only light from the halo to enter the spectrograph. From their data they calculated a red shift precisely equal to the one proposed by the French astronomers. The evidence for this red shift, however, was not convincing: it consisted of no more than a single small dip in the recorded spectrum. What is more disturbing, another group of astronomers at the Lick Observatory failed in an attempt to reproduce the observation. The validity of the result remains in doubt, but investigations are proceeding.

Meanwhile astronomers at the Mount Stromlo Observatory in Australia worked to record a line spectrum from AP Librae, an object in the Southern Hemisphere that is remarkably like BL Lacertae. They did not block off the light from the galactic core but made a number of spectra over three years, hoping to catch the core at minimum luminosity. No single spectrum alone provided unambiguous evidence, but taken together the spectra were convincing. Several weak absorption lines commonly seen in the spectra of galaxies were present. The red shift of these lines was equivalent to a recessional velocity of 15,000 kilometers per second and a distance of about a billion light-years. AP Librae is thus definitely an extragalactic object. It seems to consist of a small region of exceptional luminosity lying at the nucleus of a large elliptical galaxy.

Red shifts have since been determined with reasonable confidence for four more lacertae. All seem to be embedded in galaxies at a great distance. Red shifts of from 15,000 to 20,000 kilometers per second do not compare, of course, with the extreme red shifts of the most distant quasars. It should not be concluded, however, that quasars are generally the more distant objects. Instead the difference in measured red shift may result from observational biases: because the spectra of lacertae are so faint it is likely that only the nearest ones have been recorded. Even so, absorption lines tentatively identified in the spectra of two lacertae suggest they are moving at about half the speed of light. One object, designated AO 0235 + 164, has two sets of lines with quite different red shifts,



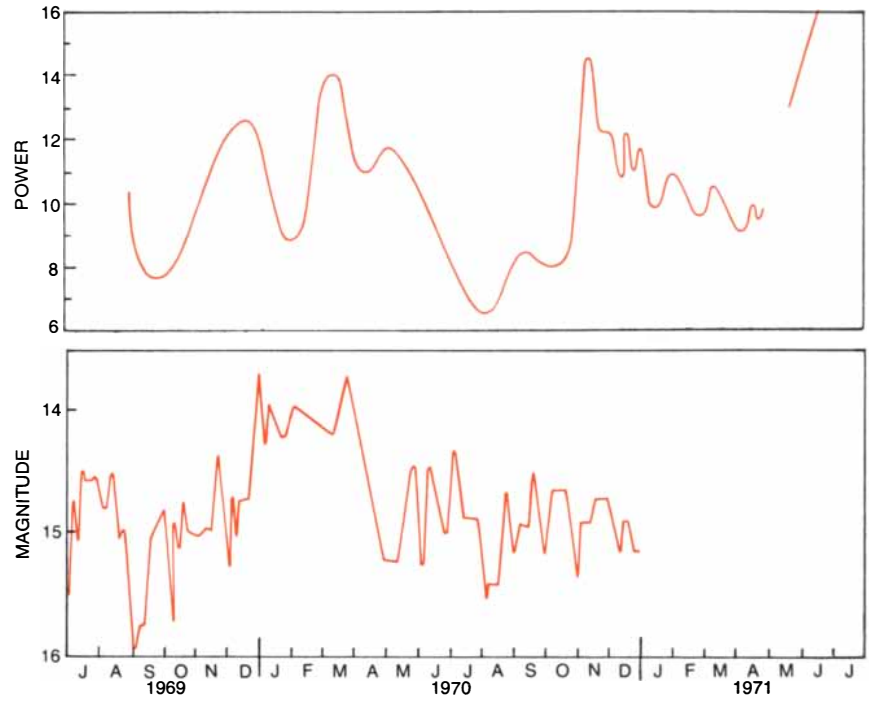
ABSORPTION LINES in the spectrum of the fuzzy halo surrounding a BL Lacertae object support the supposition that it is an independent galaxy at a great distance from our own. The BL Lacertae object is one designated Parkes 0548–322; here its spectrum, measured with the Anglo-Australian Telescope at Siding Spring Observatory in Australia, is compared with that of a normal galaxy, NGC 1512. The absorption lines are impressed on the spectrum by atoms in the comparatively cool outer envelopes of stars, and their presence in the spectrum can be regarded as evidence that the halo consists of stars. The absence of these lines in a spectrum recorded from the core of the object is consistent with the hypothesis that light is emitted in that region by a nonthermal source. For comparison the absorption lines of the BL Lacertae object are shown in coincidence with those of NGC 1512; actually they are shifted to longer wavelengths by the general expansion of the universe, which causes Parkes 0548–322 to recede from our galaxy at a speed of 21,000 kilometers per second.

equivalent to 52 percent and 85 percent of the speed of light. A dual spectrum of that kind might be produced by absorption in two intervening galaxies in the line of sight, which would require, of course, that the object itself be more distant than either galaxy. (Another but less probable interpretation is that the lines are produced by envelopes of gas blown off a galaxy at very high speed.) At the distance implied by the red shift the galaxy itself would be invisible and only the starlike core would be detected.

If BL Lacertae objects are at cosmological distances, what powers their dazzling luminosity? The ultimate source of the energy is quite unknown, although in recent years the sentiment has become increasingly widespread that it must be gravitational. In addition there is a known mechanism that might account for the final step in producing much of the radiation. That mechanism is the synchrotron process, an interaction of high-energy electrons with a magnetic field.

Electromagnetic radiation is emitted whenever an electron is accelerated, that is, forced to change its speed or its direction. Synchrotron radiation is emitted when an electron moving at nearly the speed of light undergoes the continuous acceleration entailed in following a circular path. Under terrestrial conditions that can be accomplished when electrons are accelerated to relativistic speeds (those approaching the speed of light) in a circular particle accelerator such as a synchrotron; hence the name of the radiation. In an astronomical context, synchrotron radiation is produced when a population of relativistic electrons encounters a magnetic field. The electrons are diverted by the field into helical trajectories oriented along lines of magnetic flux. The radiation is emitted tangent to these trajectories with a frequency distribution determined by the strength of the field and the energy of the electrons.

Synchrotron radiation is readily distinguished from the thermal radiation of ordinary stars. The spectral distribution of synchrotron radiation is far broader; that is why astronomical objects emitting by the synchrotron mechanism can be bright at both radio and optical wavelengths, whereas the thermal radiation of stars is largely confined to a single band of wavelengths. An even more striking distinction of synchrotron radiation is its polarization: the electromagnetic waves tend to be aligned parallel to the plane of the emitting electron's circular motion. The thermal radiation of stars is not polarized, that is, the waves are randomly oriented. It is polarization that provides the most persuasive evidence for synchrotron emission in quasars and BL Lacertae objects. In the lacertae the polarization at visible wavelengths is typically 10 percent, and it can rise to 35 percent. It follows that the magnetic fields responsible for the synchrotron emissions must have a very orderly structure.



RECORD OF VARIATIONS in the intensity of radiation from BL Lacertae has a direct bearing on the size of the radiating core region. The radio variations (*top*) were monitored for about two years at a wavelength of 2.8 centimeters; the largest fluctuations in output have a time scale of a few weeks. At optical wavelengths (*bottom*) large changes in magnitude were observed over a period of a few days. No object can change its luminosity faster than the period required for a signal to traverse the object at the speed of light. It follows that the optical emissions of BL Lacertae must come from a region only a few light-days in diameter; the radio-emitting region could be larger. In comparison the solar system is half a light-day in diameter.

The property of the BL Lacertae objects that is the most difficult to understand is the small size of the region emitting the synchrotron radiation. The size is inferred from the rate at which the luminosity of the objects is seen to vary. The maximum sizes are already uncomfortably small for quasars, and the problem is even severer for lacertae.

A good deal more observation will be required to define the range of variability in the lacertae, but estimates are already possible. BL Lacertae itself has been the most carefully monitored for rapid changes. At optical wavelengths the recorded range of intensities covers about three magnitudes, or a factor of 15. The luminosity has changed by as much as 400 percent within two days, and a flickering of a few percent has been reported on a time scale of minutes. The polarization has also been shown to vary—in both degree and orientation—over the course of a few nights. At radio wavelengths the variation is less extreme: although the range of luminosities may reach 300 percent, the time scale is weeks rather than days, suggesting that the radio emissions come from a larger volume than the optical emissions. In BL Lacertae the optical and radio variations seem entirely independent; a correlation between the two parts of the spectrum has been discovered in only one of the lacertae, OJ287. In all other lacertae and at all wavelengths the variations seem to be erratic; no patterns or periods have been perceived. Some spectacular outbursts have been recorded. In 1975 the object AO 0235 + 164 brightened by a factor of 100; another lacertae, PKS 2155–152, has increased its optical brightness six hundredfold.

The rapid optical variation in the BL Lacertae objects suggests that the core region has a diameter of no more than a few light-days, and it could be substantially smaller. (The solar system, if defined by the orbit of Pluto, has a diameter of about half a light-day.) The variations in lacertae may be 10 times faster than those of quasars, and thus the luminous core of a BL Lacertae object may be only a tenth as large as that of a quasar. In volume measure this ratio must be cubed, and so the lacertae may radiate the same energy as a quasar from only a thousandth the volume.

Principles of theoretical physics argue that the radiation of so much energy from such a small volume by synchrotron emission cannot be sustained for long. The radiation would be quenched, it seems, by the effect called the inverse Compton catastrophe.

The Compton effect is named after

the American physicist Arthur Holly Compton, who first explained it in 1922. It describes the scattering of a photon, or quantum of electromagnetic energy, from an electron that has only a small kinetic energy. The photon gives up part of its energy to the electron, and the transaction is observed as a slight decrease in the frequency (or increase in the wavelength) of the radiation. Inverse Compton scattering is observed when a photon encounters a high-energy electron. Then it is the electron that loses energy to the photon. Moreover, the change in frequency is no longer a slight one: a radio-frequency photon can be converted by inverse Compton scattering into a high-energy X ray.

Ordinarily the probability that a photon will collide with a relativistic electron is quite low and inverse Compton scattering is a rare event. When electrons and photons are confined at high density, however, the probability of collisions increases. At the density required to explain the size and luminosity of BL Lacertae objects, all the photons emitted by the synchrotron process should be scattered by electrons and converted into X rays.

The inverse Compton effect can be described in an equivalent way by imagining the interaction from the point of view of an electron. An electron in the core of a BL Lacertae object is subject to two magnetic fields. One is the original, external field that confines the electrons to helical trajectories and gives rise to the synchrotron radiation. The other is the field of the radiation itself, the magnetic field that is a component of every electromagnetic wave. At low density the radiation field is negligible, but as the density of photons increases it becomes the stronger of the two fields. The electrons then respond mainly to the radiation field, and as a result the radio waves are suppressed and X rays are emitted instead. The electrons are rapidly slowed and soon cease to radiate.

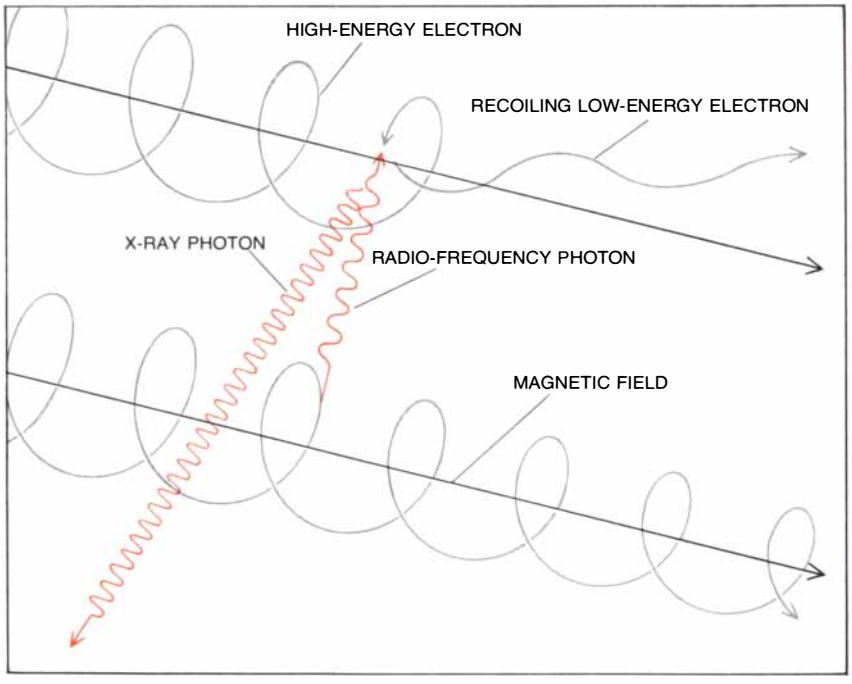
These considerations argue that BL Lacertae objects cannot be at very great distances; if they were, they would emit X rays, which are not observed, instead of radio waves, which are. Obviously this argument contradicts some of the observational evidence. A number of ad hoc measures to resolve the contradiction have been proposed, but none of them is convincing. It appears that we should not rely too heavily on theoretical arguments when the physics is not well understood.

With the discovery of numerous active galaxies, quasars and BL Lacertae objects it has been recognized that these categories blend smoothly into one another. The most variable quasars, for example, can alter their luminosity by 100 percent in a few days, a rate approaching that of the lacertae. The same quasars tend to have the comparatively red spectrum and the high optical polarization characteristic of lacertae. Similarly, some radio galaxies that have bright core regions look much like BL Lacertae objects but have emission lines in their spectra. The lines are of the same type but not of the same strength as the lines found in quasar spectra. Some nearby elliptical galaxies have no bright nucleus at visible wavelengths but do have a radio-wavelength nucleus; moreover, the visible radiation from these galaxies is sometimes slightly polarized, suggesting that a weak region of nonthermal optical radiation may underlie the more powerful radiation of normal stars. In one dramatic case a radio galaxy, 0521–36, with emission lines at optical wavelengths, brightened to become a BL Lacertae object with a featureless spectrum.

The existence of intermediate forms reinforces the conjecture that active galaxies, lacertae and quasars are all manifestations of a single process. All these objects may be galaxies with compact regions at their cores emitting intense, polarized radiation over a broad spectrum. The energy output of these small luminous regions varies over a wide range, both from galaxy to galaxy and from time to time in a single galaxy. The most powerful of the nuclei may be hundreds or even thousands of times brighter than the rest of the galaxy; that is why only the starlike core can be detected in quasars and in some lacertae, such as OJ287. When the nucleus and the parent galaxy are of roughly equal brightness, we see both, as in AP Librae and perhaps in BL Lacertae itself. When the nucleus is weak, its radiation is swamped by that of ordinary stars and the galaxy appears to be normal, although radio emissions may still betray the presence of an active core.

One possible objection to this unifying scheme is that all known lacertae are radio sources, whereas many quasars have no detectable radio emissions. In fact, the great majority of quasars may be radio quiet, perhaps as large a proportion as 95 percent. This distinction, however, may simply reflect an instrumental bias. The first quasars were detected by their unusual brightness at radio wavelengths, and so all of them were of course radio sources. Only later, when the sky was searched at optical wavelengths for bright blue and ultraviolet objects, were the radio-silent quasars discovered. It may turn out that radio-silent lacertae are also abundant, but because their optical spectra are less distinctive than those of quasars none have been found so far.

The most intriguing difference be-



INVERSE COMPTON SCATTERING is a process that would seem to prohibit regions of extraordinarily intense radio-frequency radiation such as those observed in BL Lacertae objects. The scattering takes place when a photon, or quantum of electromagnetic energy, collides with a high-energy electron. Much of the electron's energy is transferred to the photon, which is thereby converted to a much shorter wavelength, usually in the X-ray range. Under ordinary circumstances inverse Compton scattering is rare because collisions between photons and high-energy electrons are improbable. In the core of a BL Lacertae object, however, the density of both photons and high-energy electrons must be very high. Indeed, calculations based on the apparent size and luminosity of the core predict an "inverse Compton catastrophe," in which all the radiation appears as a burst of X rays that rapidly drains the electrons of their energy.

tween quasars and lacertae is that the quasars have strong emission lines in their spectra that the lacertae lack. The reason for this is not yet understood.

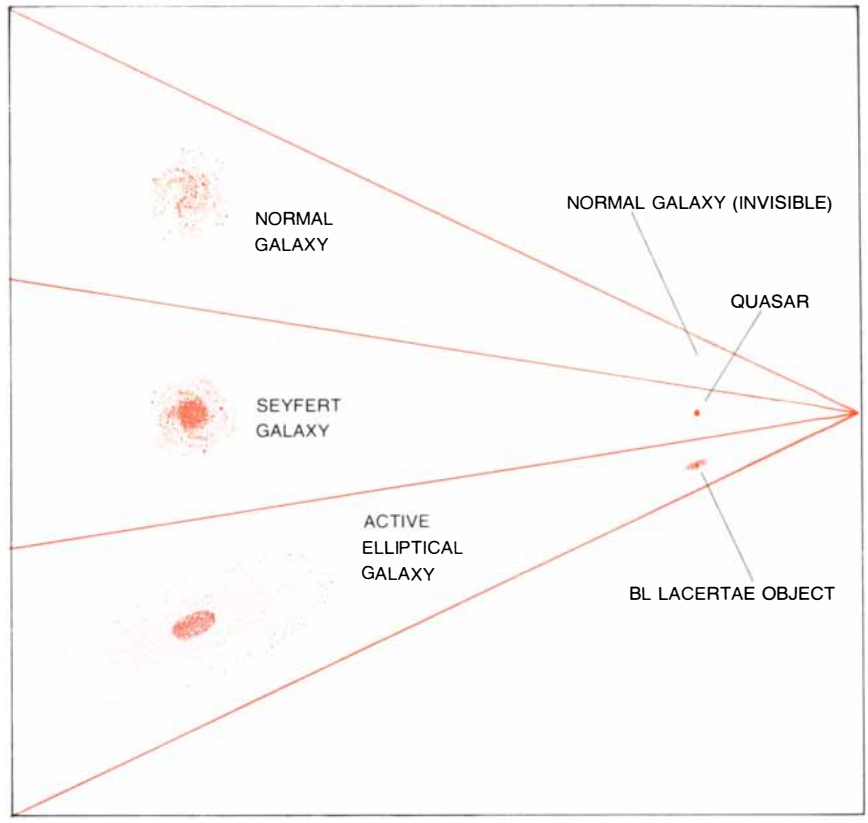
Emission lines are produced when a gas is heated to about 10,000 degrees Kelvin; at that temperature the electrons are stripped away from many atoms and the gas is said to be ionized. When an electron and an ion recombine, the electron falls through a series of quantum-mechanical states, giving up energy with each transition; the energy appears in the form of electromagnetic radiation with wavelengths characteristic of each atomic species. Since hydrogen is the commonest element, the recombination lines of hydrogen are generally the strongest.

In quasars emission lines are thought to be produced when gases are heated by ultraviolet radiation from the luminous core. Ultraviolet photons have enough energy to ionize hydrogen atoms (and other atoms as well), and they are strongly absorbed. The energy absorbed by the gas in this way could heat it and lead to a rapid expansion of the cloud. The expansion in turn might explain why the emission lines are rather broad: some parts of an expanding cloud are propelled toward the observer and some parts away, giving rise to opposite shifts in wavelength.

The absence of strong emission lines in BL Lacertae objects remains a puzzle, although at least two promising explanations have been considered. One possibility is that lacertae emit too little ultraviolet radiation to fully ionize a cloud of gas. Indeed, the spectra of most lacertae fall off rapidly in the visible band, so that it seems likely they are not as bright in the ultraviolet as quasars. On the other hand, a few lacertae, such as OJ287, have optical spectra shaped much like those of quasars, and yet they still lack emission lines. The issue might be settled by recording the ultraviolet spectrum of a BL Lacertae object, but the wavelengths in question are strongly absorbed by the earth's atmosphere.

A simpler explanation for the absence of emission lines is that the galaxies surrounding lacertae contain no gas to be ionized by the ultraviolet emissions of the core. The plausibility of this hypothesis is enhanced by the observation that when a galaxy can be resolved surrounding a BL Lacertae object, it is always an elliptical galaxy. Studies of nearby elliptical galaxies have shown that as a group they are deficient in gas. Active nuclei embedded in spiral galaxies, on the other hand, always have strong emission lines like those characteristic of quasars. In spiral galaxies, such as our own, gas is abundant.

In these observations we can perceive the outline of a synthesis relating all these bizarre objects in the sky. Quasars, in this view, are simply active nuclei in spiral galaxies that have been removed to a great distance, where their galactic nebulosity is completely obscured. At closer range the same object would be called a Seyfert galaxy. When the same kind of active core is observed in an elliptical galaxy at a great distance, it may be identified as a BL Lacertae object.



ACTIVE-CORE HYPOTHESIS provides a single possible explanation for certain unusual galaxies that are relatively nearby as well as for quasars and BL Lacertae objects. All these objects have in common the emission of intense radiation with a nonthermal spectrum; moreover, the variability of the radiation suggests that in many cases it must come from a quite small core region. Spiral galaxies that have such cores are called Seyfert galaxies; if they were removed to a great distance, they would look much like quasars, since only the bright core and not the surrounding stars would be visible. A normal galaxy at the same distance would be too faint to be seen. Active cores are also found in some elliptical galaxies and, when viewed from a great distance, they could have the appearance of BL Lacertae objects. Faint halos of stars surrounding the core would be visible only in the lacertae and the quasars closest to our galaxy.

It is an appealing synthesis, but there are objections to it. For example, some elliptical galaxies have strong emission lines and some quasars have a two-lobed radio structure of a kind that has never been found in a spiral galaxy. Both of these observations tend to associate quasars with elliptical galaxies rather than spiral ones.

In the end the foremost mystery remains: What is the source of energy for the active galaxies and the more distant objects? Since no mechanism known to terrestrial physics will do, attention has focused on exotic and speculative proposals. At one time a leading candidate was the superstar, an enormous ball of hot gas with a mass millions of times that of the sun. More recently popular allegiance has shifted to the black hole, the compact object that is the ultimate residue of matter after gravitational collapse. Many other ideas have been discussed: "white holes," lagging cores, giant pulsars, colliding star swarms, collapsing galaxies, multiple supernovas.

A successful model must account for the operation of what may be the most powerful engine in the universe, and it must fit that engine into a compartment of trifling size. A crucial test of all models is the fastest variation in luminosity that can be accommodated, since that period corresponds to the time required for a signal to traverse the emitting region. Some models have had to be discarded already because they do not allow rapid enough variations. The black-hole model can account for the highest luminosity and the fastest variations observed so far in lacertae only by postulating an object with a diameter of three million kilometers (about half again the size of the sun). An object of that size cannot fluctuate significantly in less than about 10 seconds. If faster variations were observed in a BL Lacertae object, even the black-hole model would have to be discarded.

Light-Wave Communications

The first commercial test of light-wave telephone service is under way in Chicago. The signals are dispatched over glass fibers in pulses generated by tiny solid-state light sources

by W. S. Boyle

Three months ago the Bell System began the commercial evaluation of a light-wave communication system in which messages are coded into pulses of light transmitted through hair-thin glass fibers. The new system carries voice, data and video signals over one and a half miles of underground cable interconnecting two switching offices of the Illinois Bell Telephone Company and a large commercial building in Chicago's business center. The light-guide cable, only half an inch in diameter, contains 24 fibers in two ribbons of 12 fibers each. The information capacity of each fiber is 44.7 megabits per second, meaning that the light source feeding into the fiber is turned on and off 44.7 million times per second. At this pulse rate a single fiber can carry 672 one-way voice signals; thus the 24 fibers have a capacity of 12×672, or 8,064, two-way conversations. To match this capacity with conventional pairs of copper wires would require a cable many times larger. Apart from such technological advantages, the light-guide system will save copper and greatly increase the potential capacity of existing underground duct systems.

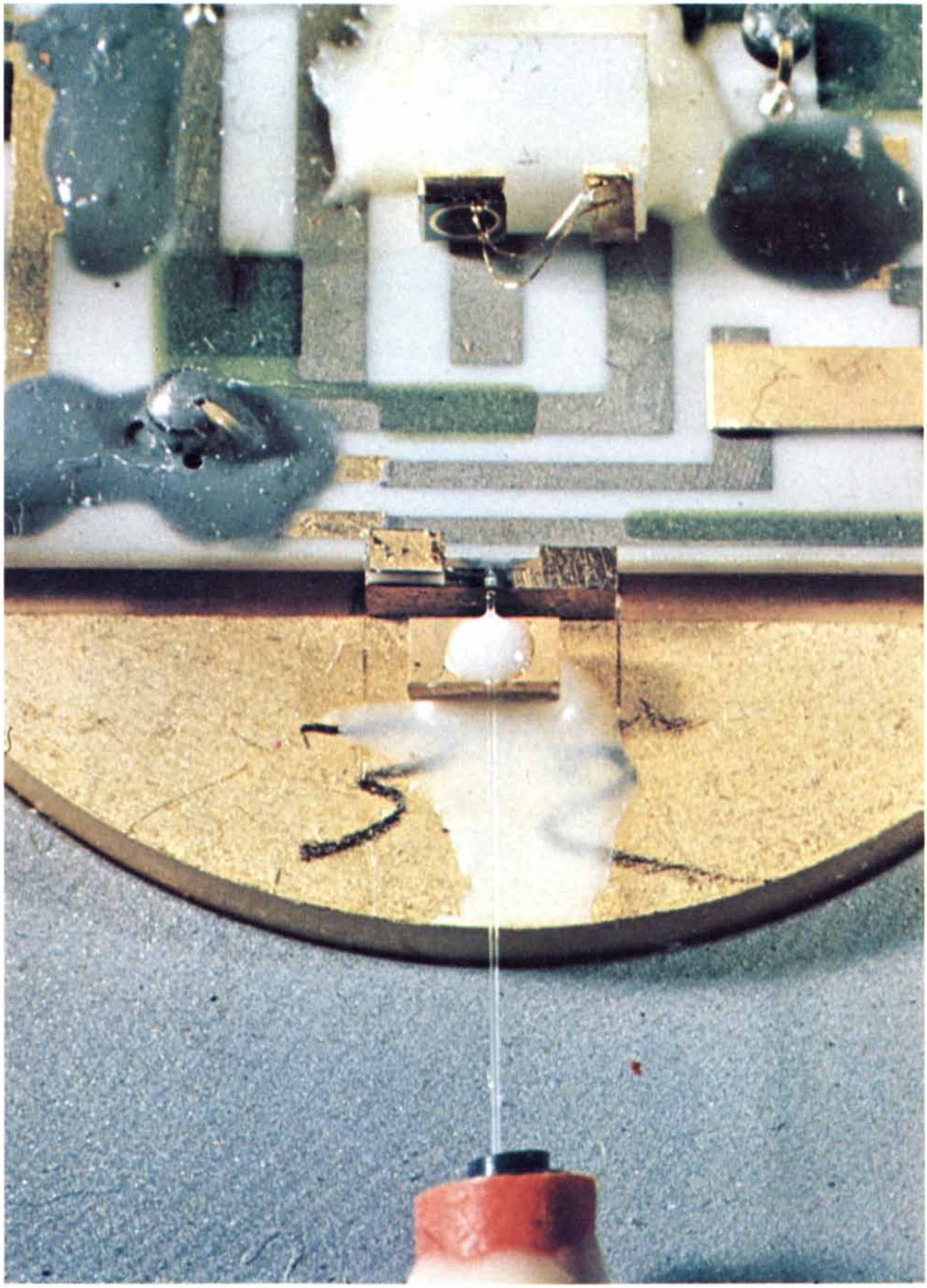
There is nothing particularly new in using light for communication. After all, the American Indians sent up smoke signals and the English built bonfires to warn of the approach of the Spanish Armada. In the 1790's Claude Chappe built an optical telegraph system consisting of semaphore stations on hilltops throughout France. The system, which reputedly could transmit messages a distance of 200 kilometers in 15 minutes, remained in service until it was superseded by the electric telegraph. In 1880 Alexander Graham Bell invented the "photophone," with which he demonstrated that speech could be transmitted on a beam of light. In one system Bell focused a narrow beam of sunlight onto a thin mirror. When the sound waves of human speech caused the mirror to vibrate, the amount of light energy transmitted to a selenium detector varied correspondingly. The light reaching the detector caused the resistance of the selenium, and therefore the intensity of the current in a telephone receiver, to vary, setting up speech waves at the receiving end. And at least until World War II it was common for naval vessels to exchange messages with Morse-coded light signals.

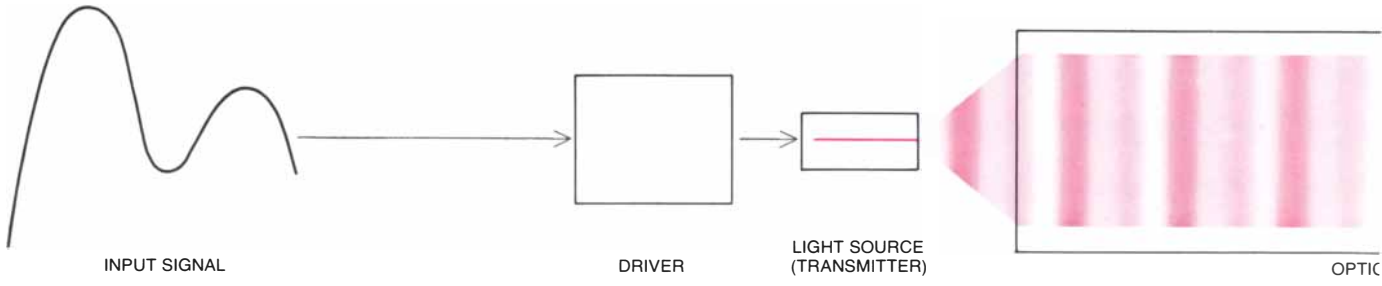
What is new today are the techniques available for generating a light beam that can be modulated at extremely high rates and, equally important, for transmitting the resulting signals through a glass fiber several miles long with an acceptably low loss of energy. The modern interest in light-wave communications dates from the first demonstration of the laser in 1960. This device, which can emit a nearly monochromatic beam of intense visible or infrared radiation, opened up a region of the electromagnetic spectrum whose frequencies were 10,000 times higher than the highest frequencies then in service for radio communication systems. Since potential information-carrying capacity increases directly with frequency, communication engineers had expended great ingenuity over many decades developing systems of ever higher frequency. From the early days of radio they had pushed useful frequencies gradually upward by about five orders of magnitude, from about 100 kilohertz (100,000 cycles per second) to about 10 gigahertz (10 billion cycles per second). Now the laser provided an increase of four more orders of magnitude to 100 terahertz (100 trillion cycles per second). By utilizing only a small part of the full range of light frequencies generated by the laser a single light-wave system could in principle simultaneously carry the telephone conversations of every person living in North America.

The early lasers, however, were cumbersome and unreliable; the best of them failed after a few months of operation. Furthermore, it soon became obvious that the concept of projecting laser beams from point to point through the atmosphere, in analogy with microwave radio-relay systems, was unsatisfactory because the signal would be attenuated by fog, smog, rain and snow. Actually it is easier to transmit light pulses reliably from Arizona to the moon than it is to transmit them between downtown and uptown Manhattan.

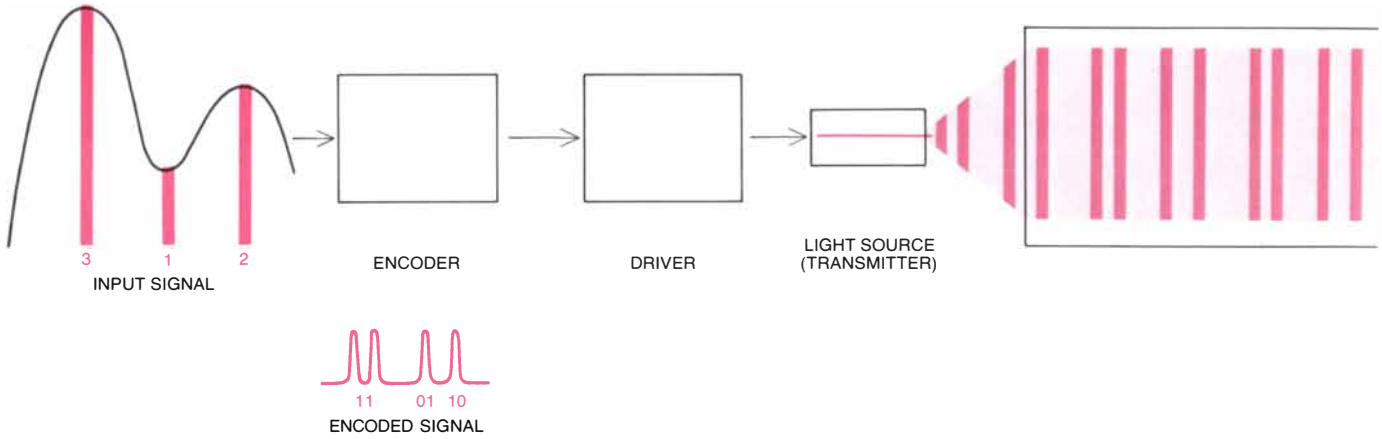
There has been steady progress in making lasers compact, reliable and long-lasting and in circumventing the transmission of the light beams through the atmosphere. Moreover, although for some demanding applications lasers are still preferred, for others a simpler and cheaper device, the high-intensity light-emitting diode (LED), is adequate. The first promising alternative to transmitting light signals through the atmosphere consisted in sending optical signals through a light pipe: a carefully fabricated tube a centimeter or so in diameter provided with optical means (possibly local variations in the density of a gas in the pipe) for bending the rays

LIGHT-GUIDE REGENERATOR CIRCUIT receives the weakened light pulses that are delivered by one optical fiber of a light-wave communication system and amplifies them into sharp new pulses for the next leg of their journey. In this photograph, which was made at Bell Laboratories, a portion of the amplifier circuit and the optical fiber that will carry the regenerated pulses are enlarged some 10 diameters. The fiber can be seen in the lower part of the picture where it emerges from the red-tipped cable housing. The end of the fiber is held snugly against a tiny gallium arsenide laser by a drop of epoxy resin. The regenerated light pulses that are produced by the laser can travel up to 14 kilometers through the fiber before they need to be regenerated again. A photodiode, housed in the white block above the laser, picks up the laser's back emission and uses it to compensate laser drive circuit for variations in temperature.





LIGHT-WAVE COMMUNICATION SYSTEMS can be designed to transmit signals coded in various ways. In the simplest "analogue" system, depicted here, the amplitude of the incoming signal is translated directly into amplitude variations in the light beam entering the optical fiber. A photodetector at the receiver converts the varying intensity of light into a corresponding electrical signal, which is then



DIGITAL CODING is being adopted for commercial light-wave communication systems because of its many advantages over amplitude, or analogue, coding. In digital coding the amplitude, or height, of the input waveform (*left*) is sampled electronically at regular intervals (*vertical bars under waveform*). For accurate representation the wave must be sampled at twice the rate of its highest-frequency component. Thus a voice signal, which has a maximum frequency of 4,000 cycles per second, must be sampled 8,000 times per second. (The sampling rate would be higher than the one suggested by this simplified illustration.) The individual sample heights are coded into sequences of binary digits: 0's and 1's. For transmission 1 can be represented by a pulse and 0 by the absence of a pulse. In a typical

wherever the pipe had to diverge from a straight line.

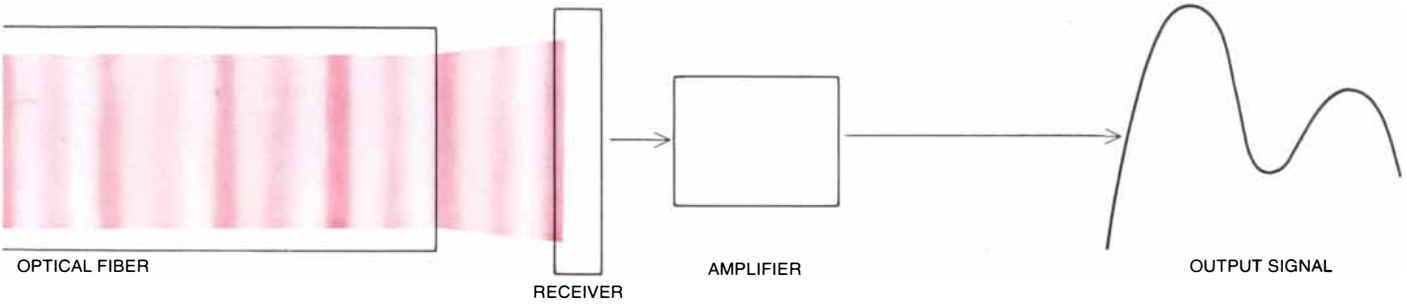
As an alternative to the light pipe, which presented many practical problems, communication engineers began studying the possibility of transmitting light through glass fibers. Bundles of glass or plastic fibers had been used for some time to carry light short distances, for example to light an instrument panel or to examine the interior of the stomach, but they were not nearly transparent enough for the purposes of light-wave communications. The materials commonly employed were less transparent than water. The glass fibers ultimately developed for communications are so transparent that if seawater were as clear as they are, one could easily see to the bottom of the deepest ocean.

Before considering the kinds of communication system one can build from lasers, light-emitting diodes and glass fibers let us examine how information from a source, such as a telephone, a television camera or a computer, is made suitable for transmission by light. In conventional "analogue" transmission systems the wave pattern of the original signal is used to modulate the amplitude of the energy entering the transmission line, in this case the amplitude of the light beam emerging from a light source and entering a glass fiber. At the far end of the fiber the light enters a photodetector that converts the varying intensity of the light into a corresponding electrical signal. The signal is amplified as needed to reproduce the incoming electrical waveform for presentation to the ear, the eye or an inanimate device such as a computer.

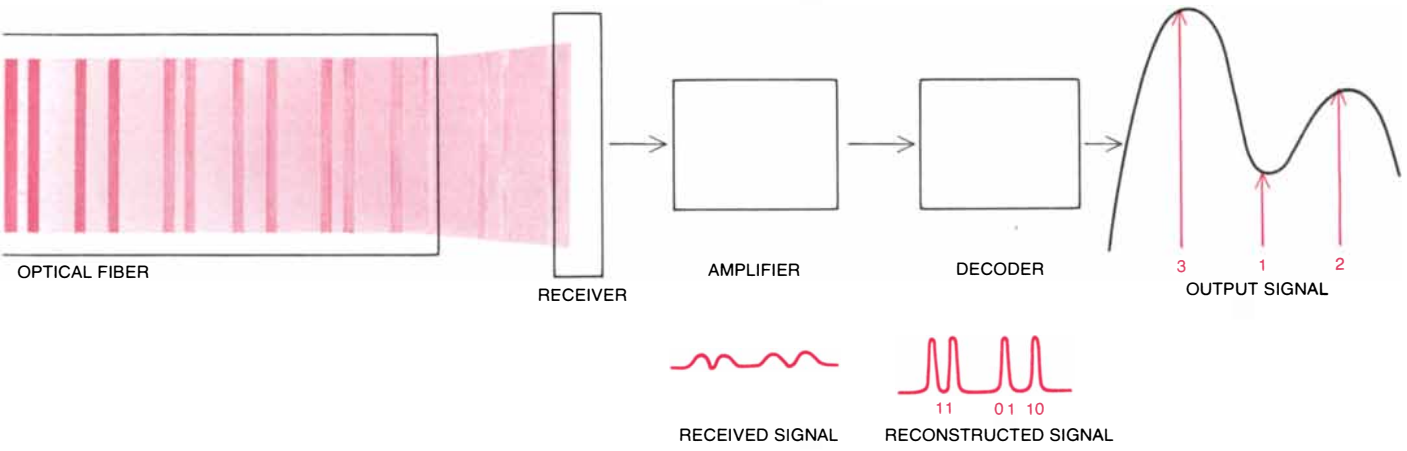
Even in the best fibers some of the light is lost by absorption and scattering, so that the strength of the light signal decreases geometrically as the signal travels from the source to the detector. For example, if the strength of the light signal falls to half its original value after the signal has traveled a kilometer, the strength will fall to a fourth of the original value at the end of the second kilometer, and so on. Thus for long-distance transmission the light source should be as powerful as possible and the detector as sensitive as possible, other things being equal.

At present this requirement is best met by high-intensity lasers and ultrasensitive photodetectors of the "avalanche" type, that is, detectors in which each incoming photon triggers an avalanche of electrons. It should be observed, however, that the maximum transmission range is much more dependent on losses in the fiber than on the power of the source or the sensitivity of the detector. For example, decreasing the loss by a factor of two will exactly double the range, whereas increasing the power of the source by a factor of two will normally increase the range by only about 10 percent (to be precise, by that length of fiber which will increase the loss by a factor of two).

The chief drawback of the analogue transmission system I have been de-



amplified just enough to reproduce the original waveform. In traveling through an optical fiber the strength of the signal decreases geometrically with distance. Moreover, signal becomes increasingly degraded and distorted, so that reconstructed waveform no longer precisely matches original. An efficient way to combat such degradation is to use a digital coding system such as the one depicted below.



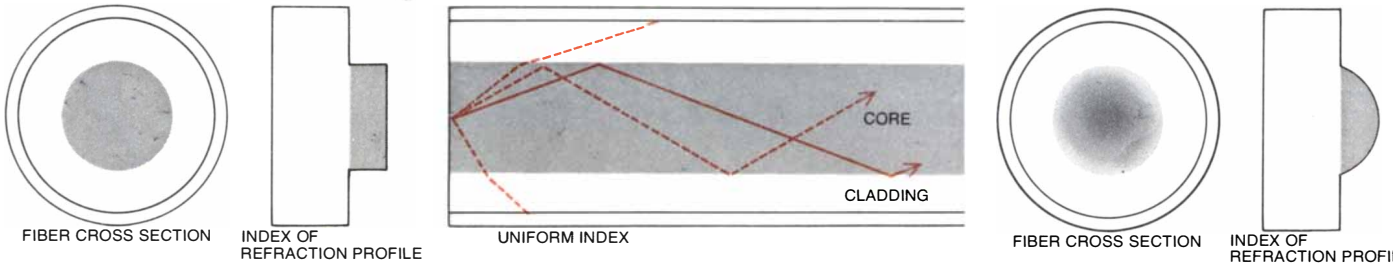
voice system the height of the waveform at each sampling point is assigned a value between 0 and 255, which requires a sequence of eight binary digits, or bits (since 256 equals 2^8). Hence to sample a voice wave for one second the digital system requires 64,000 bits (8,000 samples times eight bits per sample). In the light-wave communication system that has been operating since May of this year in Chicago, the light sources used as transmitters are designed to generate 44.7 million pulses per second and so are capable of transmitting more than 650 voice signals simultaneously. Although the light pulses are degraded in passing through the optical fiber, the pulses can be cleanly regenerated (since a pulse is either present or absent) and used to reconstruct the original waveform with high fidelity.

scribing is that if the amplitude-modulated signal is distorted in any way during its passage through the fiber, and a certain amount of distortion is unavoidable, the distortion will be superimposed on the signal that is extracted and amplified at the receiver. One of the most effective means for transmitting an essentially distortion-free signal is to encode the signal into digital form before transmitting it. This is done by sampling the amplitude, or height, of the continuous signal wave electronically at regular intervals. If the wave is to be represented accurately, it must be sampled at twice the rate of its highest frequency component. Hence a voice signal with a maximum frequency of 4,000 hertz will be accurately represented if it is sampled 8,000 times per second. The individual sample measurements are coded into binary form, represented by a series of 1's and 0's. The binary numbers are now transmitted according to a prearranged code. For example, 1 can be transmitted as a pulse of light and 0 by the absence of a pulse. At the receiver the pulses are detected and used to reconstruct the original wave.

The most important advantage of digital transmission comes in dealing with weak signals. Every detector has an inherent internal noise that corrupts the signal entering the detector to a greater or lesser degree. Thus communication engineers commonly talk about signal-to-noise ratios. The ratios are measured on a logarithmic scale to the base 10 and in units of the decibel. A decibel is defined as 10 times the logarithm of the ratio of two power levels. For example, a signal-to-noise ratio of 20 decibels signifies that the signal level is 100 times higher than the noise level. Since digital pulses are either present or absent, they can be detected with a low probability of error even in the presence of significant noise. For example, with a signal-to-noise ratio of 21 decibels only one pulse in a billion will be lost in the background noise. For analogue signals, on the other hand, any noise tends to distort the message; hence if the signal is to be satisfactorily reproduced, the signal-to-noise ratio must be much higher than 21 decibels. Typically a signal-to-noise ratio of 60 decibels is needed, that is, a signal a million times greater than the noise.

The digital transmission system's greater tolerance of noise means that digital signals can be transmitted farther than analogue signals before amplification is needed. Another great advantage of digital transmission lies in the ease with which digital pulses can be detected and regenerated. Since minor distortions in the shape of the pulse are of little consequence, the weakened pulses can be detected and regenerated without imposing stringent requirements on the amplifiers.

It has become increasingly common in telephony to send voice signals over cable or microwave transmission systems by means of digital pulses. The voice signal is sampled 8,000 times per second, with eight binary digits specify-



OPTICAL FIBERS are designed so that the core has a slightly higher index of refraction than the cladding. Thus most light rays that strike the interface between the two regions are reflected back into the core. Unless the fiber takes a sharp bend the rays will carom from side to side indefinitely. Only rays that enter the fiber at a large angle escape. If the core has a uniform index of refraction (*left*), rays that make many reflections follow a longer path and arrive behind rays that make fewer reflections. This defect, called modal disper-

ing the "height" of each sample. Since eight binary digits are capable of specifying 2^8, or 256, levels of amplitude, they provide an accurate specification of the wave pattern. This means that in order to reproduce the original voice wave, which has a frequency bandwidth of 4,000 hertz, the digital system must be able to transmit 64,000 pulses per second. The large bandwidth of light-wave systems makes it attractive to be somewhat generous in the use of bandwidth in return for a vastly improved signal-to-noise performance, which pays off handsomely in range, or the distance the signal can travel before it must be regenerated.

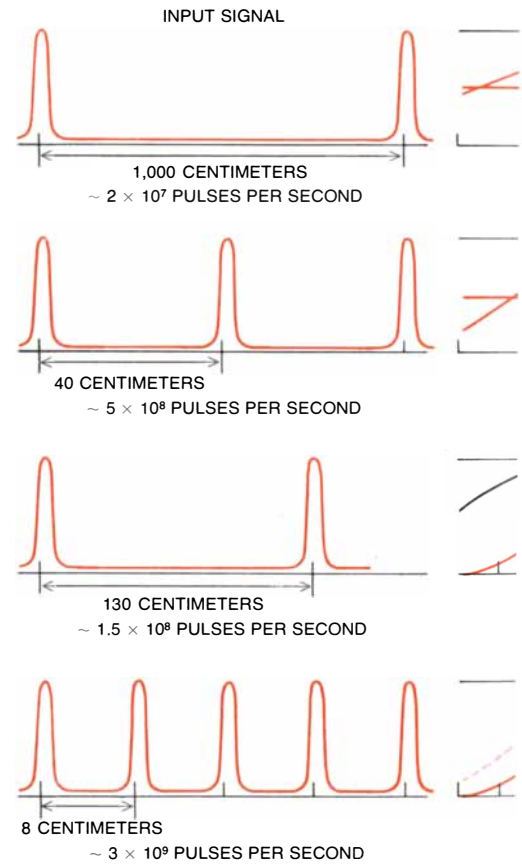
Thus we see that in a practical light-wave communication system the range depends on the power of the source, the attenuation per unit length of fiber, the noise level of the detector and the kind of modulation or coding that is employed. The capacity of the system in bandwidth, pulses per second or any other measure of information capacity depends on the speed with which the source can be turned on and off, the response speed of the detector and also on the pulse-spreading characteristics of the fiber.

Two kinds of light source are currently in service. The first is a refined version of the LED's that are found in the visual-display units of pocket calculators. For light-wave communications what is needed is a source that is not only much more intense than that of visual displays but also comparable in size to the optical fiber, which is only a few hundredths of a millimeter in diameter. In LED's designed for light-wave communications a small hole, or well, is etched into the face of the LED in order to bring the fiber as close as possible to the active region of the semiconductor junction, where the light originates. The fibers have their lowest loss in the infrared region of the spectrum, so that one selects a semiconductor material that emits infrared radiation. LED's made from gallium arsenide, which emit at a wavelength of about .8 micrometer, are satisfactory, but they would be even better if their emission wavelength were somewhat longer. Semiconductor materials that promise a better wavelength match for present optical fibers are under active investigation.

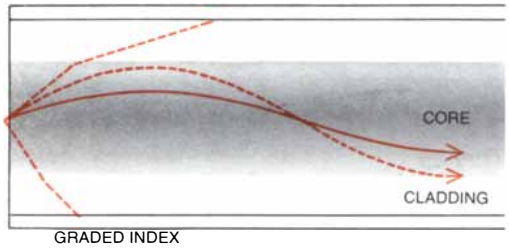
The other type of light source is the semiconductor diode laser, which has a more complex structure than the light-emitting diode. No bigger than a grain of salt, a diode laser consists of several layers of semiconductor material, each one of a different composition. The sandwich structure helps to establish the conditions necessary for laser action; it provides a region that confines the charge carriers giving off light when they recombine and also helps to guide the light in a preferred direction.

It was once difficult to lay down successive layers of material in such a laser without spoiling the crystal structure of the layers. The early devices were therefore notorious for their rapid decay in light-emitting efficiency; some failed completely within a few hours. New techniques were gradually developed so that the composite structures can now be fabricated without introducing imperfections in the crystalline layers. Accelerated aging tests demonstrate that the most recently developed devices should last for several years at room temperature. Ultimately laser diodes should be fully as reliable as other solid-state devices.

The laser source has two main advantages. The first is its directionality. Because the stimulated emission from the laser emerges in a narrow beam it is possible to couple a large fraction of its radiation directly into the end of an optical fiber. The second advantage is its small spread in color, or wavelength, which is typical of a laser source. In traveling through a light guide rays of different wavelengths travel at slightly different velocities; hence the broadening of the pulses in a light guide varies directly with the width of the band of wavelengths transmitted. Laser sources are therefore able to transmit pulses at a higher rate over a given distance than LED's, which emit a broader band of wavelengths. The spectrum of a typical diode laser has a width of only 20 angstroms, compared with 350 angstroms for a light-emitting diode of the type suitable for light-guide communications. After traveling a kilometer through an optical fiber, the laser pulse will show a dispersion in time of 200×10^{-12} second, equivalent to a dis-



TWO TYPES OF DISPERSION, or pulse-spreading, must be dealt with in light-wave communications. The first type, modal dispersion, is shown in the illustration at the top of these two pages. This set of diagrams shows how pulse-spreading limits the pulse rate and hence the capacity of optical fibers. In each case the pulse rate is selected so that the spreading amounts to half the interval between pulses. For a fiber with a core of uniform refractive index (*a*) the difference in arrival time between the leading edge and the



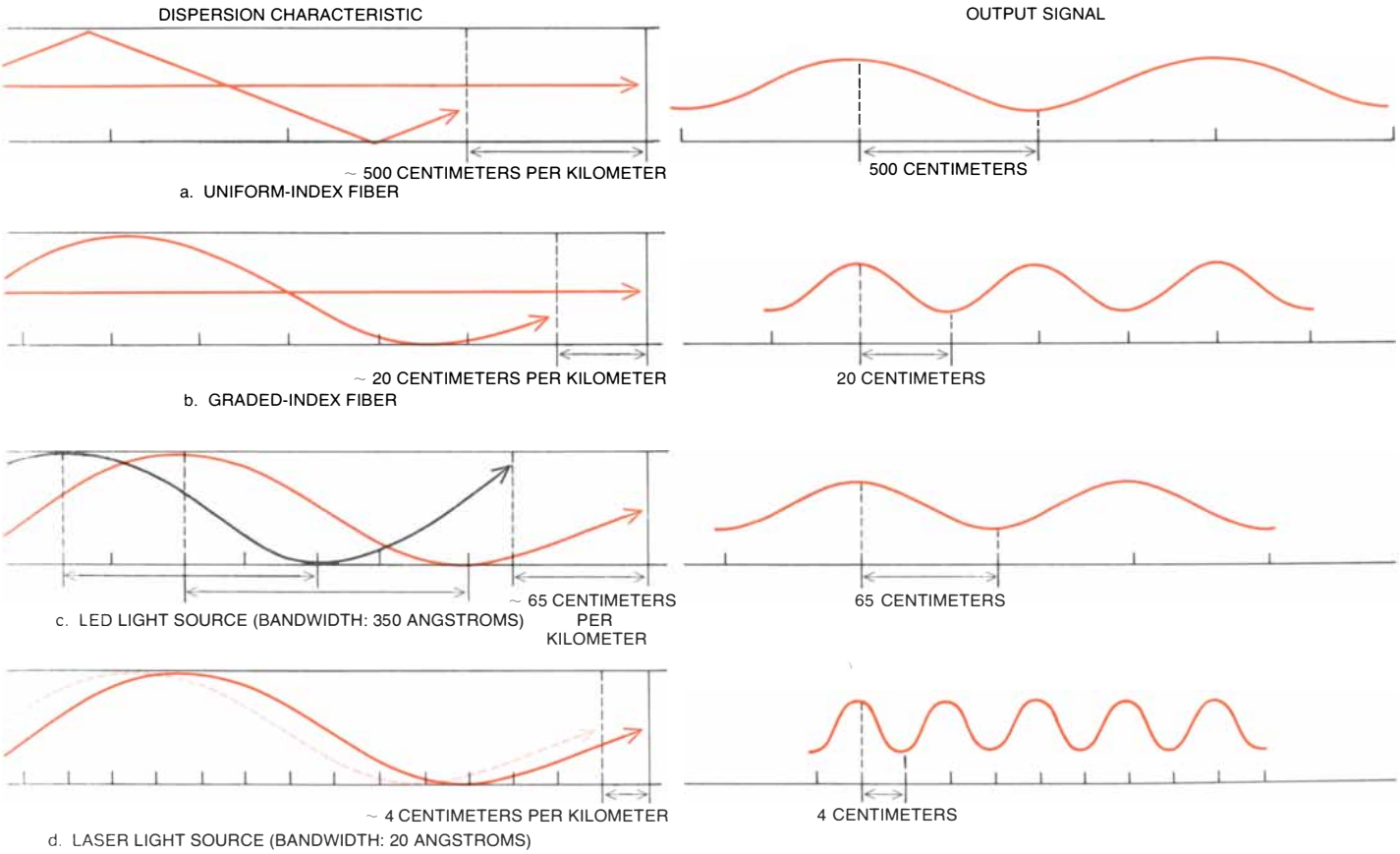
GRADED INDEX

sion, is remedied by fabricating a fiber whose index of refraction increases toward the axis (*right*). Now rays that deviate from the axis travel faster than the rays that deviate less.

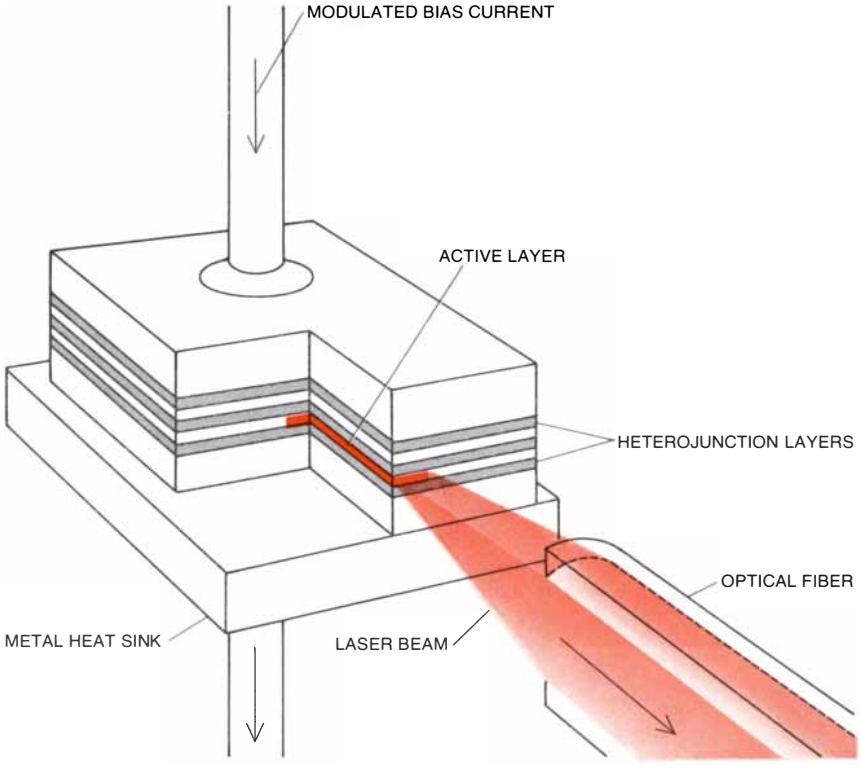
persion in distance of four centimeters at the reduced velocity with which light travels through glass. For an LED source the dispersion is nearly 20 times greater. This dispersion due to a lack of spectral purity places an important limitation on the pulse rate and thus on the information capacity of light-wave communication systems. Another major limitation is a type of pulse-spreading due to modal dispersion, which arises because some of the rays entering an optical fiber travel slightly longer paths than other rays. As we shall see, modal dispersion can be greatly minimized but not entirely eliminated.

In order to obtain the extraordinary transparency needed for a light guide the optical fibers are designed so that the light never comes near the outside surface of the fiber, where dust, scratches or contact with other surfaces would cause serious losses. Each fiber actually consists of three layers. The outer layer is a coating, usually of plastic, that provides protection from scratches and abrasion, which could weaken the fiber and lead to breakage under stress. Within the protective coating the glass fiber itself has a core region with an index of refraction slightly higher than that of its surrounding cladding. Because of this higher refractive index, rays that enter the end of the fiber at a shallow angle to the central axis are reflected back into the core when they strike the interface between the core and the cladding. Rays that enter the fiber at large angles to the axis simply escape without being reflected. One can see from geometric considerations that if a ray is reflected back into the core at its first encounter with the interface, it will continue to be confined indefinitely, provided there are no sharp bends in the fiber. Such bends can be avoided by carefully encasing several fibers in a fairly stiff cable sheath.

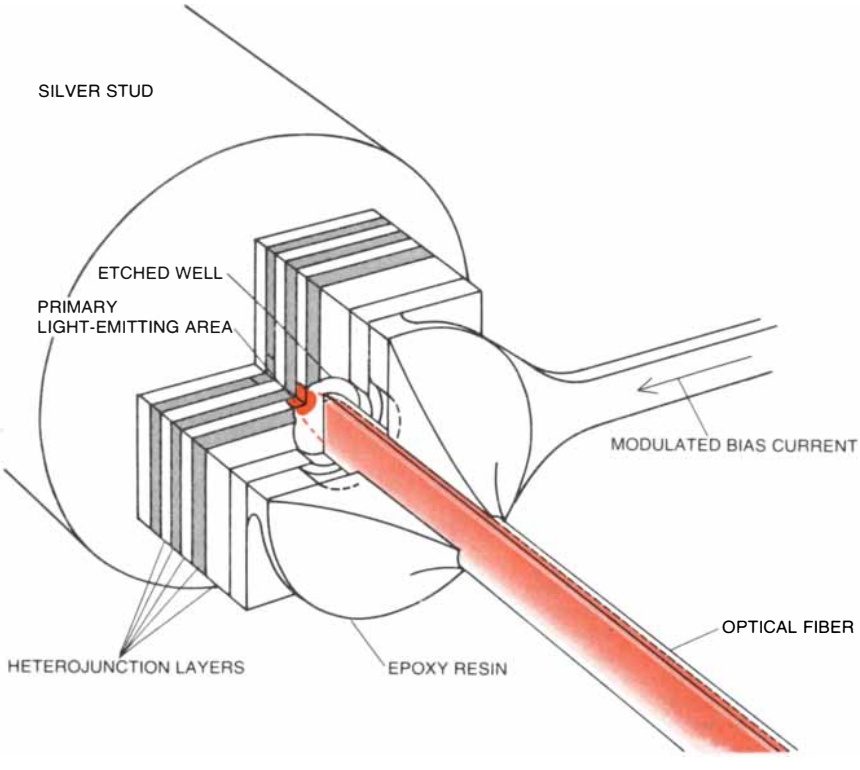
One can now visualize the source of modal dispersion: a light ray that enters parallel to the central axis of a fiber will normally travel a shorter distance than a ray that enters at an angle and caroms from side to side as it travels down the fiber channel. As a result a light pulse



trailing edge of an arbitrarily sharp pulse amounts to 25×10^{-9} second per kilometer, equivalent in distance to about 500 centimeters at the velocity light travels in glass. For the spreading to be held to half that interval the pulse rate cannot exceed 2×10^{7} pulses per second. Fibers with a graded refractive index reduce the modal dispersion by a factor of 25, to 10^{-9} second per kilometer, or to about 20 centimeters (*b*). Hence arbitrarily sharp pulses could be transmitted at the rate of 5×10^{8} pulses per second. (Distances shown for pulse-spreading are plotted in logarithmic units; the pulse intervals at the far left are plotted as simply being equal to twice the spreading distance.) The second type of pulse-spreading, wavelength dispersion (*c and d*), arises because the velocity of electromagnetic waves varies with the refractive index of the medium: the higher the frequency, the lower the velocity. The high-intensity light-emitting diodes (LED's) being used in light-wave communications have a spectral bandwidth of about 350 angstroms, centered in the infrared part of the spectrum at a wavelength of .82 micrometer. Their bandwidth is roughly equal to the interval between green and yellow in the visible spectrum. Even in a graded-index fiber with zero modal dispersion a single pulse from an LED would spread about 65 centimeters per kilometer, thus limiting the signaling rate to about 1.5×10^{8} pulses per second (*c*). With a laser light source (*d*), which as a much narrower bandwidth (about 20 angstroms), wavelength dispersion amounts to only four centimeters, allowing a signaling rate of 3×10^{9} pulses per second.



LASER LIGHT SOURCE used in light-wave communications is about the size of a grain of salt. Laser action (light amplification by stimulated emission of radiation) takes place within an assembly of "heterojunction" layers of semiconducting gallium arsenide and aluminum gallium arsenide with precisely controlled electronic characteristics. Laser beam issues from one of gallium arsenide layers and travels 40 micrometers through a sealed gas medium before entering optical fiber. Such lasers have power of .5 milliwatt at wavelength of .82 micrometer.



LIGHT-EMITTING DIODES, which are also made from heterojunction layers, are simpler, cheaper and more reliable than lasers. They are useful where narrow bandwidths are not needed and for transmission distances where their low average power of .1 milliwatt is adequate.

made up of a combination of rays becomes spread out over time.

To overcome pulse-spreading many of today's fibers have a core whose refractive index is graded, or shaped, to compensate for the different distances the rays travel. In such fibers the refractive index decreases with radial distance away from the center. In the regions of lower refractive index the light travels faster. It is possible to arrange the radial decrease in refractive index so that all rays arrive at their destination at more nearly the same time. In a fiber of uniform refractive index the pulse-spreading amounts to about 25×10^{-9} second per kilometer, which is equivalent to about 500 centimeters. Graded-index fibers now being tested in the field reduce this dispersion by a factor of 25, and in laboratory samples an improvement of 100 to one has been demonstrated.

The first high-transparency fibers were made by the Corning Glass Works out of a material whose principal component was silicon dioxide. The first successful fiber with a graded refractive index was made by the Nippon Sheet Glass Co., Ltd. In a process developed by Bell Laboratories a graded-index fiber is made by heating and collapsing a three-foot tube of quartz glass that has previously been coated on the inside with dozens of precisely controlled layers of silicon dioxide doped with germanium. Each layer is only about a hundredth of a millimeter thick. The tube is collapsed into solid rod called a preform, which is then drawn into a fiber a few kilometers long.

In the best fiber specimens the transmission losses can be as low as one decibel per kilometer, which is equivalent to 80 percent transmission of the input energy. Such low losses, however, cannot be achieved at the operating frequency of available light sources; a more realistic average loss figure is four or five decibels per kilometer, or about 30 percent transmission of the input energy. Even at this value laser-light pulses can be transmitted a distance of 14 kilometers before amplification is needed. (At that distance only 10^{-7} of the input energy survives.) Undoubtedly as the sources and detectors are "tuned" to the region of the spectrum where the fibers show a minimum loss at wavelengths somewhat longer than one micrometer, and as the fibers themselves are improved, the distance between amplifiers can be extended substantially beyond 14 kilometers.

The hair-thin light guides are readily assembled into cables. After they are coated for protection against humidity, abrasion and losses due to bending, the fibers are assembled into flat, color-coded ribbons, each ribbon containing 12 fibers. Up to a dozen ribbons are enclosed in a cable that cushions and protects the individual fibers against dam-

age in field service. Considerable ingenuity was required to devise efficient splicing methods. A technique was finally developed that can align all the fiber ends in a cable to an accuracy of within two micrometers.

Light guides offer a number of advantages over transmission by metallic conductors. Since the light in a light-guide transmission system is tightly confined to the inner core of each fiber, signals cannot leak between adjacent fibers and give rise to "cross talk." Moreover, since light guides are not affected by electrical interference from other sources, light-wave systems should show advantages in carrying information in electrically noisy environments, such as between switching apparatus in telephone central offices.

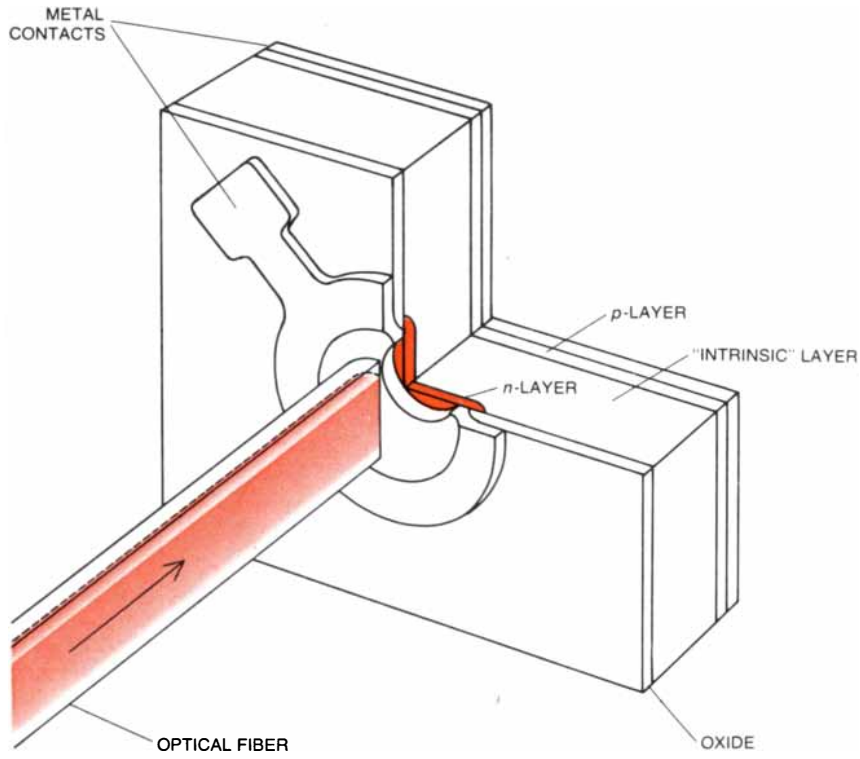
Cables for light-wave communications offer distinct savings in materials compared with metallic cables of equivalent capacity. At present the optical fibers are much more expensive than copper wire, but this is to be expected when a technologically complex new product is first put into production.

Just as there are two kinds of light source for light-wave communications, so there are two types of detector in service. Both are solid-state devices. One is a simple device of the junction type known as a PIN detector, rather similar to a solar cell, in which photons of light generate an electric current. (The letters P, I and N stand for the electronic properties of the semiconductors used in the junction of the detector.) The other device is the avalanche photodetector mentioned above. All signal detectors have a background noise that increases in proportion to their operating speed. For example, the background noise in a PIN detector increases from 10^{-11} watt when the device is operated at one megabit per second to 10^{-9} watt at 100 megabits per second. At the same operating speeds the background noise in an avalanche detector is lower by a factor of 10. It follows that the transmission distances for low-speed systems are greater than those for high-speed systems. In light-wave communications the signal detector is the first stage in a receiver module that contains the circuitry needed to adapt the signals for transmission through the existing telecommunication network.

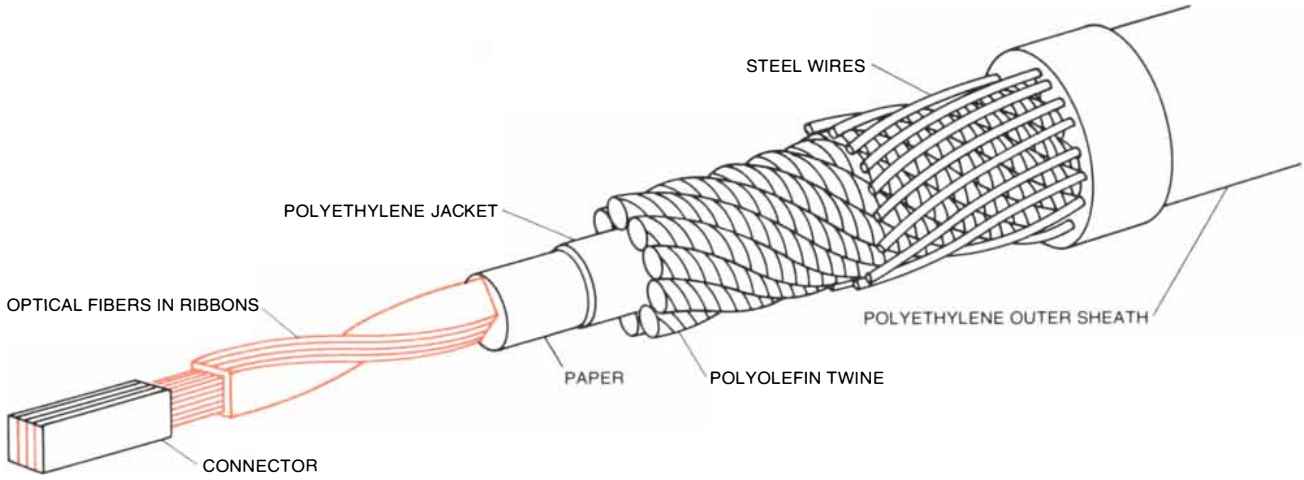
Let us now assemble these various kinds of information about sources, detectors and fiber properties and see what kinds of communication capabilities are available. First let us calculate the range for a low bit-rate system, one capable of transmitting 10^6 bits per second. In order for detection to proceed with few errors the signal arriving at the detector must be 100 times larger than the detector's internal noise. If an avalanche photodetector is employed, the arriving signal must have a power level of at least 10^{-10} watt. For maximum range we would choose a laser with a power output of 10^{-3} watt in preference to an LED, which is an order of magnitude less powerful. As we have seen, with digital coding the maximum allowed attenuation of the light passing through the fiber is a factor of 10^7, or 70 decibels. Since present production fibers have an attenuation of less than five decibels per kilometer, we can expect satisfactory transmission for a distance of 14 kilometers before amplification is needed. (If fibers with an attenuation of only one decibel per kilometer are perfected, the range could be stretched to 70 kilometers.) In practice it seems doubtful that fibers will ever be available with continuous lengths much greater than a few kilometers. Therefore the extra loss introduced at the junction of two fibers must be added to the loss figure. Present plug-type connectors introduce a loss of about .5 decibel. If one were to need six connectors for a 14-kilometer route, the additional loss would come to only three decibels. (The total loss could be held to 70 decibels by shortening the route by three-fifths of a kilometer.)

With the selection of a source, a detector and a fiber, what will the information-handling capacity of light-guide systems be? Since it is desirable to transmit at the highest possible bit rate, one must consider a number of factors. As we have seen, the noise level of the detector increases with the bit rate. Thus if the signal power was just adequate for transmission at 10^6 pulses per second, it would have to be raised by a factor of 100 for transmission at 10^8 bits per second. In addition, as the pulses get shorter and closer together their spreading as they travel through the fiber becomes an important limiting factor.

In order to simplify the calculation let us decide somewhat arbitrarily that the pulse-spreading will not be more than half the interval between successive pulses. In a graded-index fiber the pulse-spreading due to modal dispersion (the difference in path lengths) amounts to about 10^{-9} second per kilometer, which means that if one attempts to transmit



LIGHT DETECTOR at the terminal of an optical fiber generates electrons when it is struck by photons. The simplest detector is the PIN photodiode shown in this illustration. The term PIN denotes that the detector consists of an "intrinsic" (*i*) region of weakly *p*-type silicon sandwiched between layers of *p*-type and *n*-type silicon. (A *p*-type, or positive-type, material is deficient in electrons; an *n*-type, or negative-type, material has a surplus of them.) Photons of light that are absorbed in the *i* region of the sandwich give rise to electrons and "holes" (electron vacancies) that move under the action of a uniform electric field to produce an electric current. A somewhat more complex photodetector, known as an avalanche photodiode, incorporates an additional layer of *n*-type material that provides a built-in amplification process for enhancing the electrical signal. Photodiodes have inherent noise that increases with operating speed. When a PIN detector is receiving signals at a rate of 10^8 pulses per second, its inherent noise is about 10^{-9} watt. For avalanche detector inherent noise is lower by factor of 10.



LIGHT-GUIDE CABLE of the type used in last year's demonstration in Atlanta, Ga., contains 144 individual glass fibers assembled into 12 ribbons of 12 fibers each. The ribbons are stacked side by side and are surrounded by several layers of protective material, all contained within a polyethylene sheath reinforced with steel wires. Operating at a pulse rate of 44.7 million cycles per second, a pair of fibers can carry 672 two-way telephone conversations, a two-way Picturephone conversation or an equivalent volume of information of other types. Cable sections are equipped with factory-prepared connectors that are mated in the field with the aid of precision jigs.

10^9 pulses per second, the spreading is equivalent to the entire interval between pulse peaks. Therefore to maintain a separation of half an interval the signaling rate cannot exceed $.5 \times 10^9$, or 5×10^8, pulses per second. This is the limit if we have a laser source, which is so nearly monochromatic that pulse-spreading due to wavelength dispersion can be ignored.

If we choose an LED light source, however, wavelength dispersion becomes the limiting factor in the signaling rate. For an LED the wavelength dispersion amounts to 3.5×10^{-9} second per kilometer, a figure 3.5 times larger than that for modal dispersion. In order to hold the spreading below half the interval between successive pulses the signaling rate with an LED source must therefore be slightly less than a third of the rate permissible with a laser, or 1.4×10^8 pulses per second. Naturally as the desired transmission distance is increased the signaling rate must be proportionately reduced. For example, for a nominal transmission distance of 10 kilometers the rate for the laser source would have to be reduced tenfold to 5×10^7 pulses per second, or approximately the rate (4.47×10^7) actually selected for the Chicago installation. These simple calculations illustrate what can be achieved with today's technology and also give a feeling for the kinds of design choice that can be made among range, capacity and device complexity. Undoubtedly there will be significant improvements in the future.

There are many promising areas of application for the new light-guide technology. For example, television signals could easily be carried over a single fiber, thereby opening up new possibilities for both entertainment and business purposes. Buildings could be "wired" with almost invisible fibers to provide internal communication services. The parts of computers could be interconnected with fibers. It is in telephony, however, that one can expect some of the first important applications.

Today much of the copper cabling that interconnects metropolitan telephone-switching centers goes through underground ducts where space is at a premium. Adding new duct space is both costly and inconvenient. Light-wave communication systems, with their high capacity and small size, could make better use of the existing underground ducts and help to postpone the need for new ones. Moreover, since adjacent switching centers in many cities are less than seven kilometers apart, light-wave systems might not require any amplifiers in manholes to boost signals along a typical route.

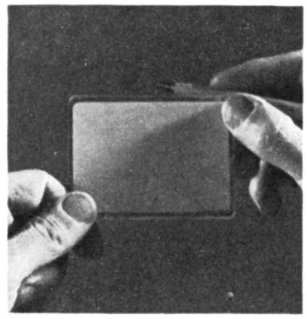
Before the completion of the Chicago installation Bell Laboratories and the Western Electric Company tested a prototype light-wave system under simulated field conditions last year in Atlanta. Two light-guide cables 640 meters long, each containing 144 fibers, were pulled through standard underground ducts and subjected to tests simulating a typical urban telecommunication environment. The installation work did not break any of the fibers, and the pulling operation, which required the negotiation of sharp bends, did not degrade the performance of the light guides. As in the present Chicago system, each pair of fibers carried the equivalent of 672 two-way voice channels. The light sources were gallium aluminum arsenide lasers operating at a rate of 44.7 million bits per second. At the receiving end the light pulses were converted into electrical signals by avalanche photodetectors.

As part of the Atlanta experiment the ends of some individual fibers were joined to create a continuous communication path about 70 kilometers long. With the help of 11 regenerators, or amplifiers, virtually error-free transmission was achieved in one direction over the full path for a sustained period. The Chicago installation closely follows the Atlanta experimental system except that LED's are being used in addition to lasers as light sources.

Apart from some references to the anticipated reduction of fiber losses in the future, everything I have described here is based on current technology. It would be contrary to all previous experience to believe we shall not witness further dramatic developments. For example, a number of industrial and university investigators are conducting experiments with integrated optics, which include techniques for processing light signals within thin films, the optical equivalent of integrated microelectronic circuits. Such optical circuits may someday eliminate the need for converting light pulses to and from electrical signals in amplifiers along transmission paths. In addition both theoretical and experimental work is proceeding on the possibility of switching light pulses directly, obviating the need for first converting the light signals into their electrical equivalent. The hope is to develop optical switches to replace the present electromechanical and electronic devices, thereby making it possible to connect telephone calls in greater numbers and at higher speeds than ever before.

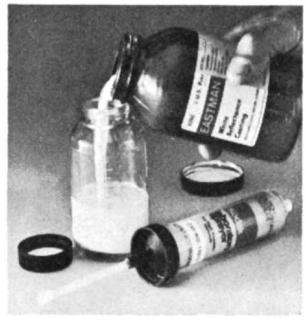
Other Kodak products

(A few nonphotographic things we make because . . . well, because we know how and there seems to be a need . . .)



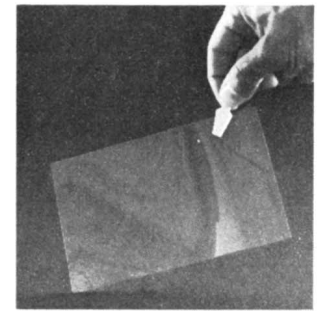
KODAK IR Phosphor—A bit of secret technology uncorked a year after the war in 1946. Still arouses flurries of interest among, for example, those who work with infrared-emitting gallium arsenide diodes when they hear there is such a product which we freely offer for sale to anybody who finds out about it. In effect, translates infrared in the 0.7-1.3 μm range into visible light. We supply the phosphor coated on plastic-enclosed paper or transparent base. It stores up energy by simply being left around in daylight or under fluorescent room light. The phosphor then releases the energy as an orange glow that shows the infrared beam structure with resolution better than 5 lines/mm. Pattern can be photographed. Under continuous IR impingement and with the maximum energy that the phosphor can store, the glow can last up to 60 seconds. Emission persists for 50-60 nanoseconds after the IR goes off. Just recharge and reuse.

Stocked in 2" x 3", 4" x 5", 8" x 10", 14" x 17", and 20" x 24". Orders and inquiries to Special Products Sales, Kodak Apparatus Division, Rochester, N.Y. 14650 or phone 716-325-2000, ext. 63579.



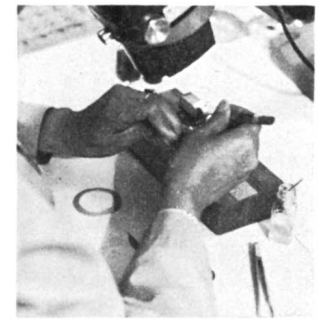
EASTMAN White Reflectance Coating—Even as all cats are black in the dark, so all whites are gray when there is something whiter to compare them with. Instrumentation for measuring optical reflectance properties simply compares the test surface with something assumed to be a perfect diffusing reflector. Such a standard was traditionally prepared by depositing the white smoke from burning magnesium. The microcrystals of MgO soon develop color centers from ambient UV or particle radiation and lose UV reflectivity. TiO_2, the great white opacifier in the visible, is murky indeed in the UV. That leaves this particular product of ours. It's highly purified $BaSO_4$ in a binder and solvent for spray application.

Lab supply houses that sell EASTMAN *Organic Chemicals sell it as* EASTMAN *6080 and can get you a copy of Kodak Publication No. JJ-32 that gives detailed application directions. The stuff seems to have no luminescence at all.*



EASTMAN Q-Switch Acetate Sheet—Cellulose acetate in which we have dissolved our dye Bis(4-dimethylaminodithiobenzil)nickel to produce an optical density of 0.36 ± 0.02 at 1.06 μm, the wavelength of neodymium lasers. This absorbance in the laser cavity permits a buildup of the population inversion of energy states in the Nd^{3+} ions to loose a large pulse when the dye temporarily loses its absorbance. The ability of this dye to do this over and over again without breaking down has been utilized for several years now, but hitherto it has had to be dissolved in a liquid. That is less convenient than having it in a sheet with optical quality of 24 fringes (or fewer) of sodium light per cm^2 and which can handle a peak power density of 300 megawatts/cm^2.

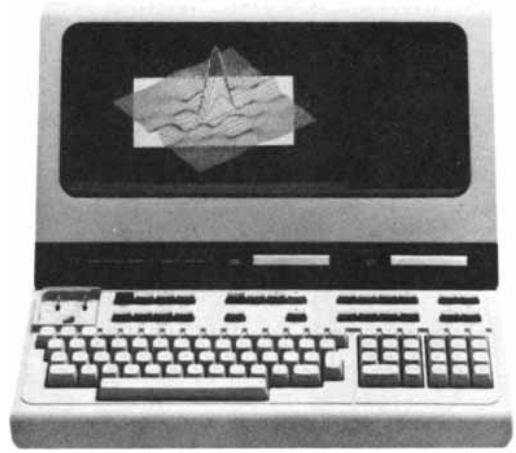
$214.50 for 13 cm x 16.5 cm sheet, subject to change without notice. Orders also accepted for densities other than 0.36. Phone 716-458-4080 and ask for EASTMAN *No. 15064.*



An Mn-Zn ferrite and the service of installing it in broadcast video magnetic recording equipment for 1,000 hours of high-grade service (instead of the "normal" 350-hour average life)—A ferrite is a coordination compound of one or more metals with iron oxide. Spin Physics, Inc. is a corporate grandchild of ours, a division of Eastman Technology, Inc., which is a subsidiary of Kodak. Spin Physics has spent many millions of dollars in developing this particular ferrite and in learning how to design with it and fabricate it. Building on that base, Spin Physics has then proceeded to manufacture or refurbish a substantial portion of all the recording heads currently being fed telemetered analog data. Now Spin Physics has turned from the temples of high technology to the broadcasting industry as an outlet for its services in upgrading the performance and operating life of nonphotographic picture-recording equipment.

Spin Physics' address is 11633 Sorrento Valley Road, San Diego, Calif. 92121.

Kodak



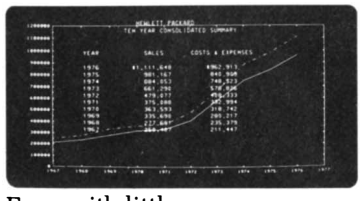
HP introduces a low-cost graphics terminal with interactive alphanumeric capabilities.

The new HP 2648 brings graphics to our growing family of intelligent terminals. It offers high performance at a low cost, and is an invaluable tool for automatic plotting, schematic diagramming, and computer-aided design. It helps you visualize your ideas from concept to execution, and its built-in microprocessor lets you perform a wide range of graphics without any CPU help at all.

The HP 2648 graphics terminal introduces a constellation of bright ideas. Here are some of them:

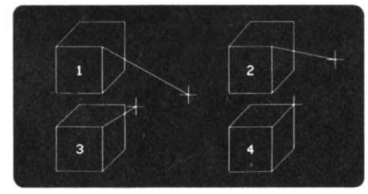
ZOOM AND PAN

Magnify any area of the display up to 16 times (16X) to investigate or modify dense or complex areas. Pan in any direction for close-up viewing of the magnified display, without affecting display memory.



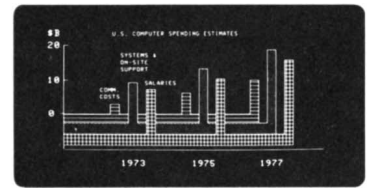
AUTOMATIC PLOTTING

Even with little or no programming knowledge, you can plot tabular data automatically. A simple menu helps you enter your data parameters on the keyboard, after which a single keystroke plots the data instantly.



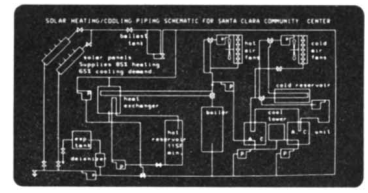
RUBBER BAND LINE

Trial graphics are a snap with the rubber band line. You can anchor a line at any point, then use the graphics cursor to move the other end of the line anywhere you wish. As you move the cursor, the line automatically conforms.



USER DEFINABLE AREA SHADING

You can specify any pattern that an 8- by 8-bit cell can generate to make clear visual definitions of separate areas with similar shapes, as in bar charts and mechanical or architectural drawings.



INDEPENDENT DISPLAY MEMORIES

Alphanumeric and graphic data can be shown independently or at the same time, as in this process control schematic. And because alphanumeric and graphic displays each have separate memories and cursor controls, you can change one without disturbing the other.

Other advantages of the HP 2648 include: storage of graphics and alphanumeric data on 110k byte magnetic tape cartridges; user-definable keys that can be programmed to speed repetitive jobs; and self test at the push of a button.

Price of the HP 2648 starts at $5500*.

Four-digit signature displayed by the HP 5004 signature analyzer (left) greatly simplifies the identification and correction of faulty digital circuits at the component level.

After the honeymoon, how do you support a microprocessor?

If you're contemplating the conjugation of a microprocessor with a product you are developing, you now have an alternative to costly board-exchange programs in the event of faulty circuit components: HP's new 5004 signature analyzer.

Until now, tracing faults in a digital circuit (the type of circuit in which microprocessors reside) has been difficult. The time-honored signal tracing techniques that a field service technician can use to troubleshoot and repair *analog* circuits just don't work with digital circuits, where voltage measurements have no meaning and all waveforms look alike. Instead, manufacturers have had to base field service not on component repair but on much more expensive board exchange or replacement programs.

Signature analysis, a new measurement technique development by HP, provides component-level service capability for digital circuits. Here's how it works. When the modest requirements of signature analysis are designed into a product, the HP 5004 signature analyzer automatically characterizes the bit stream associated with a data node as a four-digit hexadecimal signature. A compressed "fingerprint" of the data present at the node, each signature is unique for a specific good node; any fault in the data stream—even one bad bit out of thousands—generates an erroneous signature. By comparing the displayed signature with the correct one noted on the product schematic, a service technician can easily spot a faulty node, trace it back to a bad component, and repair it without replacing the board—just as with an analog circuit.

The HP 5004 generates the signature by compressing the data stream in a linear feedback shift register. With this technique, there is a 100 percent certainty that the HP 5004 will detect a single-bit error, and a 99.998 percent certainty that it will detect *any* error—regardless of the length of the stream or the subtlety of the fault. It even detects time-related faults such as mid-cycle displaced bits, and speed-related failures in assembled systems.

For the manufacturer, the bottom line is a substantial reduction in field service costs for microprocessor-based products and high-speed state machines. Signature analysis increases development costs by only about 1 percent, an increase that is more than offset by lower manufacturing costs due to reduction in circuit boards, interconnections, and production-line troubleshooting time.

If this sounds interesting, send in the coupon to receive HP Application Note 222, *A Designer's Guide to Signature Analysis*, which tells how to implement signature analysis in the design of a product.

Price of the HP 5004 is $990*.

HEWLETT hp PACKARD

1503 Page Mill Road, Palo Alto, California 94304

For assistance call: Washington (301) 948-6370, Chicago (312) 255-9800, Atlanta (404) 955-1500, Los Angeles (213) 877-1282.

Mail to: Hewlett-Packard, 1503 Page Mill Road, Palo Alto, CA 94304.
Please send me further information on

( ) HP 2648 graphics terminal
( ) HP 5004 signature analyzer

Name _____

Company _____

Address _____

City _____ State _____ Zip _____

*Domestic U.S. prices only.

00747

SCIENCE AND THE CITIZEN

Women in Science

Between the mid-1960's and the mid-1970's the number of women entering all fields of natural science and engineering in the U.S. approximately doubled, but qualified women scientists and engineers still suffer inequities of opportunity, salary and status. Such are the conclusions of a study by Dorothy Zinberg of Harvard University, published recently in *Trends in Biochemical Sciences.*

Zinberg's study, a sequel to one she had conducted for the period between 1960 and 1966, showed that between 1966 and 1974 the overall number of women enrolling in four-year colleges and universities rose from 38.2 percent of the female high school graduates to 44.2 percent. The percentage of women undergraduates majoring in scientific or engineering subjects increased three times faster than the overall enrollment of women. Since 1969 the percentage of undergraduate men majoring in the physical sciences and engineering has actually decreased.

On the graduate level the number of women earning doctoral degrees in science and engineering rose from 744 in 1965 to 2,590 in 1974, an increase of almost 250 percent. Moreover, whereas in 1965,7 percent of all science and engineering degrees were awarded to women, by 1974 the figure had risen to 14 percent.

In professional life, according to Zinberg, women have been substantially assisted by "affirmative action" legislation, which specifies that all U.S. organizations that receive Government funding or do business with the Government must file an annual report on the number of women and minority members they employ. If women and minority members are underrepresented in an organization, the employer must set timetables and goals for correcting the imbalance and must make every effort to avoid future discrimination. As a result of the legislation qualified women have been recruited in many universities and organizations to serve on policymaking committees and boards. In 1975 the American Physical Society and the American Association for the Advancement of Science each elected a woman president; a year later the American Chemical Society did the same. Zinberg writes: "It is quite clear that men, by which I mean the older generation of scientists in either totally male or predominantly male organizations, are more aware of the existence of women scientists. The women, in turn, have gained self-confidence, and affirmative action has enabled women to raise their expectations and sense of entitlement."

In spite of the clear gains women scientists and engineers have made in the past decade, various studies have revealed a less encouraging aspect of the picture. First, in general women scientists and engineers earn substantially less than their male colleagues at the same job level, and the salary differentials seem to be increasing with time. In colleges and universities during the academic year 1974–75 faculty women earned an average of $2,820 less than faculty men; the next year the disparity had increased to $3,096. In industry women with bachelor's degrees in science were offered an average of between $216 and $1,332 less than men science graduates in their initial job proposal. As women progress in their career the salary gap continues to widen: today women biologists with Ph.D.'s have a mean salary that is less than 75 percent of the mean for their male peers.

Second, in general women scientists and engineers are more likely to be found in the less highly rated colleges and universities. More women have been appointed to part-time or untenured positions over the past decade, but the percentage of women with full-time, tenured appointments has increased almost imperceptibly, and in some fields it has decreased. A survey of 184 chemistry departments in American universities revealed that in 1976 only 2 percent of the full-time positions were held by women. In industry only 5 percent of the chemists are women. Yet 20 percent of all bachelor's degrees in chemistry go to women, 22 percent of all master's degrees and 11 percent of all Ph.D.'s.

"In the decade 1966–1976 there has been a continuous rise in the numbers of women studying science and completing Ph.D.'s," Zinberg concludes. "And undoubtedly the lot of women scientists has improved. Yet as the negative findings indicate, the struggle has only begun." Zinberg believes the biggest barrier to the advancement of women in science is the economic environment: at the same time that larger numbers of women are becoming scientists, the growth rate of the scientific enterprise has decreased. "Were the pace of expansion in the 1970's similar to that of the late 1950's and early 1960's," she observes, "many of the problems ... would have disappeared. At present the employment situation is difficult for everyone." Should the barriers discourage young women from pursuing scientific careers? "Hardly," Zinberg writes. "Although [I am] no Pollyanna, I believe that with continued efforts, legislative enforcement, and some good will, a more egalitarian profession will evolve. Career decisions should be based on talent and interest."

Enrichment by Centrifugation

The decision of the Carter Administration to postpone plans for the commercial reprocessing and recycling of spent nuclear fuel and to slow the U.S. program to develop a plutonium-based fast breeder reactor means among other things that this country will have to place increased reliance in the next few decades on nuclear power plants that are based on standard light-water-reactor designs. Such reactors normally consume uranium fuel that is "enriched" to contain about 3.3 percent of the readily fissionable isotope uranium 235. (Natural uranium is composed of .7 percent uranium 235 and 99.3 percent nonfissionable uranium 238.) In order to meet the anticipated demand for more enriched-uranium fuel President Carter announced in his energy address to Congress in April that the U.S. must increase its capacity "to produce enriched uranium fuel for light-water nuclear power plants, using the new centrifuge technology, which consumes only about 1/10th of the energy of existing gaseous-diffusion plants."

The gas-centrifuge approach to the problem of separating uranium isotopes, which Carter has evidently selected to provide the additional production capacity needed for enriched uranium, was first considered as an isotope-separation method in the course of the U.S. atomic-bomb project during World War II. The investigation of the centrifuge option was dropped in 1944, however, in favor of the gaseous-diffusion process, which proved capable of producing large amounts of high-enriched, or weapons-grade, uranium (typically containing at least 90 percent uranium 235). The giant gaseous-diffusion plants built in Oak Ridge, Tenn., and Portsmouth, Ohio, have continued to supply not only high-enriched uranium for military purposes but also low-enriched uranium for power plants in the U.S. and abroad.

Federally supported research on gas centrifugation as a means of separating uranium isotopes was resumed in the early 1960's, largely as a result of preliminary work done at the University of Virginia. By 1968 the results of studies conducted at Virginia and elsewhere were considered sufficiently encouraging to convince the Atomic Energy Commission to begin the design of a facility for testing the reliability of "high capacity" centrifuges working with gaseous uranium hexafluoride (UF_6). The resulting test facility, built at Oak

White rum and Canada Dry® tonic. The making of a "classic."

Through the years the summer heat has inspired an endless variety of drinks, but few have ever attained the rank of "classic."

So it pleases us immensely to observe that white rum and Canada Dry tonic shows all the signs of becoming an enduring favorite.

We base that feeling on the fact that a national taste test has shown us that Puerto Rican white rum is preferred over gin and vodka for taste and smoothness. And that white Puerto Rican rum is the fastest growing distilled spirit in America.

In addition we know that more drinkers prefer Canada Dry than any other tonic.

The two of them together are an unbeatable team—with all the ingredients for the making of a "classic."

The reasons that lie behind these preferences are clear enough. Drinkers appreciate quality.

Breeding shows

White rum, for example, is the beneficiary of a great deal of care and attention. More so than either gin or vodka. Only white rum, of the three, is aged for smoothness, spending a minimum of a year in oak casks, by Puerto Rican law. By the time it emerges, not even the faintest trace of a rough edge remains.

A perfect combination

Canada Dry, for its part, is the perfect partner. Providing its own distinctively smooth taste, created by a unique blending of citrus flavors and quinine.

We suggest that you sample the pleasures of white rum and Canada Dry tonic for yourself. When you taste this new "classic" you'll discover what a tonic drink was meant to be.

PUERTO RICAN RUMS

For free "White Rum Classics" recipes, write: Puerto Rican Rums, Dept. SC-5, 1290 Avenue of the Americas, N.Y., N.Y. 10019

All this. All standard. In our lowest priced car.† The Honda Civic 1200 Sedan.

Our standard basis for comparison:

- Front-wheel drive with transverse mounted engine.
- 1237cc overhead cam 4-cylinder engine.
- Four wheel independent MacPherson strut suspension.
- Dual-diagonal braking system.
- Power-assisted, self-adjusting front disc brakes.
- Rack and pinion steering.
- Full carpeting.
- Reclining front bucket seats.
- Tinted glass.
- Opening rear quarter windows.
- Combination light switch and turn signals on steering columm.
- Trip odometer.
- Two-speed electric windshield wipers with electric washers.
- Coolant recovery system.
- Flow through ventilation.
- Owner tool kit.

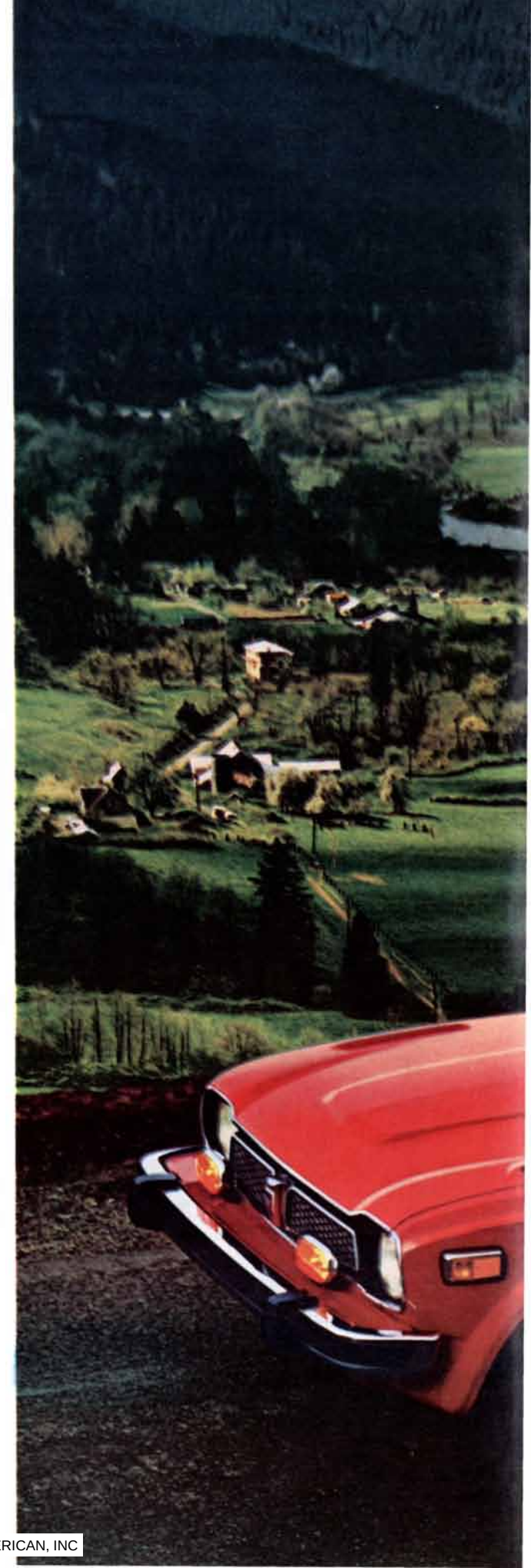
The Honda Civic 1200 Sedan.

Some Honda Models may be in short supply in your area.

Civic 1200 (1237cc) (not available in Calif. and high altitude counties)		1977 EPA Mileage Estimates* Highway	City
Sedan	4-Speed	43	28
Hatchback	4-Speed	43	28
	Hondamatic	29	23
Civic CVCC® (1488cc)			
Sedan	4-Speed	50 (46)	39 (35)
Hatchback	4-Speed	50 (46)	39 (35)
	Hondamatic	37 (34)	32 (28)
5-Speed	Hatchback	54 (51)	41 (34)
Wagon	4-Speed	41 (37)	30 (28)
	Hondamatic	32 (32)	27 (25)
Accord CVCC® (1600cc)			
Hatchback	5-Speed	48 (47)	38 (33)
	Hondamatic	31 (32)	26 (25)

† Manufacturer's suggested retail price excluding freight, taxes, license and options.

*EPA ESTIMATES. The actual mileage you get will vary depending on the type of driving you do, your driving habits, your car's condition and optional equipment. For high altitude models, see your dealer. California estimates shown in parentheses.



Is the best defense a good offense?

Ask the men who fly the world's best fighter: the F-15 Eagle.

To U.S. Air Force pros whose business is national defense, the F-15 Eagle is a winner.

At all altitudes and all speeds, in any weather night or day, whatever the challenge from whatever foe, the best fighter aircraft in the world today is the F-15 Eagle.

F-15 EAGLE
MCDONNELL DOUGLAS

EQUAL OPPORTUNITY IN PROFESSIONAL CAREERS. SEND RESUME: BOX 14526, ST. LOUIS, MO. 63178

Ridge, went into operation in 1971 and has been used to study several different types of high-capacity machines designed to "cascade," or operate in series, and to be potentially competitive with the gaseous-diffusion process.

As a result of these and other studies the Energy Research and Development Administration (ERDA) in February selected one of the two leading designs to serve as the "program machine" for testing in a new advanced test facility, which is expected to be in full operation at the Oak Ridge laboratory by next spring. The machines in the new facility will have a significantly greater capability (measured in separative work units, or SWU) than the earlier machines.

ERDA is also planning to build an even bigger "centrifuge plant demonstration facility" in which a large number of cascading centrifuges will be operated in a configuration that duplicates as closely as possible an actual enrichment-plant layout and operating conditions. Construction of the demonstration facility has already begun at a site adjacent to the existing gas-centrifuge development facilities in Oak Ridge. A full-scale centrifuge plant, with a rated capacity of 8.8 million SWU, has also been announced by ERDA. The plant will be built in Portsmouth, in part to compensate for the cancellation of the Portsmouth Add-on Diffusion Plant.

According to Ewin B. Kiser, Jr., assistant manager for development and planning of ERDA's Oak Ridge operations, the continuing progress in centrifuge technology "clearly supports the view that the centrifuge is a competitive process for an enrichment plant capable of operation in the 1980's." For example, ERDA's estimates of the relative economics of uranium enrichment by means of the Portsmouth Add-on Diffusion Plant and a gas-centrifuge plant of equal size showed that the capital-cost estimate is about 22 percent higher for the centrifuge plant than for the add-on diffusion plant but that the unit cost of separation work by the centrifuge plant is projected to be about 30 percent less than that by the diffusion plant. "With 20-mill power and a fixed charge rate of 11 percent per year," Kiser says, "these estimates indicate that about 45 percent of the cost of separative work from diffusion plants is due to capital recovery and about 53 percent of the cost is due to power.... In the case of the centrifuge process the capital-recovery component represents about 77 percent of the cost of separative work. The remaining 23 percent is primarily operating cost with only about 3 percent due to power costs, since the gas-centrifuge process uses only about 4 percent of the power required by a gaseous-diffusion plant."

One of the main advantages of the gas-centrifuge approach to isotope separation is expected to be its adaptability to an expandable plant design. In contrast to the situation with gaseous-diffusion plants, increased capacity can be added to an existing gas-centrifuge plant in comparatively small increments in order to keep pace with the rising demand for enriched uranium.

Resonances of Uranus

The unexpected discovery that Uranus is surrounded by at least five rings, not unlike those of its neighbor Saturn but not nearly as wide, was made with the aid of a 91-centimeter (35.8-inch) telescope carried aboard the Kuiper Airborne Observatory on a flight over the Indian Ocean in March. The original purpose of the flight had been to gather information about the atmosphere and the precise shape of Uranus by accurately recording the light curve of a ninth-magnitude star (SAO 158687) as it was occulted by the planet and then reappeared some 25 minutes later. An aircraft was used both because it could fly above obscuring clouds and because it was not certain that the narrow path of the occultation would intersect the location of any land-based observatories. Four of the planet's rings interrupted the star's light so briefly—for barely a second—that their width can be no more than about 10 kilometers. The fifth ring seems to be somewhat broader, perhaps 50 to 100 kilometers. In the letter to *Nature* in which they reported their discovery, J. L. Elliot, E. Dunham and D. Mink of Cornell University asked a central question: "Why are the rings of Uranus so narrow—in contrast to the broad rings that surround Saturn?" (The bright central ring of Saturn is about 26,000 kilometers wide.)

A theory to account for the narrowness of Uranus' rings was not long in emerging. Less than three weeks after getting the discovery letter *Nature* received a possible explanation from S. F. Dermott and Thomas Gold, also of Cornell. Dermott and Gold pointed out that in the case of Saturn it is necessary to explain the separations between the three principal rings and the thin gaps in the two outer rings. The generally accepted explanation is that the separations and the gaps represent lanes that are continuously swept free of particles by resonances between the orbital motions of the particles in the rings and the orbital motions of the largest of the 10 satellites of Saturn that travel around the planet outside the rings.

Reversing this explanation, Dermott and Gold calculate that the three most massive of Uranus' five satellites—Ariel, Titania and Oberon—are capable of setting up resonances that would force particles circulating around Uranus, in cooperation with the pressure of sunlight on the particles, to occupy six narrow orbits at distances agreeing closely with those measured during the occultation of SAO 158687. Dermott and Gold suggest that the broad outermost ring is actually two closely spaced rings. This ring occulted SAO 158687 for about seven seconds before the star disappeared behind the planet, giving rise to two closely spaced minimums in the star's light. The ring is peculiar, however, in that it occulted the star for only about three seconds after the star had reappeared from behind the planet. The discovery of the rings was facilitated because at the time of the occultation the rotational axis of Uranus was pointing roughly in the direction of the earth. Since the rings lie in a plane perpendicular to that axis, they presented themselves as a series of concentric circles.

Reprieve from Drowning

Few events are more dangerous to life than an interruption of breathing or blood circulation that interferes with the central process of animal metabolism: the respiratory gas exchange that brings oxygen to the tissues and removes carbon dioxide. Although the skin, the muscles and the gut can survive without circulation for more than an hour, the heart and the brain are acutely sensitive to asphyxia. Suffocation or heart failure can kill a human being within a few minutes, and the human brain suffers irreversible damage if it is deprived of oxygen for more than four minutes.

It is not surprising, then, that the body responds with heroic measures to the threat of asphyxia. When a man's face is immersed in cold water, a reflex is activated that immediately slows the heartbeat and constricts the peripheral arteries, shunting the oxygen-containing blood away from the extremities and the gut to the heart and brain. Called the diving reflex, this physiological response was first identified in birds and mammals that are specialized to stay under water for substantial periods without breathing [see "The Master Switch of Life," by P. F. Scholander; SCIENTIFIC AMERICAN, December, 1963]. The diving reflex has since been found in all vertebrate animals in response to the threat of asphyxia from any one of a number of causes. Only recently, however, has the importance of the human diving reflex to clinical medicine been appreciated.

Martin J. Nemiroff, a physician at the University of Michigan Hospital, has concluded that the diving reflex can enable many people to survive submersion in cold water for long periods. Although human beings soon become unconscious after involuntary submersion, the diving reflex apparently allows the heart to maintain a feeble but constant supply of blood to the brain. In cold water the oxygen need of the tissues is reduced, extending the possible time of survival

without external oxygen to a period of as long as an hour.

Nemiroff based his conclusions on 50 near-drowning incidents in Michigan waters, 13 of which involved submersion in water below 70 degrees Fahrenheit. Of these cold-water cases nine were successfully resuscitated without brain damage or other ill effects, two died of lung infections contracted in the water and two suffered brain damage. All had gone beyond the four-minute oxygen-deprivation limit after which, it has been thought up to now, irreversible brain damage must occur.

The record for survival without injury after lengthy submersion is held by an 18-year-old Jackson, Mich., college student. In March, 1975, he was trapped in his car after it had broken through the ice of a frozen pond and rolled over. The student later recalled struggling, inhaling water and losing consciousness. (The fact that the car had rolled over eliminated the possibility that it might hold a pocket of air for him to breathe.) When he was brought to the surface after 38 minutes in the icy water, he showed no life signs and was declared dead at the scene. As he was being put into an ambulance for removal, however, he emitted an agonal gasp: a kind of involuntary belch often associated with dying. Resuscitation efforts were then begun immediately. After a high-speed ambulance trip to the University of Michigan Hospital, two more hours of resuscitation efforts and 13 hours of respiratory support, the student regained consciousness and was able to recognize his mother. Two weeks later he left the hospital and returned to college. Another survivor in the study was a physician who was submerged for 15 minutes in a lake near Ann Arbor, Mich. After resuscitation he showed no mental impairment, and he has since resumed his practice.

The key factors for surviving submersion without permanent injury appear to be the length of time spent under water, the temperature of the water, the age of the individual (the diving reflex is more pronounced in children) and the promptness of the resuscitation efforts. Since the protective effect of the diving reflex is lost as soon as the individual is taken out of the water, resuscitation should be started immediately at the water's edge. In the Michigan study one fatality was submerged for only four minutes, but his resuscitation was delayed between 10 and 15 minutes until he reached the hospital.

A similar study of drownings and near-drownings in Florida showed that in waters warmer than 70 degrees F. four minutes seems to be the limit for submersion without death or brain damage. In Michigan and many other states, however, the temperature of many bodies of water, particularly in deeper areas, is well below 70 degrees. Nemiroff's findings clearly have important legal and moral implications for physicians and rescue workers. They also imply that many people who have been declared dead from drowning could have been resuscitated if rescue workers had known it was possible.

From Reckoning to Writing

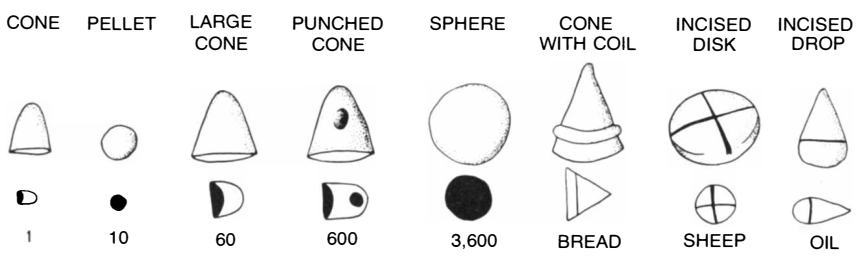
Why was writing invented? In the case of two relatively late examples, the undeciphered Linear A script of Crete and its successor, Linear B, the principal motivation was apparently the keeping of records. The tablets written in both systems bear numbers, weights, measures and the names of various commodities. The oldest-known writing—Sumerian tablets dating back to between 3500 and 2900 B.C.—also seems to have had record keeping as its major purpose. Now a scholar of Middle Eastern records and writing suggests that certain geometric clay objects found at sites from western Turkey to eastern Iran, particularly in Mesopotamia, served as record-keeping tallies 5,000 years earlier and were the precursor of Sumerian writing.

Writing in *Syro-Mesopotamian Studies,* Denise Schmandt-Besserat of the University of Texas at Austin observes that large numbers of small objects made of fired clay in a variety of shapes, many of the objects a centimeter or less in the largest dimension and few larger than 2.5 centimeters, have been uncovered at Middle Eastern archaeological sites. For example, at Jarmo in Iraq many such objects were found; 106 were cone-shaped, 206 disk-shaped and 1,153 spherical. Among other shapes were tetrahedrons, pyramids, incomplete spheres and double cones joined at the base. At Susa in Iran the temple archives included fist-sized clay balls that rattle when they are shaken; broken open, they proved to contain numerous smaller clay objects, most of them cones, disks, spheres and tetrahedrons.

Schmandt-Besserat, working in the field and with museum collections, noticed a correlation between the geometric objects and certain arbitrary signs (as opposed to obvious pictographs) on Sumerian tablets. For example, she equates small cones with a conical indentation signifying the number 1 made by pushing the blunt end of a stylus into wet clay at an angle, and small spheres with a circular indentation signifying the number 10 made by pushing the blunt end of the stylus straight down. Larger cones match a larger conical indentation (the number 60); punched cones, a conical indentation together with a punch mark (the number 600). Spheres more than 2.5 centimeters in diameter match larger circular indentations on the tablets (the number 3,600). Some of the geometric objects might therefore have served as numerical tallies. What about the others?

An incised triangle with a line drawn parallel to one side is a character, designated No. 535, that appears on the tablets discovered at the Sumerian site of Warka in southern Iraq. It stands for bread. Parallel to this character among the geometric objects is a cone with a thin clay coil along the base. Character No. 761 (sheep) is a circle enclosing a cross; its parallel among the geometric objects is a disk with an incised cross. Character No. 733 (oil) is a semicircle joined to a triangle; its parallel among the geometric objects is a teardrop with an incised line around the hemispheric end. Sixteen additional parallels between Sumerian tablet characters and the geometric forms have now been noted by Schmandt-Besserat.

The transition from the use of three-dimensional tallies to the development of writing Schmandt-Besserat sees coming about as follows. At about 8500 B.C. the increasingly settled way of life in the Middle East brought into existence a system of recording transactions by means of clay tallies, some with numerical values and others representing different commodities. Several millenniums later, at about 3500 B.C., "bills of lading" came into use: a series of tallies enclosed in a clay ball, a tamperproof document accompanying each shipment of goods. Not long thereafter, in order to bypass the need to open each clay bill of lading, two-dimensional symbols representing the enclosed geometric forms were impressed on the surface of the ball. From this stage it was a short step to making the bill of lading in tablet form, using the two-dimensional symbols and abandoning the geometric tallies altogether. In short, reckoning had become writing.



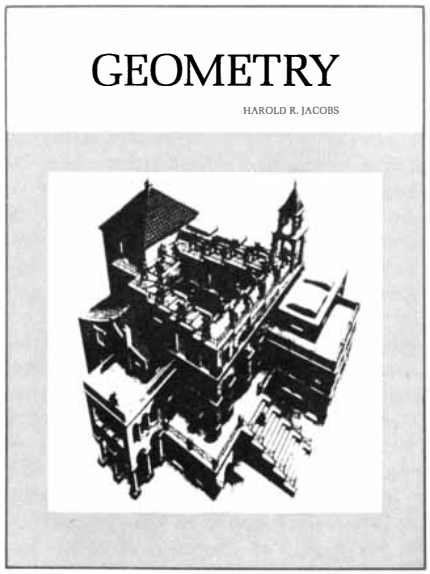
Sumerian signs **(bottom)** *are compared with geometric objects 5,000 years older* **(top)**

"This is the most clever geometry book I have ever seen!"
—Thomas E. Lepperd, Evanston Township High School, Evanston, Illinois

"Another typical Jacobs masterpiece!"
—Al Bigger, Auburn High School, Auburn, Alabama

"This book may single-handedly save geometry in the high schools."
—Patrick J. Boyle, Santa Rosa Junior College, California

Strong words about a strong book



GEOMETRY
1974, 701 pages, 1100 illus.,
ISBN 0-7167-0456-0, $9.00
Teacher's Guide $5.00
Transparency Masters $40.00
Test Masters $2.50

Also by Harold Jacobs
MATHEMATICS
A Human Endeavor
A Textbook for Those Who Think They Don't Like the Subject
Adopted by more than 2000 high schools
Foreword by Martin Gardner
1970, 529 pages, 634 illus.,
ISBN 0-7167-0439-0, $8.50
Teacher's Guide $5.00

"Einstein could have used Mr. Jacobs' book and still not have become so bogged down in the run-of-the-mill education that he could have missed the theory of relativity. Remember, Einstein said 'common sense is the collection of prejudices acquired by age 18'—no prejudices would be acquired by the geometry student and quite a few would be eradicated through this text."
—James J. Dolan, Chaminade Preparatory School, Canoga Park, California

"This is a book that belongs in every geometry class in the country. It has something to offer students at every level, combining wit and imagination with some very solid mathematical principles."
—Diana Brown, Wheeler High School, Marietta, Georgia

"This is a fascinating book with much more than the usual theorems. The cartoons and photographs make this a geometry book that is enjoyable reading. Most important, the problems are very good—they make one *want* to think and understand. . . . Congratulations, Freeman and Jacobs have done it again."
—Mark Saks, Community College of Philadelphia, Pennsylvania

The underlying theme of *Geometry* is a simple one: the principles of geometry may be more easily learned when they are shown to have immediate, practical applications in the student's world. *Geometry* is an exciting blend of humor, clear explanations, and an uncompromising attention to basic concepts—a blend that really does inspire students. We think that once you've examined this book, you'll have some strong words of praise too.

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

The Flow of Heat from the Earth's Interior

A global heat-flow map can be drawn on the basis of thousands of individual field measurements on continents and ocean floors. The heat-flow pattern is interpreted in terms of plate tectonics

by Henry N. Pollack and David S. Chapman

"Straight through the dim and open portal we entered unopposed, and I, eager to learn what part of Hell's bowels those burning walls enclosed, began to look about." So did Dante write of his descent into the Inferno. Miners of later centuries might well have believed he was describing their daily working environment, since it has been widely observed that without proper ventilation and cooling a mine gets hotter with depth. There are of course many other indications that the earth's interior is hot, the most obvious being an erupting volcano. Only slightly less spectacular are the handful of areas around the world that display hydrothermal activity, such as the hot springs, steam vents and geysers of Yellowstone National Park. One of the fundamental axioms of physics, embodied in what is known as Fourier's law of heat conduction, is that heat flows from the warmer parts of a body to the cooler ones. It can therefore be inferred that since the temperature increases with depth in the earth's crust, there is a flow of heat outward from the earth's interior.

The transfer of heat within the earth and its eventual passage to the surface by conduction through the crust play a fundamental role in all modern theories of geodynamics. In the 19th century the earth's internal heat also figured significantly in the protracted debate over the age of the earth between William Thomson (Lord Kelvin) and several of his scientific contemporaries. Kelvin's dissertation at the University of Glasgow in 1846, titled "Age of the Earth and Its Limitations as Determined from the Distribution and Movement of Heat within It," was the first of a long series of papers in which he laid out the argument that the earth's thermal gradient (the rate at which the temperature increases with depth) would continue to diminish with time as the earth cools following its formation and solidification from molten rock. By determining the earth's thermal gradient from measurements in mines and boreholes, he maintained, one could tell how long the earth had been cooling and so could determine the age of the earth. Records of temperatures at various depths in mines could be found in mining journals, and Kelvin supplemented them with measurements of his own in Scotland. His conclusion was that the temperature, at least to modest depths below the surface, increased at a rate of between 20 and 40 degrees Celsius for every 1,000 meters of depth. To Kelvin this relation indicated that the earth had been cooling for only a few tens of millions of years, a period far shorter than many geologists and biologists of the time thought necessary for the development of the known stratigraphic and fossil record. The ensuing debate spanned half a century and pitted Kelvin against such prominent evolutionists as Charles Darwin and Thomas Huxley.

Kelvin's calculation was based on the assumption that the heat being lost by the earth was drawn from the reservoir of heat left over from the earth's originally molten condition. That assumption, which was essentially unchallenged for decades, was to be the undoing of all estimates of the earth's age based on the measurements of its heat. Three observations made Kelvin's estimate of the age of the earth based on its initial heat no longer tenable: the discovery of radioactivity by Henri Becquerel in 1896, the observation by Pierre Curie in 1903 that the radioactive decay of certain isotopes liberates heat and the confirmation by Robert Strutt in 1906 that common rocks found in the earth's crust contain sufficient amounts of radioactive isotopes to yield a significant fraction, if not all, of the earth's observed heat flow.

Measuring Heat Flow

How much heat is the earth losing today as a result of conduction from its interior? The global average is close to .06 watt per square meter of surface, or about 30 trillion watts over the entire planet. The amount of energy arriving from the sun is almost 6,000 times greater, and it is completely dominant in establishing the temperature of the earth's surface. The flow of heat from the interior is scarcely a trickle; the heat conducted through an area the size of a football field is roughly equivalent to the energy given off by three 100-watt light bulbs. The evolution of the earth covers vast reaches of time, however, and a trickle of energy over aeons can do significant geological work, such as making continents drift, opening and closing ocean basins, building mountains and causing earthquakes. The geographic variation in the flow of heat from the earth's interior is not great: most measurements lie within a factor of three around the mean value. The patterns of heat flow in continental regions differ from those in oceanic ones, but the average heat flow through both

DRAMATIC EVIDENCE of the power of the earth's internal heat to mold the geology of the surface is provided by this photograph of the 1971 Mauna Ulu eruption on Kilauea volcano in Hawaii. The photograph was made by Wendell A. Duffield of the U.S. Geological Survey as he was standing on the rim of the Mauna Ulu crater and looking almost straight down. The bluish gray background is the comparatively cool crust of partly solidified basaltic lava that forms on the surface of the hot liquid-lava "lake" below. The jagged orange streaks are cracks through which molten lava is upwelling. The entire scene, Duffield points out, is analogous to the spreading of new sea floor from mid-ocean ridges visualized in the plate-tectonic model.

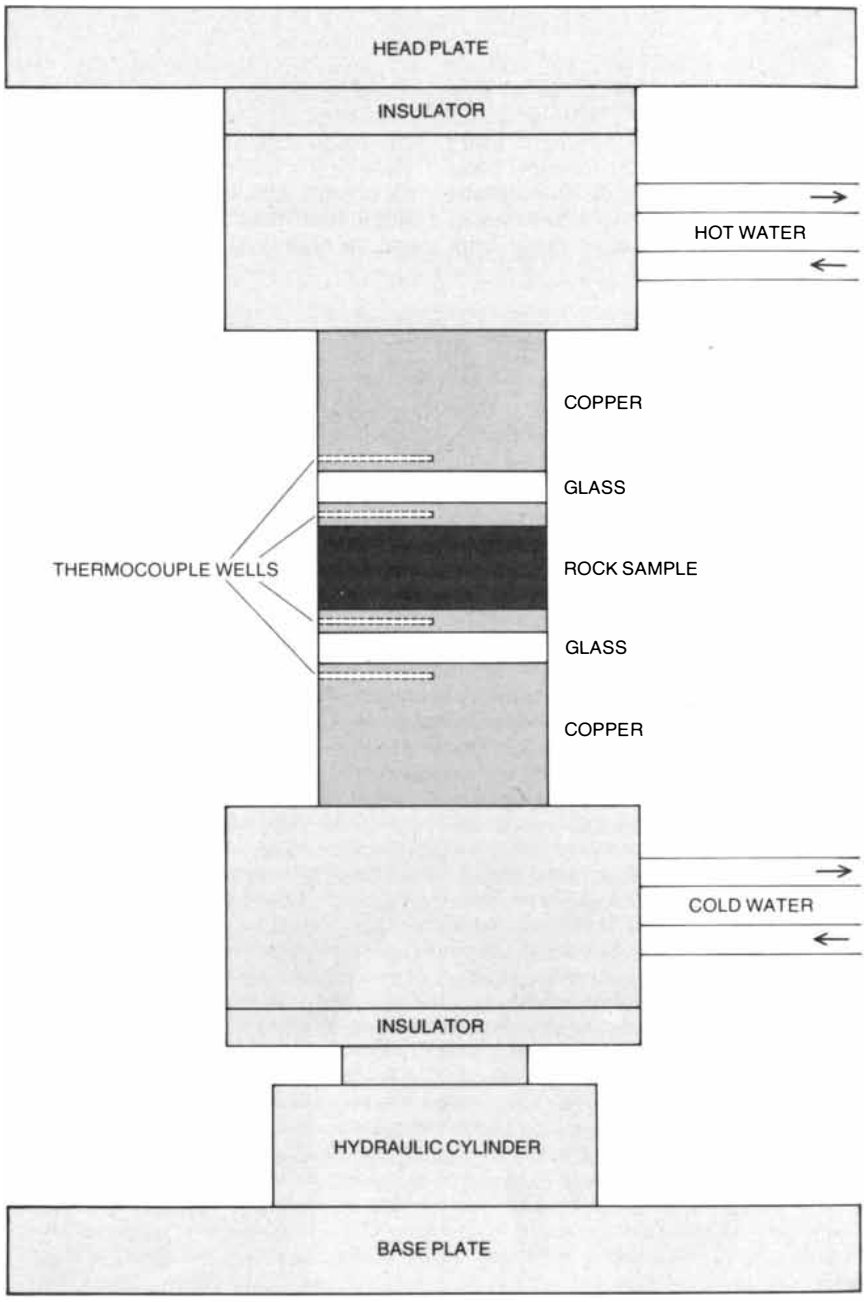
DRILL RIG IN ZAMBIA, originally set up to bore into the earth in search of copper, provided one of several "holes of opportunity" used by the authors and their co-workers as part of their program to obtain heat-flow measurements in areas of Africa and South America where the existing data are sparse. The earth's temperature is taken by lowering an electrical-resistance thermometer (called a thermistor) down the borehole, making measurements at several depths. The records are used to establish the rate at which the temperature of the rock increases with depth, a local quantity known as the geothermal gradient. The Zambian heat-flow measurements were carried out four years ago while the authors were on leave from the University of Michigan and were based at the University of Zambia. More than 50 such drill holes were surveyed at eight different Precambrian geological sites in the country. The results of the survey were interpreted by the authors as indicating the presence of anomalously warm material only a few tens of kilometers below the surface of the earth.

is surprisingly similar. Some areas, such as Iceland, exhibit an extraordinary heat flow, and geothermal areas of this type can be tapped as an energy resource.

If heat is being transported through the earth's crust by thermal conduction, the amount of heat in transit is equal to the product of the temperature gradient times the thermal conductivity (a property of the rock that describes how easily it transmits heat). Any experimental study of the earth's heat flow is concerned with measuring these two quantities. On continents temperature gradients are measured by lowering sensitive electronic thermometers—thermistors—down drill holes or by measuring the temperature of the rock at different levels in mines. The process of drilling a hole disturbs the thermal equilibrium at the site; hence several weeks or months are allowed to lapse between the drilling and the measuring. Even after the disturbance has become negligible, subsurface temperatures are disturbed by such effects as the daily and annual fluctuation in the surface temperature, unevenness in vegetative cover, unevenness in topography, the movement of ground water, the uplift or erosion of the surface and variations in climate. Most of these disturbances diminish to an acceptable level beyond depths of a few tens of meters; some, however, can extend to several hundreds of meters. Although reliable heat-flow measurements can sometimes be made in holes as shallow as 50 meters, most workers who make such measurements prefer to do so in holes that are 300 meters or more in depth.

On the ocean floor, where sediments are comparatively soft and the blanket of seawater provides an environment of almost constant temperature, the drilling of a hole is unnecessary. There temperature gradients are determined by plunging a long cylindrical probe several meters into the soft sediment and measuring the temperature at one-meter intervals with fixed thermistors.

For measurements of thermal conductivity two methods are widely used. For hard continental rock a sample from the drill hole is cut and polished in the form of a disk and is inserted into a column between silica disks of known conductivity. A constant temperature difference maintained between the ends of the column causes a flow of heat through the sample and the silica standards, and the measurement of the relative drop in temperature across the components of the stack yields the thermal conductivity of the sample. For softer continental rocks and marine sediments a thin needle is inserted into the sample and is heated along its length. From a record of the rise of temperature with time the thermal conductivity can be calculated easily.



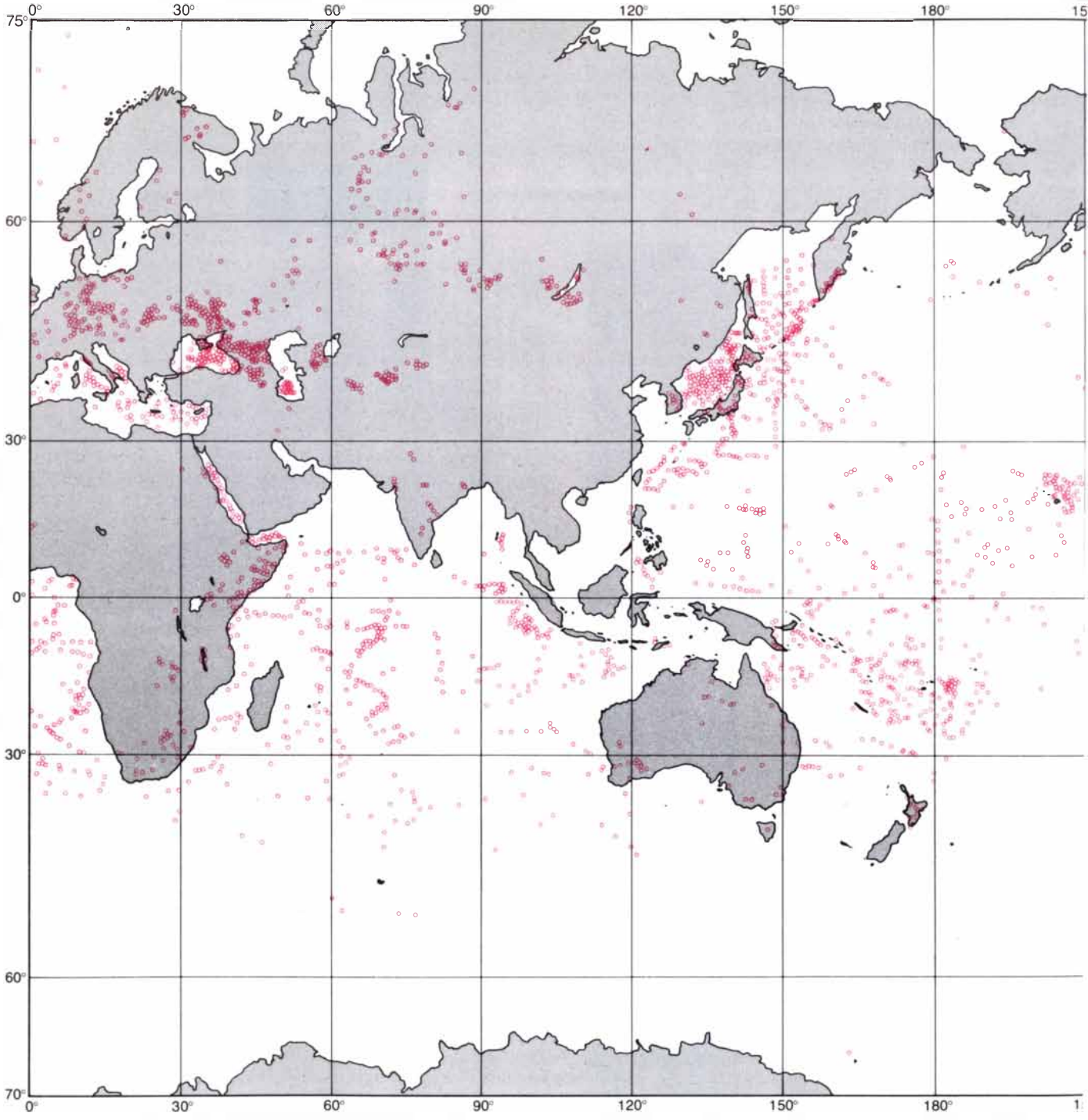
THERMAL CONDUCTIVITY of a hard rock sample obtained from a drill hole is measured in the apparatus depicted here. The sample, cut and polished in the form of a disk, is inserted in a column between silica disks of known conductivity. Constant temperature differences maintained between the ends of the column cause a flow of heat through the sample and the silica standards. The sample's thermal conductivity is determined by measuring the relative drop in temperature across the components of the stack. The heat flow at the drill site is equal to the product of observed local temperature gradient times thermal conductivity of the rock.

More than 5,000 such heat-flow measurements have been reported in a recent compilation by Alan M. Jessop of the Canadian Department of Energy, Mines and Resources and by John G. Sclater and Michael A. Hobart of the Massachusetts Institute of Technology. Although the number of measurements is sufficient for several types of regional analysis, the data set is still geographically uneven: more than twice as many measurements have been made at sea as on land. The middle-latitude oceans, North America, Europe and Australia are quite well surveyed, whereas large areas of the high-latitude oceans and of South America, Africa, Asia and Antarctica have no measurements at all. In the past four years our group from the University of Michigan has helped to remedy some of this geographic imbalance on the continents by conduct-

ing heat-flow measurement programs in Zambia, Niger and Brazil.

An analysis of the global data without regard for specific site location or geologic setting indicates a fairly wide distribution of results asymmetrically spread around a modal (or most commonly observed) value of 50 milliwatts per square meter [*see top illustration on page 66*]. Individual values range from near zero to several hundred milliwatts per square meter, the latter being located mainly within geothermal areas associated with the worldwide system of mid-ocean rifts. Subdividing the data into continental and oceanic regions reveals similar modal values for both sets; the oceanic data, however, have a wider distribution than the continental data, and the high level of asymmetry forces the means, or average values, for each of these regions well above the modes.

This gross grouping of heat-flow measurements has been useful in the past, and the similarity between continental and oceanic measurements has served to stimulate much discussion among geophysicists and geologists who had expected quite different results in the two settings. As with many other aspects of

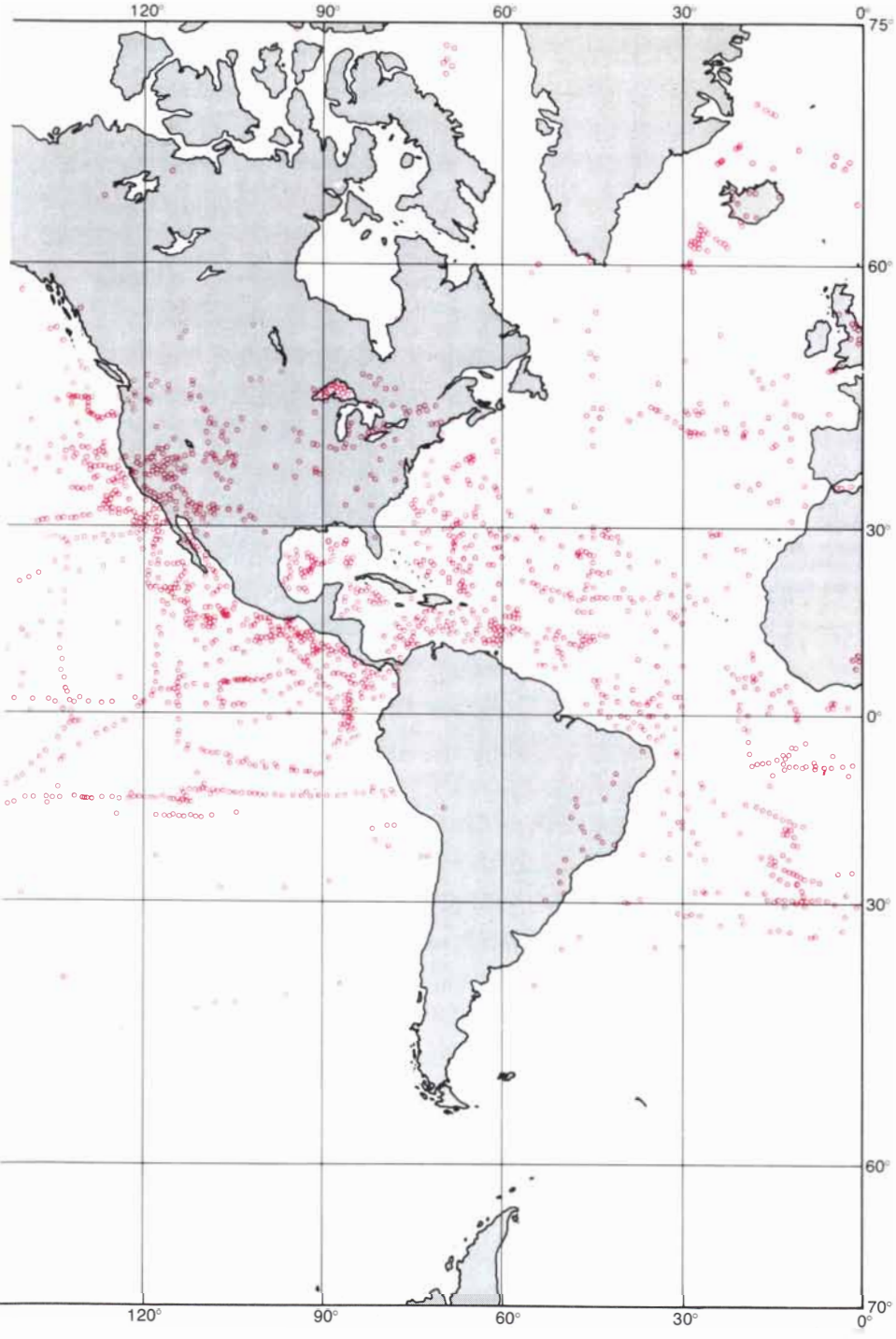


UNEVEN GEOGRAPHIC DISTRIBUTION of oceanic and continental heat-flow measurements (*colored circles*) is evident on this world map, which is based on one prepared by the National Geophysical and Solar-Terrestrial Data Center. The number of such measure-

earth science, however, heat-flow observations now find a compelling new interpretation in terms of the concepts of sea-floor spreading and plate tectonics.

Heat Flow and Plate Tectonics

According to plate tectonics, the lithosphere, the outer shell of the earth, is made up of a dozen or so rigid plates that are being moved about on the earth's surface. Wherever plates are moving apart the gap is filled by hot material flowing upward from the earth's underlying mantle. This material accretes to the edges of the separating plates; the accreting edges form the mid-ocean ridges. The new rock cools as it moves away from the ridge. Across the plate from a ridge one usually finds a great oceanic trench, which marks a site where older and cooler plate is subducted, or recycled back into the interior. Frictional and conductive heating of the plate in the subduction zone melts part of it, and the melted fraction rises buoyantly to the surface to form the volcanoes and island arcs typically arrayed behind the trenches. Such subduction processes, together with other forms of plate interactions, give rise to thermal metamorphism, the generation of volcanic magma and mountain building on continents.

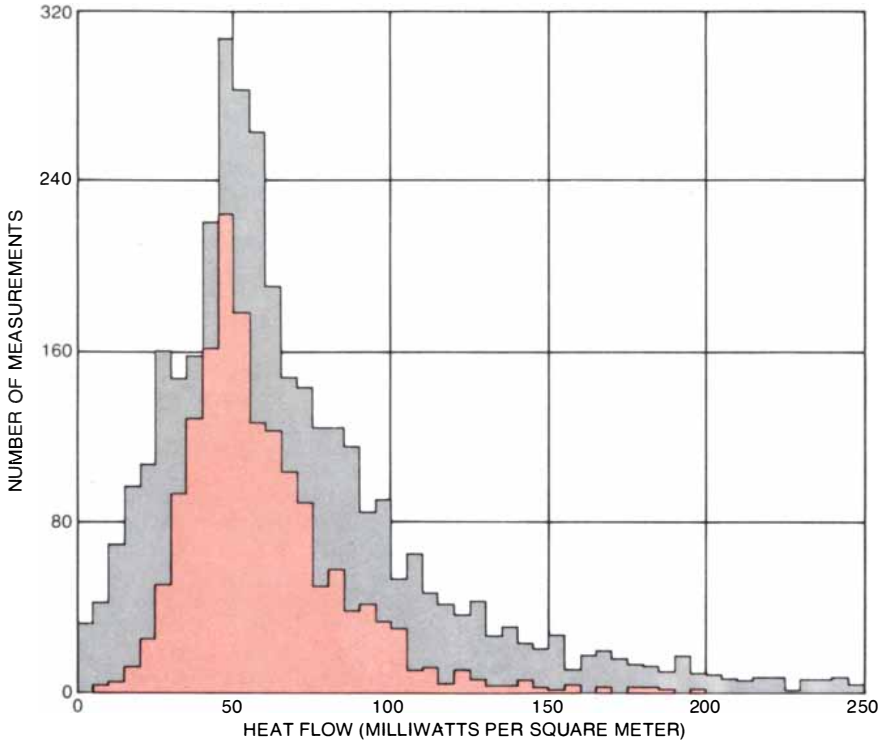


ments has increased rapidly in recent years: from 47 in 1954 to 1,162 in 1962 to more than 5,000 at present. More than twice as many measurements have been made at sea as on land.

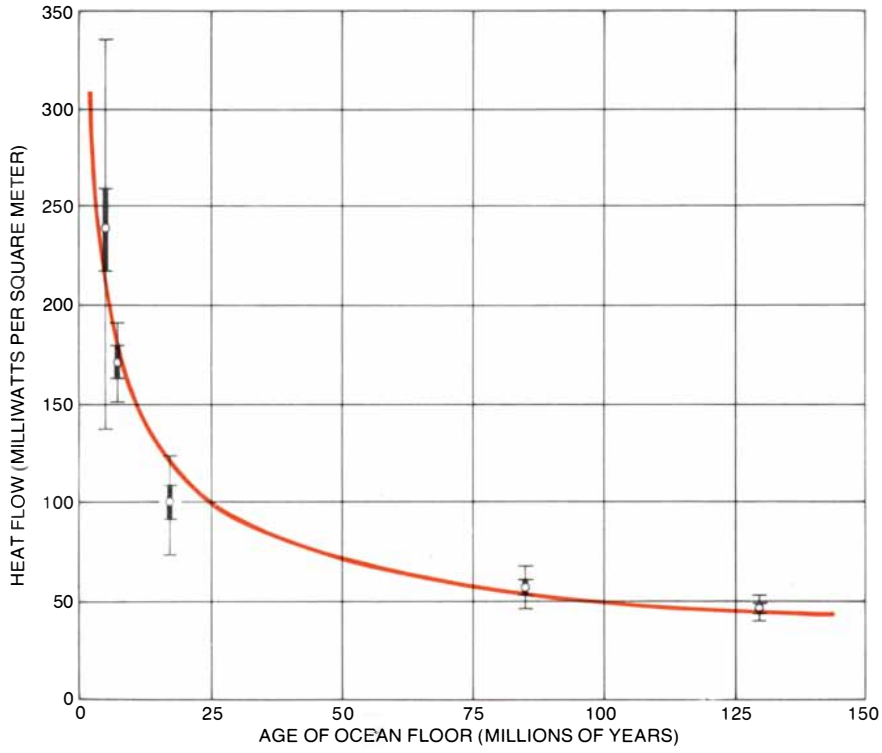
One first looks to the oceanic plates with their comparatively simple geology to obtain evidence for the thermal model of plate evolution. Edward C. Bullard of the University of Cambridge, who reported the first marine heat-flow measurements for the Atlantic in 1954, noted at that time the near-equality of the mean heat flow from the continents and the ocean floor. Today, with the addition of some 3,500 measurements in ocean floor of all ages, it is possible to see a systematic decrease of heat flow with increasing age and hence depth [*see bottom illustration on next page*]. For those sites where a thick, impermeable cover of sediments prevents the removal of heat by seawater circulating through the fractured oceanic crust, heat-flow measurements agree extremely well with predictions based on mathematical models of a cooling plate.

Such models of plate cooling also explain the broad topographic features of the ocean floor. The newly formed crests of mid-ocean ridges are typically 1,000 to 3,000 meters below sea level, whereas the oldest ocean basins are 5,500 meters below sea level. Thus in 200 million years the sea floor subsides by about 3,000 meters. The reason is that the recently accreted rock is hot and thermally expanded, whereas the older material has cooled and contracted. The match between the observed topography and the theoretically calculated topography is so good that it seems unlikely the subsidence will be explained in any way other than in terms of a simple cooling model.

The concept of the thermal evolution of an oceanic plate may also provide an answer to a long-standing puzzle in geology: What gives rise to transgressions of the sea onto continents? The Upper Cretaceous period was one such time of great marine transgression. Starting about 100 million years ago the sea level rose; it crested between 90 and 70 million years ago and withdrew from the continents about 60 million years ago. From the spacing of magnetic lineations on the ocean floor it can be shown that the Upper Cretaceous was also a period of rapid sea-floor spreading. Since the cooling and subsidence of an oceanic plate is time-dependent, an increase in



STRIKING SIMILARITY is seen in the asymmetrical distribution of heat-flow values for both continents (*color*) and oceans (*gray*). Most of the observed values are in the range between 20 and 120 milliwatts per square meter, with the global average being at about 60. The modal, or most commonly observed, values for continents and oceans (*peaks of profiles*) are both closer to 50. It is not known at present whether the near-equality of the continental and the oceanic heat-flow measurements is a fundamental characteristic of the movement of heat within the earth or is merely a coincidence arising from incomplete sampling. The data for this chart were compiled by Alan M. Jessop of the Canadian Department of Energy, Mines and Resources and by John G. Sclater and Michael A. Hobart of the Massachusetts Institute of Technology.



HEAT FLOW DECREASES with increasing age of ocean floor, as indicated here by the distribution of measured heat-flow values for five selected areas on the floor of the Pacific Ocean. Data points give the mean value for each age grouping. Heavy bars show the probable error; light bars show standard deviation about the mean. Measurements agree extremely well with a theoretical estimate of heat flow expected from a cooling plate of oceanic crust (*colored curve*).

the spreading rate would have broadened the oceanic ridge and increased its volume. This in turn would have reduced the water capacity of the ocean basins and displaced the sea onto the continents. The subsequent regression was apparently caused by a reduction in the rate of sea-floor spreading that began about 85 million years ago.

Above subduction zones the heat-flow patterns are more complex, but they nonetheless provide important clues to the subduction process. A pattern generally observed at subduction zones, and particularly well documented for the Japan arc system, is one of low heat flow near the oceanic trench and very high heat flow to the landward side of the island arc [*see illustration on opposite page*]. The pattern suggests that the top part of the cool subducting plate acts as a heat absorber, causing the band of low heat flow observed adjacent to the trench. Deeper in the subduction zone frictional and conductive heating are sufficient to melt part of the plate, yielding as a product the volcanic island arc itself and the augmented heat flow behind the arc.

The western U.S. provides another example of an elongated zone of low heat flow adjacent to a region of magmatic activity and high heat flow [*see illustration on page 68*]. Here, however, there is at present no active major subduction zone nearby. David D. Blackwell of Southern Methodist University and others have suggested that this zone represents a fossil heat-flow pattern, established some tens of millions of years ago when subduction was active along the boundary between the Pacific plate and the North American plate.

Continental heat flow in areas removed from plate boundaries also falls into recognizable patterns. Measurements on continents now number about 1,700, and from these data one can draw two major conclusions. First, there is a general decrease in heat flow with the increasing age of a geologic province [*see top illustration on page 73*]. This result is similar to that for oceans, but the time scale is apparently quite different. Whereas oceanic heat flow drops below 50 milliwatts per square meter after 100 million years of cooling, on continents one finds such heat flow in geologic settings four or five times older.

The second major result is that for large areas of continents there is a clear relation between surface heat flow and the radioactivity of the surface rocks. That continental rocks, granites in particular, generate significant quantities of heat by the spontaneous disintegration of radioactive elements has been known since early in this century. In 1968 A. Francis Birch, Robert F. Roy and Blackwell, all then working at Harvard University, demonstrated that when heat-flow measurements are plotted with respect to radioactive heat generation for

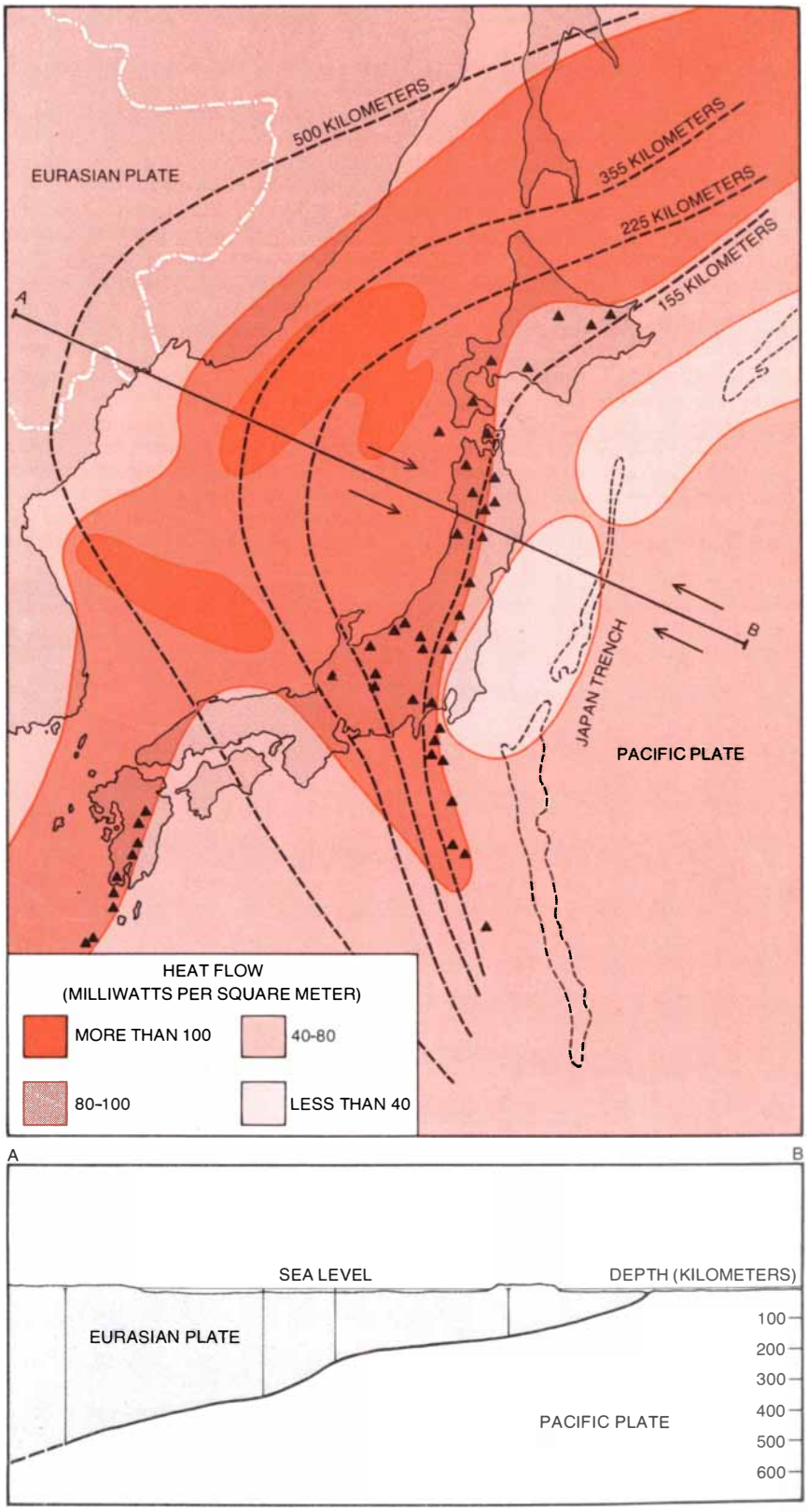
the rocks at various sites, the plotted values fall along a straight line [*see bottom illustration on page 73*]. Different lines were obtained for the eastern U.S., the "basin and range" geologic province of Nevada and Utah, and the Sierra Nevada region, but within each region a linear relation holds. This finding implies that for a given region the heat flow at the surface has two components: a crustal component that varies from site to site according to the local radioactivity and a deeper component that originates in the earth's mantle and is uniform for all sites in the region. Since 1968 this relation has received much attention. Arthur H. Lachenbruch of the U.S. Geological Survey has looked into its consequences for the distribution of heat-producing isotopes in the earth's crust, and he has explained why the concentration of such isotopes should be expected to diminish exponentially with depth.

The variation of the mantle-derived component of the observed heat flow between different provinces has been less well studied, but on the basis of the limited data available we have speculated that in most heat-flow provinces there is a regular partitioning of the heat flow, with about 40 percent of the mean surface flux coming from within the zone of crustal enrichment and 60 percent coming from below. This partitioning, if confirmed, suggests that the average heat production of the continental crust should vary inversely with its age, because in general the older provinces display less heat flow. Such a relation can be explained with a model in which radioactivity diminishes with depth, on the assumption that the older geologic provinces have been eroded to greater depths than the younger ones.

Another intriguing finding has been reported recently by Tom Crough of Stanford University and V. M. Hamza of the University of São Paulo. They show that when one subtracts the heat-flow contribution originating in the zone of crustal enrichment, the remaining heat flow continues to show an inverse dependence on the age of the province, but the time scale of this residual heat flow appears to be one of a simple cooling process, much like what is observed in the oceans. The cooling on the continents is apparently much further along, however, and it must have penetrated more deeply there. Could it be that we are seeing residual heat from a tectonic event 600 million years ago? If we are, the implication is that such events must involve at least the outer 500 kilometers of the earth in order for any residual heat to be making its way to the surface today.

Global Heat Flow

Let us now turn our attention to the broad features of the thermal field of the

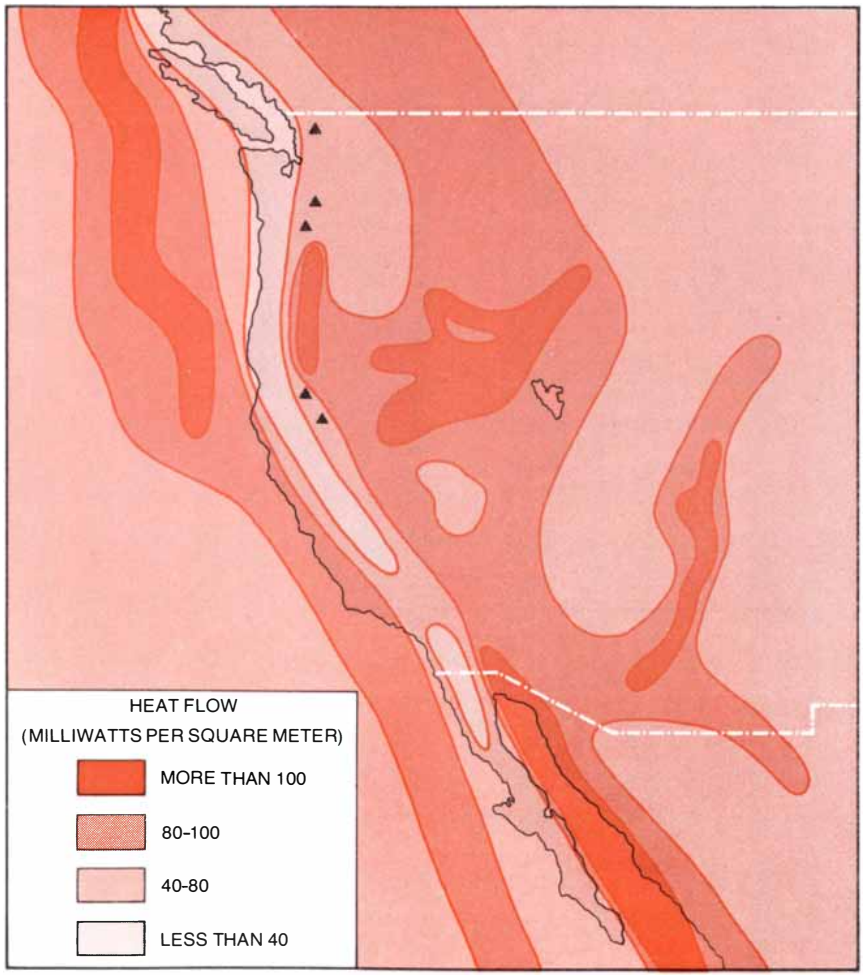


DISTINCTIVE HEAT-FLOW PATTERN is produced by the subduction of the tectonic plate underlying the Pacific Ocean as it dives under the islands of Japan. The arrows in the map at the top indicate the relative convergence of the Pacific and the Eurasian plates, and the broken lines show the depth to the subducted slab. The corresponding depths are also indicated in the cross-sectional diagram at the bottom, which is drawn approximately to scale along the line *AB*. The low heat-flow zone (*lightest color*) observed between the Japan trench and the island arc suggests that from the surface to a depth of about 120 kilometers the cool subducting plate acts as an absorber of heat from the earth's mantle. The volcanoes of the island arc (*black triangles*) and the region of high heat flow (*darkest color*) between Japan and mainland Asia result from frictional heating and partial melting deeper in the subduction zone.

entire earth and in doing so combine results from both the continental and the oceanic regions. Before 1974 several attempts had been made to plot the observed variations in heat flow on a global scale. In spite of the growing number of heat-flow measurements, however, there were still large areas of the globe where no data had been gathered. A mathematical representation of the global distribution of heat-flow measurements was desirable in order to correlate regional heat flow with other geophysical phenomena, such as the earth's gravitational field. The analysis was beset with difficulties because of the need for extensive extrapolation into unsurveyed areas.

By 1974, however, the relation between heat flow and age for both continental and oceanic regions was well established. Could these known correlations not be utilized to make estimates of the probable heat flow in unsurveyed areas and so guide the heat-flow contouring on a global map? They could if geological maps were available showing the ages of all continental and oceanic regions. Although such maps had existed for the continents for some time, it was not until 1974 that Walter C. Pitman, Roger L. Larson and Ellen M. Herron of the Lamont-Doherty Geological Observatory summarized on a single map the detailed ages of all the oceanic regions based on magnetic anomalies on the ocean floor and the recorded reversals of the polarity of the earth's magnetic field. Soon after obtaining the map we divided the entire earth into grid elements five degrees on a side and proceeded to assign to each element a heat-flow value based on the relation of heat flow to tectonic age and the fraction of different age groups present in the grid element. In effect we were creating a synthetic estimate of heat flow in unsurveyed areas. The full data set, comprising observations supplemented by estimates, could then be fitted by appropriate mathematical functions and plotted with a minimum of distortion [*see top illustration on pages 74 and 75*].

The new world heat-flow map constructed in this way showed for the first time on a global scale variations in heat flow that had been recognized in regional surveys. All the major oceanic-ridge systems can be seen as heat-flow highs, as are Alpine Europe, much of western North America and the marginal basins of the western Pacific. The Galápagos spreading center and the Chile Rise appear as bulges on the dominant East Pacific Rise. Regions of low heat flow include all the major continental shields and sedimentary platforms and the oldest oceanic regions.



FOSSIL SUBDUCTION ZONE appears to account for the elongated region of low heat flow observed adjacent to a magmatic region of high heat flow in western North America. Sea-floor spreading under way at present in the Gulf of California and at the Gorda and the Juan de Fuca oceanic ridges results in high heat flow offshore. Inland the parallel belts of low and high heat flow mark the shallow and deep parts respectively of a subduction zone that was active during the early Cenozoic era. Although the subduction of the oceanic plates off central California ceased more than five million years ago, a heat-flow pattern similar to that seen in the currently active Japan arc system persists. Black triangles again denote recent volcanoes.

The Thickness of the Lithosphere

The determination of the thickness of the lithosphere has until recently been a seismological endeavor. The seismologist's method is to observe the dispersive effects of a given region of the earth on earthquake surface waves that propagate across it. The analysis of dispersion patterns yields information about the elasticity of the crust and the upper mantle; in particular the pattern of dispersion can confirm the existence and locate the position of a zone where seismic waves travel at low velocity. Such a zone is probably a manifestation of a region of partial melting in the mantle, and many geophysicists identify this zone as the asthenosphere postulated in the plate-tectonic model. Since the lithosphere is what lies above the asthenosphere, the depth to this low-velocity zone is equal to the thickness of the lithosphere.

It has been known for some time that under the old Precambrian shields of the continents the seismic low-velocity zone either is absent or is deep and only weakly developed. In contrast, the young and active geologic provinces such as those in the western U.S. have a shallow and well-developed low-velocity zone. Recently a number of reports have appeared suggesting that the depth to the low-velocity zone under the oceans also increases with the age of the ocean floor, implying that the oceanic lithosphere progressively thickens with time.

The depth at which partial melting takes place in the mantle in a given region depends on the temperature at which the rock of the mantle begins to melt and on the variation of temperature with depth. The depth profile of the actual temperature, called the geotherm, in turn depends strongly on the heat flow. Thus with the aid of considerable extrapolation surface heat-flow

The legend continues...

The Mercedes-Benz 280SE Sedan: Heir to a Classic.

This obvious heir to the legendary Mercedes-Benz reputation for elegance has the look and size of the classic 450 Series Sedans—with the performance of a most unusual fuel-injected 6-cylinder engine.

The 280SE: Subtle dimensions in a 6-cylinder sedan.

The 280SE ("SE" for Super Einspritz, or injection) is powered by the only fuel-injected double-overhead camshaft, six-cylinder engine available in the United States. It uses a simplified Continuous Injection System for injecting fuel that offers distinct advantages over delicate electronic injection systems.

The system automatically compensates for weather conditions, engine load and altitude—and meters precisely the right amount of fuel to the engine at precisely the right instant.

The engine to which the CIS fuel-injection system is wedded is the famous 6-cylinder Mercedes-Benz double-overhead camshaft power plant. The double-overhead camshaft configuration provides smoother engine operation at all speeds and eliminates the need for pushrods and lifters.

Suspense-free suspension

The advanced front suspension system of the 280SE derives from the famous C-111 high-speed research vehicle. It has many advantages: Extremely precise steering control with instant response, almost no tilting of the car in cornering and braking, straight-line control under heavy braking and rough road conditions, and better road-holding ability under all conditions.

Matched to the remarkable fully independent front suspension is a fully independent rear suspension that does its share in creating the 280SE's incomparable roadability. It helps eliminate the suspense you may have experienced in rigid-axle cars.

Trunk—18.2 cubic feet of usable space.

The 280SE Sedan maneuvers full left to full right in just 2.7 turns of the steering wheel. Its precisely engineered power-assisted linkage, coupled with a zero-offset steering mechanism, ensures the 280SE's driver of exquisite maneuverability and tautly controlled handling in any road situation.

Mercedes-Benz pioneered the use of dual-circuit power disc brakes on *all* four wheels, starting in 1968. In the 280SE Sedan, the front disc brakes are internally ventilated.

The look of a leader

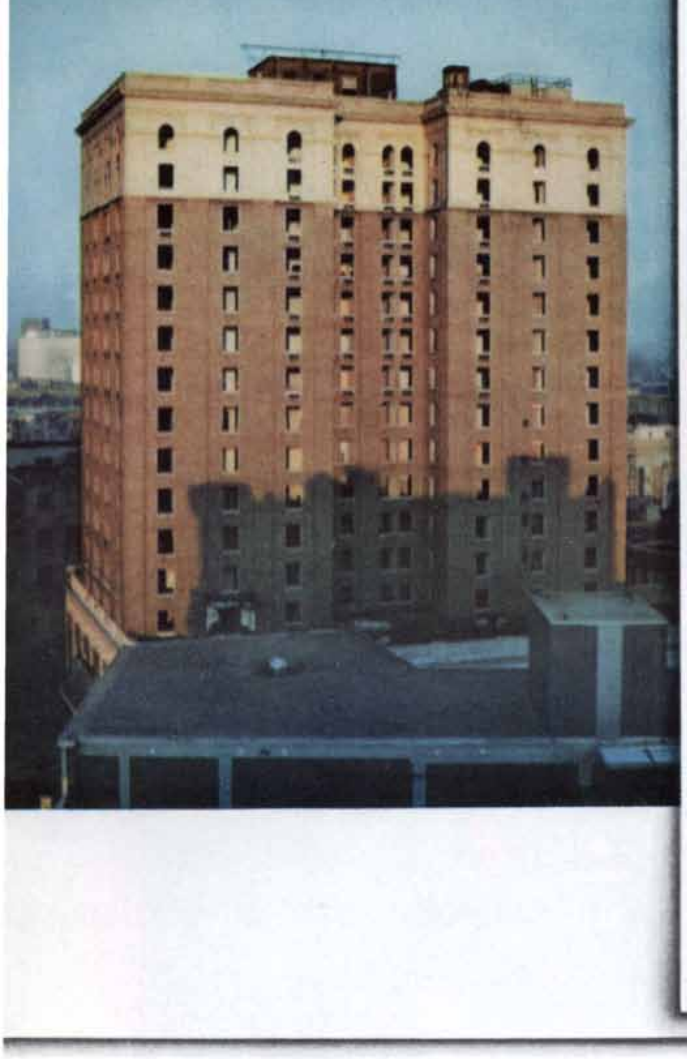
As with all Mercedes-Benz cars, both exterior and interior of the 280SE have a distinctive elegance that bespeaks leadership. The elegance derives naturally out of the car's engineering; it is not imposed.

Everything fundamental to your safety and comfort is standard with the 280SE; Mercedes-Benz has never viewed these most important factors as "optional extras." That is because Mercedes-Benz engineers create only cars that are totally integrated engineering designs—from their power plants to suspension and steering to dashboard instruments.

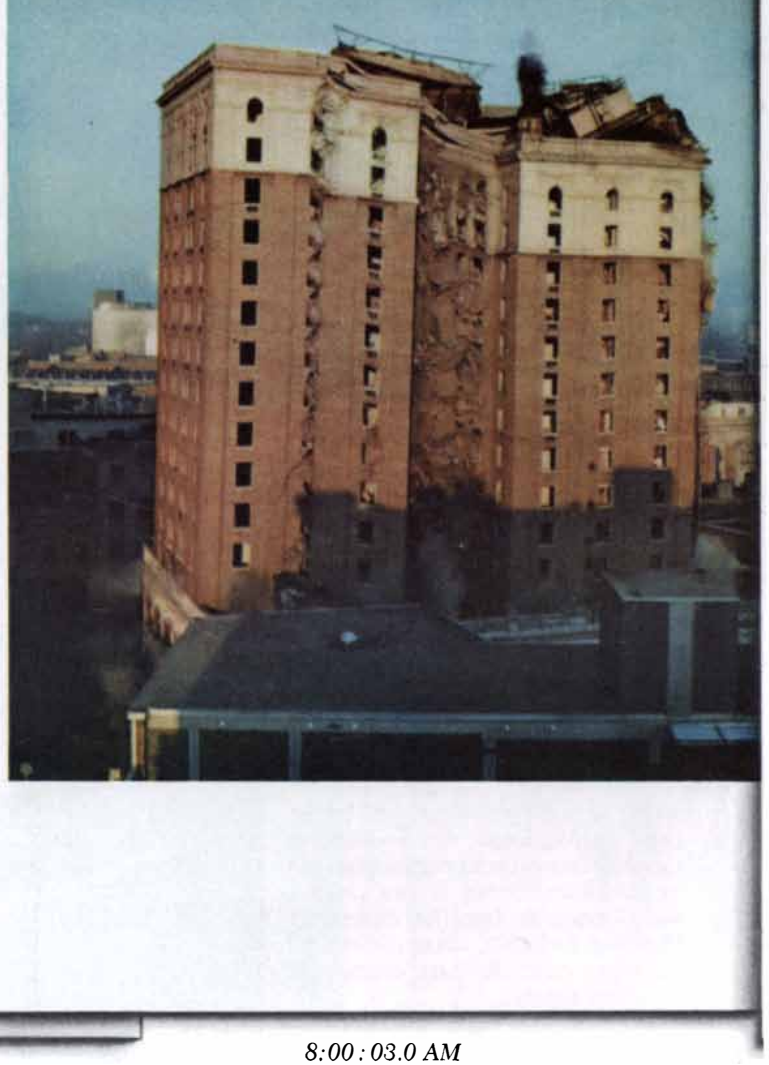
Arrange a drive soon with an authorized Mercedes-Benz Dealer. It may, we suggest, spoil you for any other luxury sedan.

Mercedes-Benz

Engineered like no other car in the world.



8:00 : 01.5 AM



8:00 : 03.0 AM

"When every job you do falls flat, you need all the document you can get." —*John L*

The president of Controlled Demolition, Inc. thinks the SX-70 is a dynamite camera.

And he's an expert.

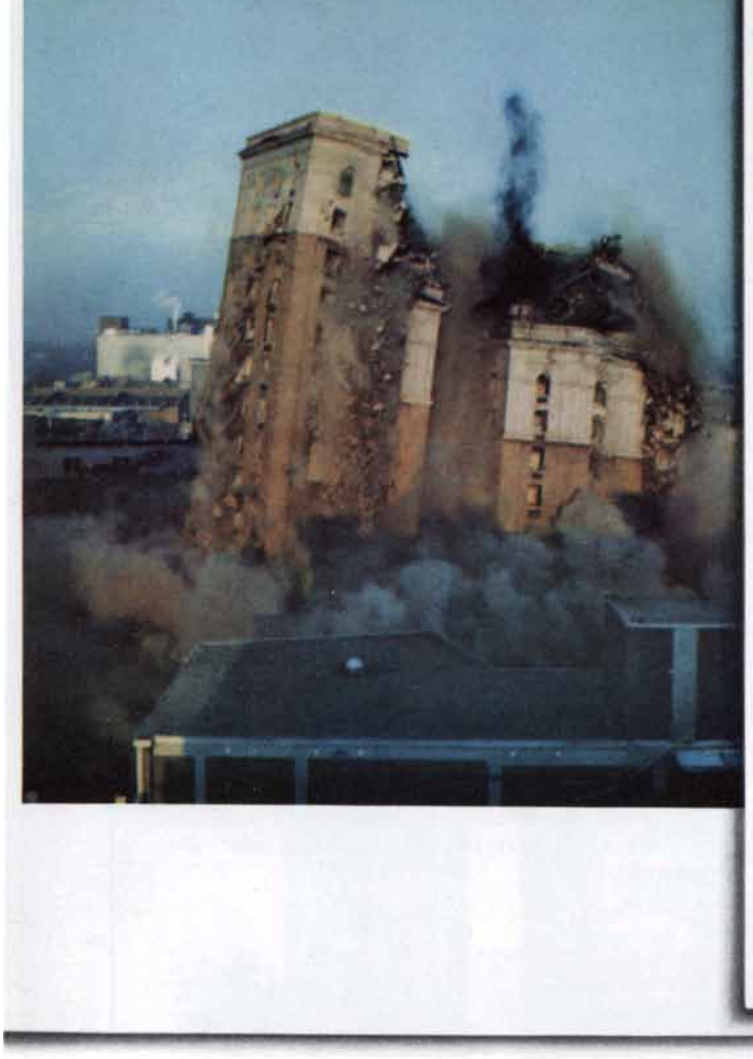
He uses precisely applied explosives to reduce huge edifices to neat piles of rubble. Without so much as jarring neighboring structures.

And he uses the SX-70 Land camera to photograph every aspect of the work. For his own records. And to help sell his services to potential customers.

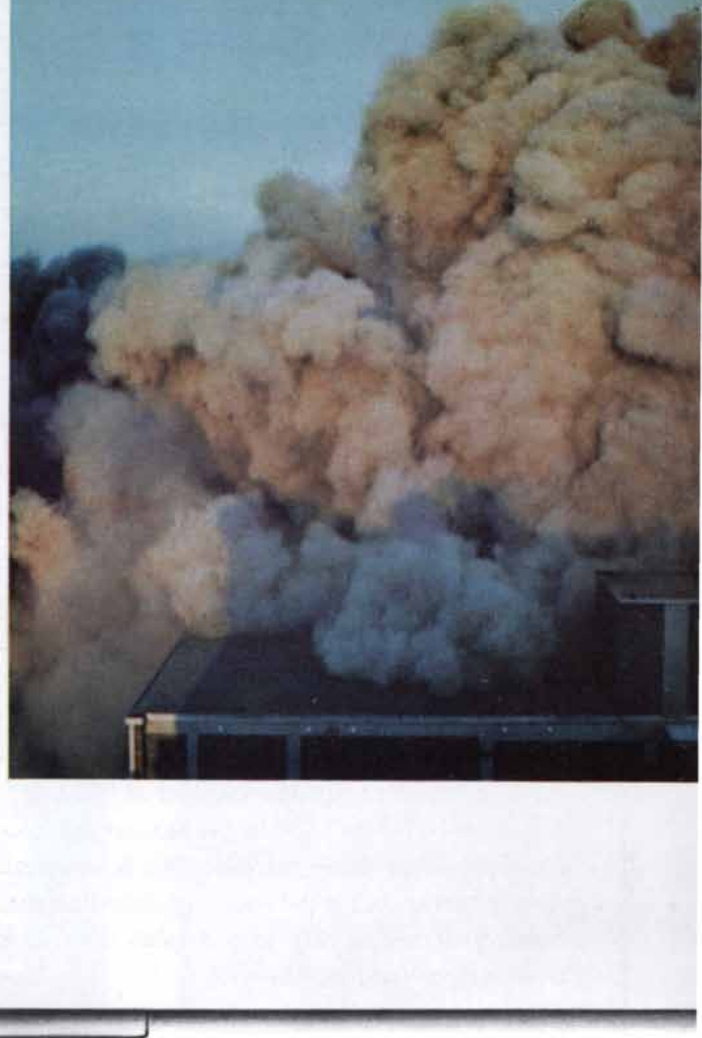
Why the SX-70? It's simple:

With this one camera, he gets instant aerial photos of prospective demolition sites (for planning purposes). And close-ups from 10.4 inches of dynamite placement locations (to help him brief his crew).

He also gets sequential photos at 1.5-second intervals of the demolition in progress (for publicity purposes).



8:00 : 04.5 AM



8:00 : 06.0 AM

ıtion

Loizeaux

The pictures are virtually grainless. So he can analyze the smallest details under a magnifying lens.

And he gets all these pictures, on the spot, at the push of a button. Without worrying about light conditions. Or shutter speeds. Or development timing.

Some business people used to think that photography was a complicated way to make records.

But the SX-70 has exploded *that* theory.

The SX-70 camera from Polaroid makes instant records for any business.

SCIENTIFIC AMERICAN
Filmstrips

Rufus Porter
A Scientific American

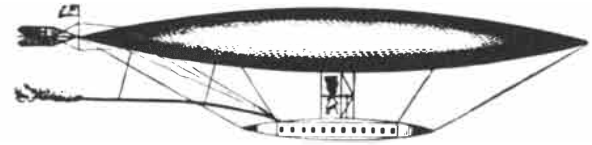
Rufus Porter—artist, inventor, journalist and founder, in 1845, of SCIENTIFIC AMERICAN—should be better known to his countrymen. His exuberant murals adorn 100 New England homes and are finding their way into museums as treasures of American art. His long list of inventions was crowned before 1850 by his demonstration of a self-propelled flying machine. His SCIENTIFIC AMERICAN spoke for the buoyant rationalism that committed this country to industrial revolution from early in the 19th Century.

The 20-minute filmstrip for DuKane projectors *Rufus Porter, A Scientific American* [$22, shipping included] brings this Yankee Leonardo back again to life. In his own colorful words and pictures it recounts the story of his untiring enterprise and places him in the context of the era in American history he so well personified.

Recommended for High School and Junior High School classes in American History, Social Studies and Art History. Available in filmstrip and tape cassette for DuKane projectors.

"We have been frequently solicited by enterprising mechanics to engage in the publication of a new scientific paper, for the advancement of several new, curious and useful arts which have recently been discovered and introduced . . ."

"It is not infrequently the case that the productions of self taught artists far surpass in excellence those of regular bred artists who have studied with the most popular masters . . ."



"It may be anticipated that within a few months these aerial machines may be seen soaring in various directions and at different elevations, some above the clouds, and others, like swallows, sailing leisurely just above the surface of the earth."

SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue, New York, N.Y. 10017, Dept. RP

Please send me _____ copies of *Rufus Porter, A Scientific American,* a filmstrip presentation for DuKane projectors, at $22 (shipping included). I understand I may return same and receive a full refund if I am not completely satisfied.

Enclosed is my check for $ _____.
Please include appropriate sales tax.

_____
Name (please print)

_____
Institution (if any)

_____
Address

_____
City and State Zip

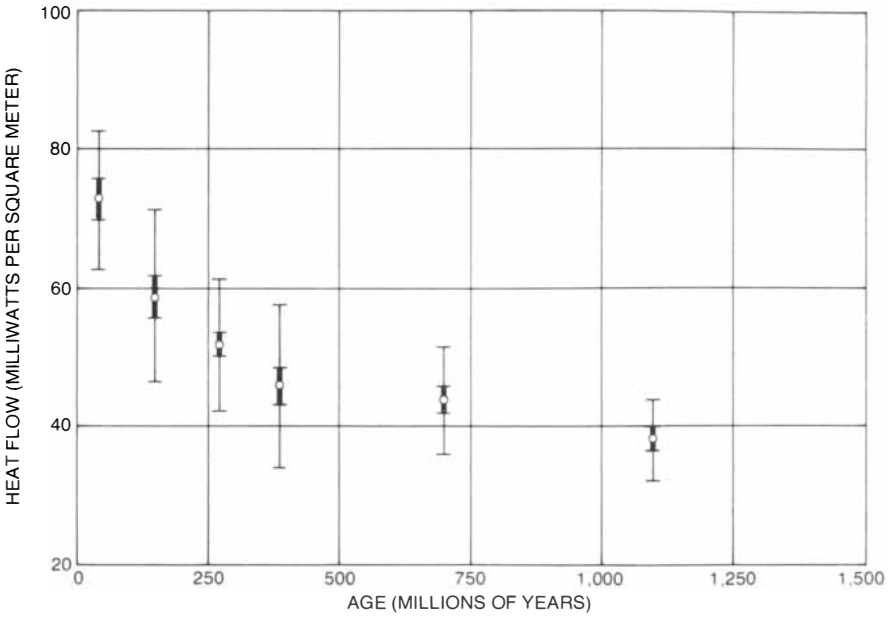
data can be used to predict the thickness of the tectonic plates.

Since direct measurement of temperatures in the earth is limited to the top 10 kilometers of the crust, the extrapolation of temperatures to depths of 100 kilometers or so involves several assumptions. One needs to know how the thermal properties of the rock vary with temperature, how radioactivity is related to depth and for oceanic regions how the oceanic plate cools after it is formed at the ridge. Recent laboratory measurements and field observations have provided enough data for the construction of detailed models, so that one can calculate characteristic geotherms for both continental and oceanic regions with some confidence [*see illustration on page 76*]. The depths to partial-melting conditions predicted from such calculations agree well with the seismologists' results from their surface-wave studies. Both the heat-flow measurements and the seismological data indicate that oceanic plates thicken as they age, from a few kilometers soon after their formation at a ridge to 100 kilometers or more in the oldest ocean basins, where the heat flow is low.

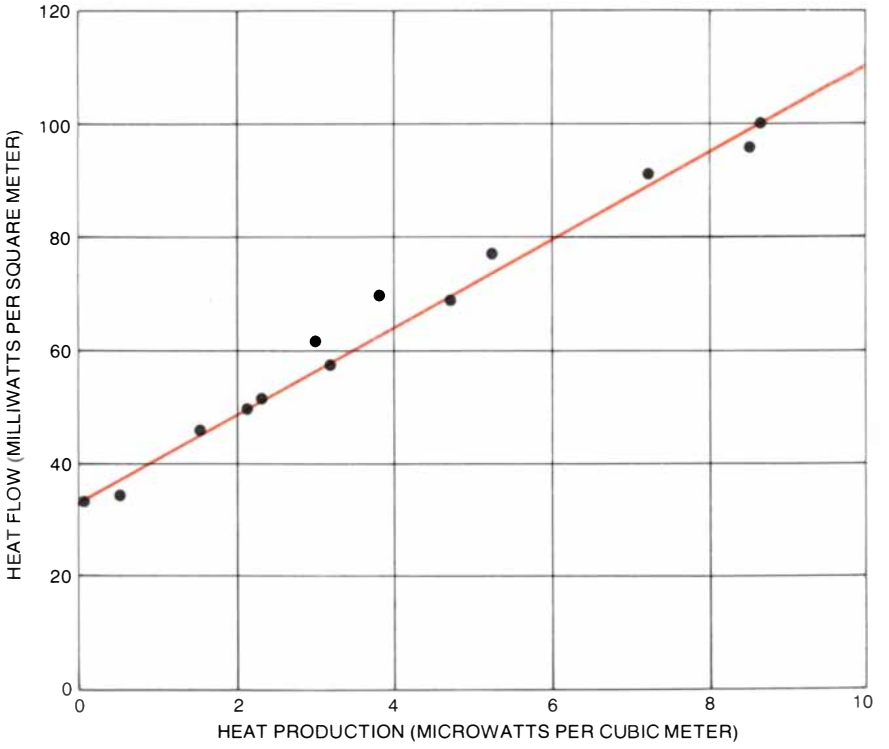
The continental portions of the tectonic plates also show a systematic variation in thickness, from 40 kilometers in young geologic provinces whose heat flow is high to several hundred kilometers under continental shields whose characteristic heat flow is much lower [*see bottom illustration on next two pages*]. For some shields the geotherm does not intersect the mantle's melting curve at any depth, and so in a strict sense the asthenosphere should not exist under the shields. In these areas thick lithosphere would be coupled directly to the deeper interior, acting as an "anchor" to retard the motion of the plate system. More realistically, we expect that the plate constituting a shield does decouple from the deeper interior, probably at a depth at which the geotherm makes its closest approach to the melting curve of the mantle. In a sense, then, the continental shields are at present probably dragging anchor.

Thermal History of the Earth

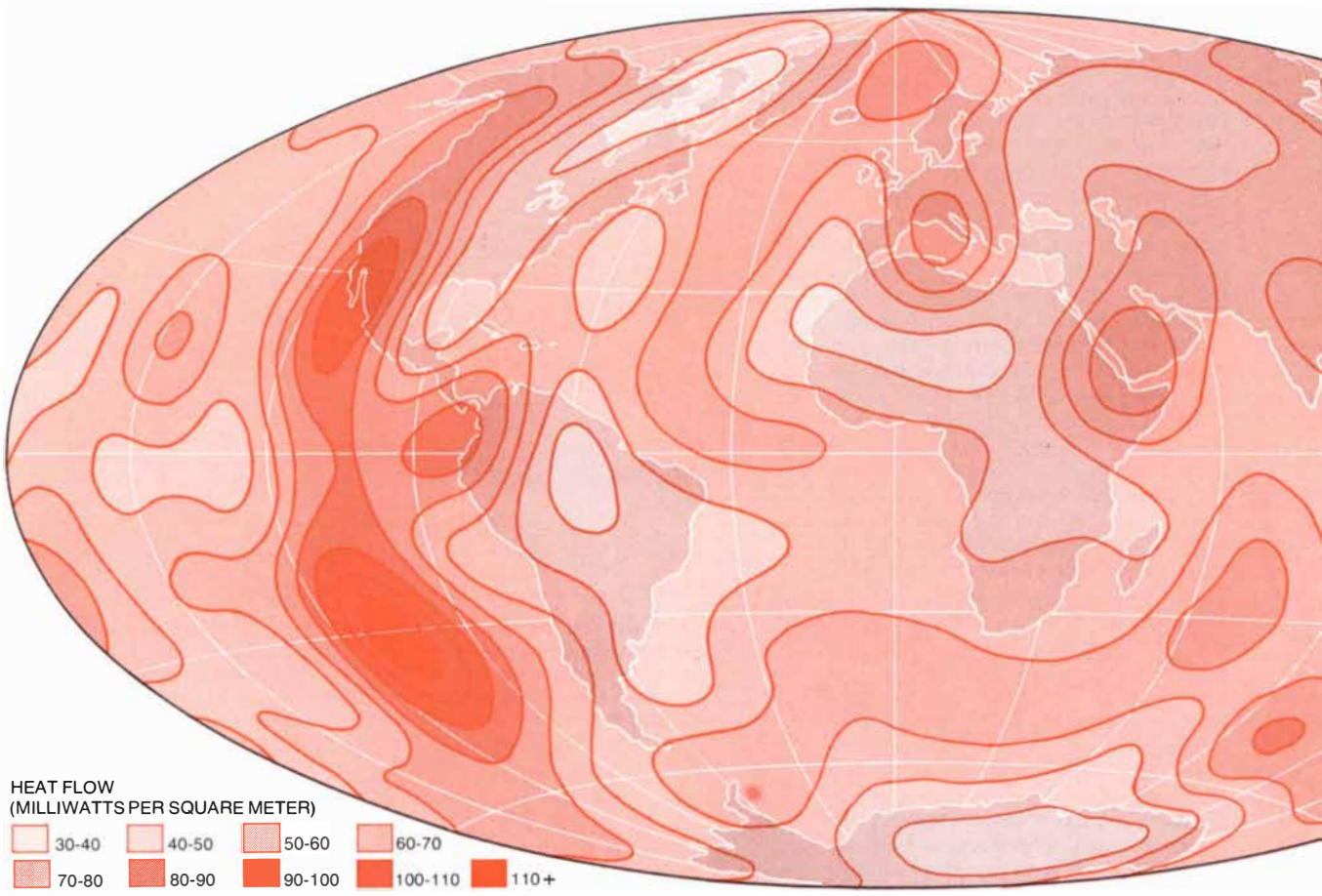
What can be said about the thermal state of the earth in times long past? Most treatises on this subject begin by telling why the problem is a difficult one and then offer a series of disclaimers in case one or another assumption should turn out to be invalid. (The lesson of Kelvin has not been lost on others!) Nevertheless, some observations can be ventured that make possible certain broad inferences, even though the details necessarily remain indistinct. One first must note the abundance within the earth's crust of the principal heat-pro-

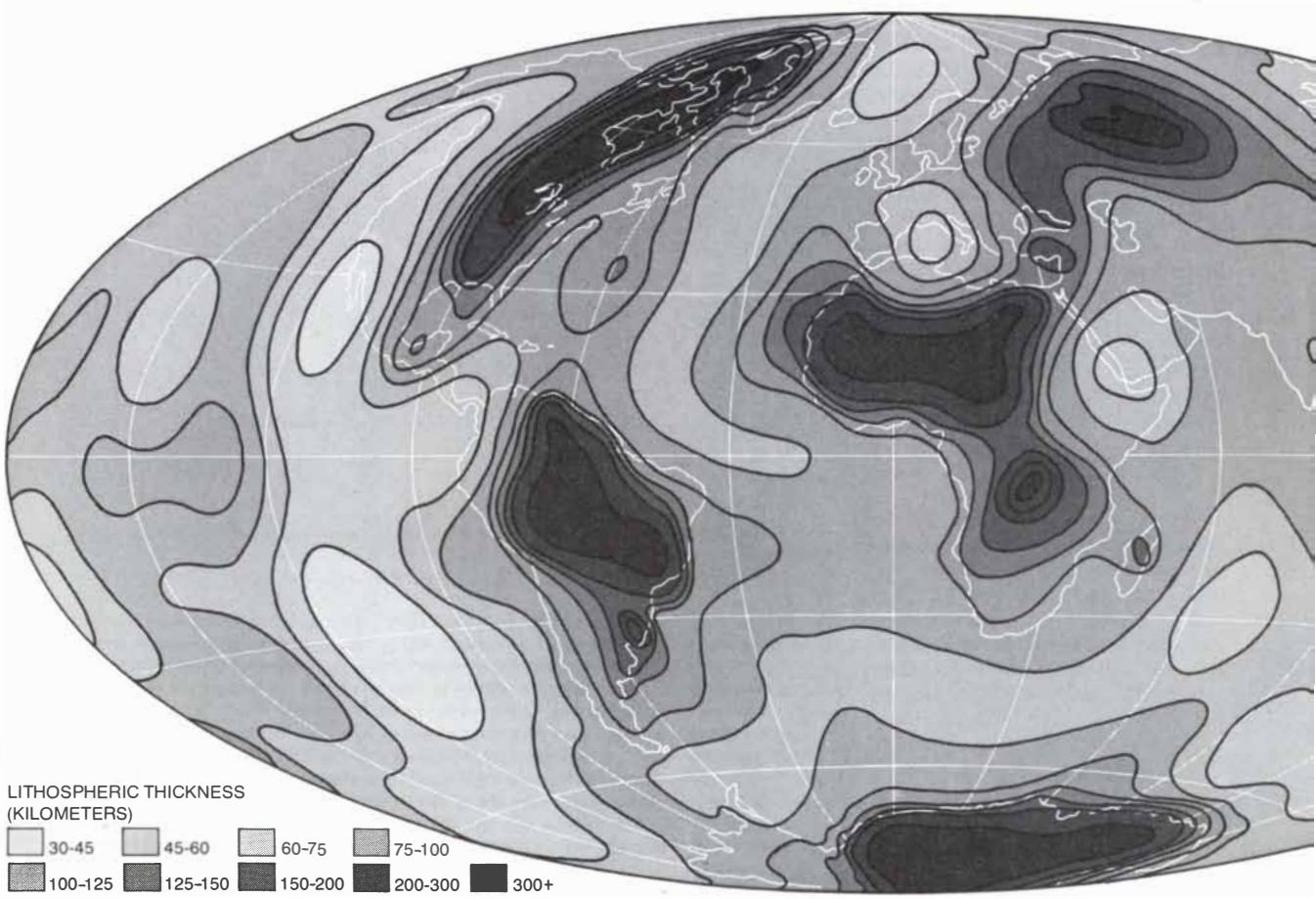


ON CONTINENTS the distribution of heat-flow values also shows an apparent relation to the age of the last major tectonic event that affected the region, just as the distribution of oceanic heat-flow values is associated with the time elapsed since volcanic rock began to cool following the extrusion of mantle material at an oceanic ridge. In the case of continents, however, the measurements indicate that the loss of heat after the tectonic event extends over an interval that is four or five times longer than the decay time for heat flow from the oceanic plate.



CHARACTERISTIC LINEAR RELATION is frequently discovered in plotting heat flow against heat production for a particular region, as can be seen in this graph for data obtained in the eastern U.S. (*black dots*). Such variations in heat-flow data for different continental regions can often be attributed to differing concentrations of heat-producing radioactive isotopes in the outer few kilometers of the earth's crust. The point where the colored line intercepts the vertical heat-flow axis indicates the amount of heat flow coming from below this zone of crustal radioactive-isotope enrichment. The slope of the line tells how deep within the crust the enrichment persists. The slope and intercept values together serve to characterize different heat-flow provinces. The variation of heat flow within a province is governed by the regional variability of the crustal isotopic enrichment. The differences in heat flow between provinces arise principally from variations in the amount of heat flow coming from below the crust.



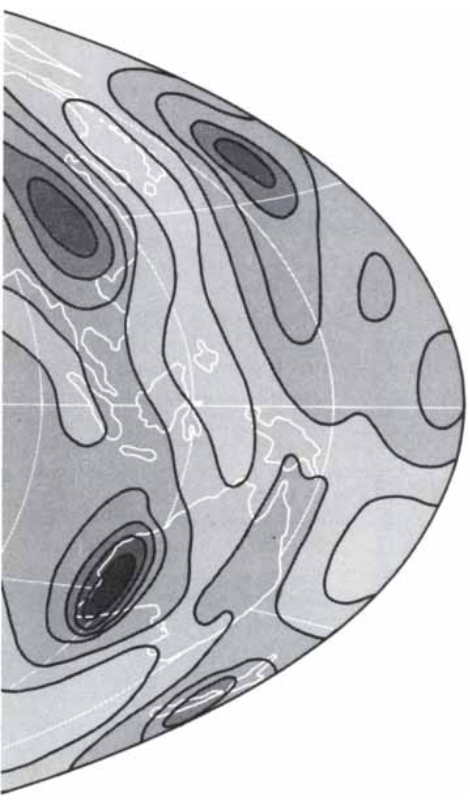


WORLD HEAT-FLOW MAP, constructed by the authors on the basis of the available observations supplemented by estimates, shows variable patterns of heat flow on a global scale. The main zone of high heat flow (*darker shades of color*) is in the eastern Pacific Ocean off Central America and South America. This zone coincides with the East Pacific Rise, a major oceanic ridge where new sea floor is being extruded and carried away quite rapidly to yield a comparatively broad band of high heat loss. The oceanic ridges in the Atlantic Ocean and the Indian Ocean are spreading more slowly and hence result in a narrower zone of above-average heat flow. Other regions of fairly high heat flow include the marginal ocean basins of the western Pacific, which overlie active subduction zones, the western cordillera of North America, where subduction ceased between five and 10 million years ago, and Alpine Europe. The principal regions with below-average heat flow (*lighter shades of color*) include all the ancient Precambrian shield and platform areas of the continents and the oldest parts of the ocean floor.

ducing radioactive isotopes: thorium 232, uranium 238, potassium 40 and uranium 235. The continental crust averages 40 kilometers in thickness, less than 1 percent of the earth's radius, yet its endowment of these heat-producing isotopes is great enough for 40 percent of the heat flow at the earth's surface to arise within the crust itself. The concentration of isotopes in the oceanic crust is less, but it still represents a significant enrichment. The implication of this upward concentration is that there has been a major geochemical segregation within the earth. The fact that continental rocks more than 3.5 billion years old show this enrichment indicates that the segregation took place very early in the evolution of the earth, in all likelihood at the same time that the earth differentiated into a dense metallic core and a lighter silicate mantle.

Significant information about the thermal and tectonic processes in the earth's interior comes from a consideration of how certain physical properties of the earth, such as its strength and viscosity, change with temperature. Viscosity is a measure of the ability of materials, including solids, to flow; a highly viscous material approaches rigidity, whereas a low-viscosity material is much more like a fluid. Elevated temperatures generally promote a lower viscosity. At the surface of the earth and within the lithosphere the rocks are comparatively cold and stiff, but deeper in the earth the increase of temperature with depth almost certainly is accompanied by a decrease of viscosity, which eventually enables the interior to behave like a fluid. Accordingly the interior is likely to be dominated by fluidlike movements driven by density differences of both compositional and thermal origin, in contrast to the purely conductive thermal regime that exists within the lithosphere. In the early evolution of the earth it was probably the gradual reduction of viscosity as the earth was warmed by the radioactive heat that began the process of density stratification giving rise to the core and the upward concentration of the heat-producing isotopes. The rearrangement of the earth's mass as the core settled liberated gravitational energy, which must have accelerated the process. The formation of the earth's core was a unique mechanical and thermal event in the history of the earth, unmatched in scope or drama by the events of later aeons, in spite of the current preoccupation of the earth sciences with contemporary geodynamics as embodied in plate tectonics.

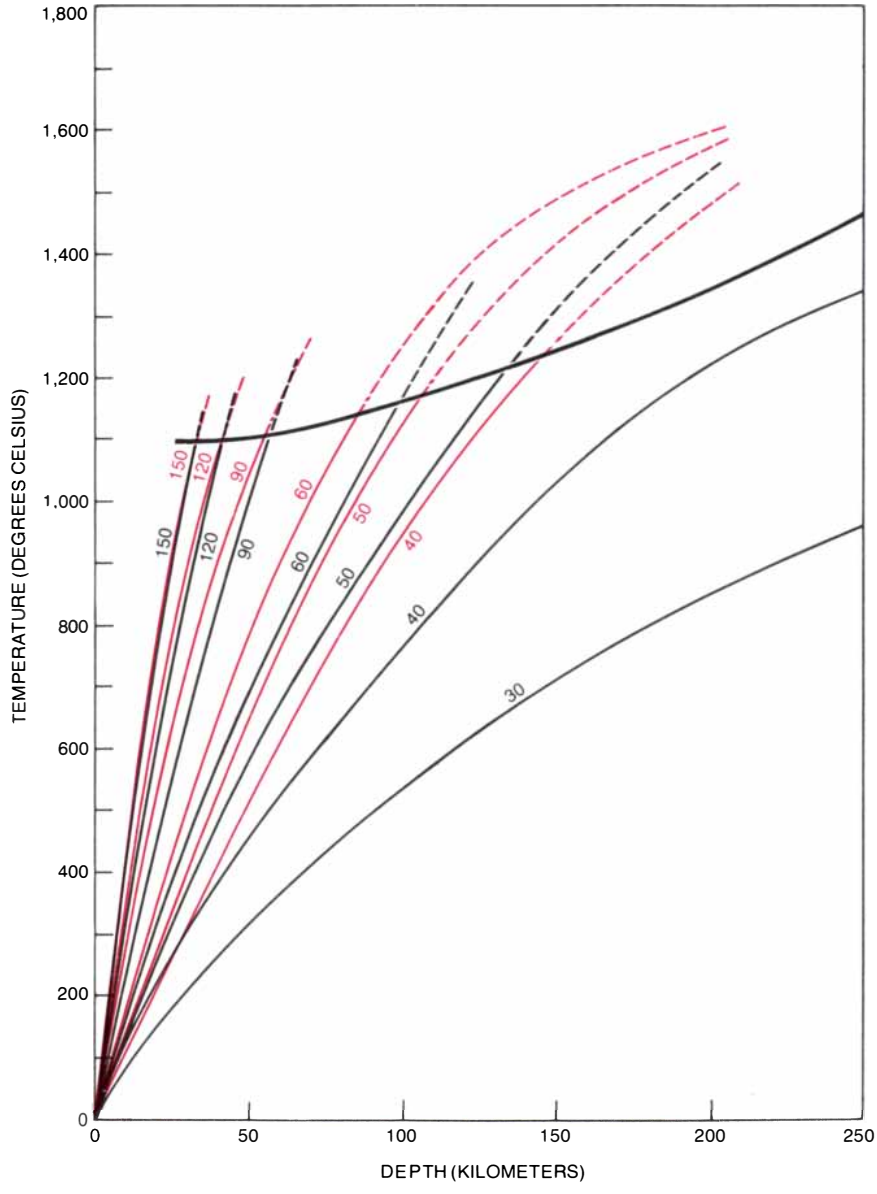
The long-term thermal evolution of the earth is closely linked to the abundance and life span of its heat-producing isotopes. For an isotope to be important in the earth's thermal history, it must be abundant enough and its radioactive half-life must be long enough for it to contribute significant amounts of heat over times comparable to the age of the earth (4.6 billion years). Only the isotopes mentioned above (thorium 232, uranium 238, potassium 40 and uranium 235), with their respective half-lives of 14.1, 4.51, 1.26 and .71 billion years, satisfy these requirements. Taking their relative abundances into account, and calculating the rate of heat generation three billion years ago, one finds that 2.2 times as much heat was being generated by radioactive decay then as is being generated now. This enhanced heat production was probably reflected in a commensurate increase in the heat flow at the earth's surface, from which one can infer that the lithospheric plates then were probably thinner, more easily fractured and hence probably smaller in area but greater in number than the plates of today. The asthenosphere un-



THICKNESS OF THE EARTH'S LITHOSPHERE, or rigid outer shell, can be estimated by selecting for each five-degree-square region of the earth's surface the "geotherm" that corresponds to the mean surface heat flow (*see illustration on next page*). The depth at which the geothermal curve intersects the incipient melting curve determines the thickness of the lithosphere in that region. As the map shows, the lithosphere is thinnest along the oceanic ridges and other regions of high heat flow and thickest under continental shields. According to authors, "there seems to be little doubt that this great thickness of comparatively cold and stiff rock imparts the long-term stability that has come to be associated with the Precambrian shields."

derlying the plates was probably in a more vigorous state of activity than it is at present.

In the future the lithosphere will continue to thicken and the asthenosphere will become more viscous, owing to the continued cooling of the earth and the slow decay of its radioactive heat sources. The motion of the thickening plates will become more sluggish and retarded, although interruptions in this long-term trend similar to the fragmentation and dispersal of the "supercontinent" of Pangaea over the past 180 million years should be anticipated. As the continental shields continue to thicken and to develop substantial viscous anchors one can expect the motion of the plates eventually to cease, bringing to an end the plate-tectonic phase of the earth's evolution. Thus for the diminishing band of earth scientists who still adhere to a nonmobile view of the earth there may be some small solace in the fact that the earth will eventually conform to their concept of it. They must be patient, however, since that time is probably some two billion years hence.



TWO TYPICAL FAMILIES of geotherms (curves representing the variation of the earth's temperature with respect to depth) are shown here for a continental province (*black*) and an oceanic province (*color*). The members of each family of geotherms are labeled according to the heat flow (in milliwatts per square meter) produced at the earth's surface. As might be expected, the temperature found at a given depth in a region of high heat flow will be higher than the temperature at the same depth in a region of low heat flow. Heavy black curve most geothermal curves intersect represents the temperature at which rock will begin to melt in the earth's mantle. The depth at which such melting is observed is variable, depending on the heat flow and the geotherm for that region. Under some continental regions with low heat flow (in particular the stable Precambrian shields) there is probably no melting. Many geologists and geophysicists believe the base of the rigid lithosphere is defined by the onset of partial melting.

Experience the excitement of owning the world's finest *personal* computer – IMSAI 8080

Waiting for you — all the incredible performance and power of the IMSAI 8080. And at a price you would normally pay for a fine home music system.

Introduced less than two years ago, the IMSAI 8080 is sold worldwide and acknowledged as the finest *personal* computer available.

WORLD OF USES

The IMSAI 8080 is a superbly engineered, quality computer. It is versatile, expandable and powerful, putting literally hundreds of applications and uses at your fingertips. Imagine sitting at your desk and enjoying interaction with your IMSAI 8080! Press the *on* switch and you're ready for game playing, research, education, business applications, or for laboratory instrument control. It has all the power you need, and more, to make your application ideas come alive.

GROWS WITH YOU

The IMSAI 8080 is designed for many years of pleasure. With its open-chassis engineering you can expand your system by adding peripherals and interfaces. The 22-slots and 28 amp power supply mean that you can plug-in today's, plus *tomorrow's* modules.

A full line of peripherals can be added as you need them, including our new Intelligent Video Interface; small and large printers, and a single interface that lets you attach multiple devices including a cassette tape recorder. Expect the latest, exciting equipment from IMSAI. We are committed to leadership in this expanding technology.

EASY TO PROGRAM

Using our higher level languages you can easily and quickly write your own applications for the IMSAI 8080. Technically knowledgable? Use our assembly language to develop sophisticated and unique software.

If you're thinking personal computer, treat yourself to the very best – IMSAI 8080

Send for free four-color brochure or $1 for catalog. Call or write for name of nearest dealer.

IMSAI®

The Standard of Excellence in Microcomputer Systems.

IMSAI Manufacturing Corporation
14860 Wicks Blvd. San Leandro, CA 94577
(415) 483-2093 TWX 910-366 7287

Kangaroos

These hopping marsupial mammals have evolved in relative isolation for some 25 million years. Their adaptive strategies closely parallel those of the hoofed mammals of the semiarid Old World grasslands

by T. J. Dawson

Kangaroos have a certain fascination because they differ so much from the usual notion of what constitutes a mammal. They stand apart from other large mammals because they rear their young in a pouch and because they hop. The explanation often given for these odd characteristics is simply that kangaroos are marsupials and marsupials are primitive mammals. This is no answer but reflects an impression that seems to have arisen, particularly among people who live in the Northern Hemisphere, from statements made about the Virginia opossum, the only marsupial in North America. The opossum is frequently described as an archaic primitive mammal, little changed since the time of the dinosaurs. There is perhaps some basis for such statements: fossils of marsupials that coexisted with dinosaurs toward the end of the Mesozoic era, some 100 million years ago, have been classified as belonging to the same genus as that of the Virginia opossum (*Didelphis*). But even if the opossum can in some ways be considered a "living fossil," the same cannot be inferred for the kangaroos. These Australian marsupials represent a recent and successful specialized adaptation to a changing world environment.

The family Macropodidae (literally "big feet") consists of some 45 species, including not only the five or six species of large kangaroos but also smaller hopping relatives ranging from the wallabies, the next largest, down to the rabbit-sized rat kangaroos. The group has evolved from small forest-dwelling browsing animals into predominantly open-range grazing ones. This development represents a recent adaptive radiation in Australia comparable to the radiations of nonmarsupial mammals elsewhere in the world. The rise of the macropodids, possibly the most successful marsupial group among the animals of Australia, was probably related to the spread of grassland in the interior of Australia in the Miocene epoch, between 10 and 25 million years ago.

Similar events were taking place in other parts of the world in the Miocene. It was a time of expansion for the even-toed ungulates: the mammalian order Artiodactyla. We know a good deal about the biology of the artiodactyls, particularly those in the suborder of ruminants, because they include such domestic animals as sheep, goats and cattle. The adaptations found in kangaroos represent solutions to the problems of life in a grassland environment that differ from the solutions arrived at by the artiodactyls. Moreover, even when the solutions arrived at by the marsupials and the nonmarsupials are similar, they embody many differences in detail. A major theme of this story is how these differences in evolution and adaptation to life on the open grasslands bear on the supposed primitiveness of the marsupials.

A brief review of the evolution of mammals will help to establish the relations between the marsupials and other mammals. It is believed all mammals are descended from a mammal-like reptile that lived late in the Triassic period, more than 200 million years ago. Soon thereafter, about 180 million years ago, the stocks leading to the two principal groups of mammals separated. One group was the prototherians, which consist of the egg-laying mammals of the order Monotremata; among the modern representatives of the order are the platypus and echidna of Australia. The other group was the therians, which consist of the marsupials and the placental mammals, that is, mammals characterized by a "true," or chorioallantoic, placenta. Thus the marsupials are more closely related to the placentals than to the monotremes, which retain many reptilian features in addition to egg-laying.

The initial radiation of the therian mammals early in the Cretaceous period stemmed from tiny insect-eating animals and was undoubtedly based on their adaptations to the newly developing flowering plants of the period and their pollinating insects. The separation that gave rise to the marsupials and the placentals took place at this time, some 130 million years ago. Information from a variety of sources suggests that the marsupials developed in North America and were the dominant therians in that region for most of the Cretaceous period, or until about 70 million years ago.

The placentals apparently developed initially in Asia and only reached North America late in the Cretaceous, when there was no oceanic barrier between the two land masses. This invasion of North America was followed by a major evolutionary radiation of placentals that coincided with the extermination of all but one of the previously abundant marsupial lines. Competition with the invading placentals has been suggested as the reason for the great marsupial extinction, but other causes may have been involved, since this was also the time when the dinosaurs finally died out.

Tracing the evolutionary pathways from the tiny insectivorous marsupials of the Cretaceous in North America to the kangaroos of the grasslands and open forests of Australia has long presented a problem. The emergence of evidence supporting the theories of continental drift and plate tectonics, however, has now given many clues to how the marsupials reached Australia. It appears that even though marsupials were disappearing from North America at the end of the Cretaceous, they had by that time spread widely through South America. There the marsupials radiated further. It is from this fauna that the Australian marsupials were apparently derived, at a time when South America was linked to Australia by Antarctica. The latest possible date for this further spread is about 50 million years ago, when Australia and Antarctica became separated.

The fossil record in Australia in this period is impoverished, so that we have no direct information about the original marsupial immigrants. Michael Archer of the Queensland Museum has suggested, however, that they came in several waves. His proposition is based on similarities between the teeth of Australian marsupials and those of South Ameri-

can marsupials, both living and extinct. Whatever the form of the earliest Australian marsupials, by the start of the Miocene, 25 million years ago, the process of differentiation had established their major groups. One of these groups was the macropodids. Their earliest-known fossil representative is a small rat kangaroo from a Miocene formation in the region of Lake Eyre.

One of the most significant aspects of faunal change in Australia over the past 10 to 15 million years has been the increase in abundance and diversity of macropodids. Several lines of kangaroos resulted from this relatively recent radiation. For example, one fossil line consists of giant, short-faced kangaroos with a skull and teeth specialized for heavy browsing. Another line had comparatively high-crowned cheek teeth, adapted for grazing. It was this line that gave rise to all the living groups of large kangaroos.

In short, the kangaroos are far from being primitive and unchanging mammals. In terms of the geologic time scale they have only recently arrived on the scene, evolving in response to new and changing environments, notably the spread of semiarid grasslands. Given the genetic heritage of their insectivorous ancestors of 130 million years ago, how did the two groups of grazing therian mammals, the kangaroos and the ruminants, differ in their adaptation to the same kind of specialized environment?

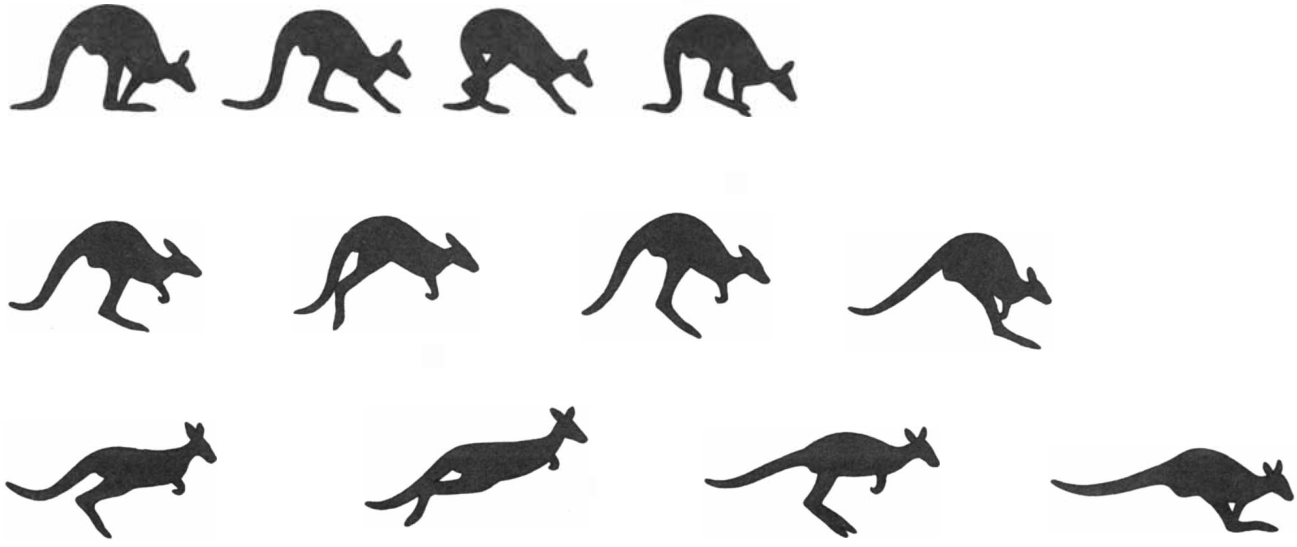
The rearing of young in a pouch is one obvious difference between the kangaroos and their placental counterparts. To understand why the kangaroos retain this basic marsupial feature it is necessary to consider marsupial reproduction. The marsupial female rears her young after a pregnancy that is shorter than that of placental females of comparable size. As a result her young are born at a virtually embryonic stage. Not all marsupials have a marsupium, or pouch, but all of them have young that are firmly attached to a teat for a period after birth that corresponds roughly to the latter period of intrauterine development among placentals.

There is some controversy over the evolution of the bearing of live young among both the marsupials and the placentals. Since the monotremes still lay eggs, however, it is assumed that the bearing of live young evolved only after the early mammals had become well established. The discussion therefore revolves around whether the bearing of live young evolved once, before the separation of the marsupials and the placentals, or evolved independently in the marsupials and the placentals. Actually in the monotreme anteater *Tachyglossus aculeatus* the leathery egg remains in the uterus (and grows markedly) for 27 days and is then laid into a pouch, where it hatches after 11 days. The intrauterine component of this form of reproduction therefore bulks quite large. It is not hard to visualize the next evolutionary step to the deposition in the pouch not of eggs but of small young. This speculation is supported by Jason A. Lillegraven of the University of Wyoming, who suggests that all primitive therians were born helpless and needed intensive maternal care. Marsupials have retained this pattern, but most placentals have extended the period of internal gestation and give birth to more developed young. The ungulates have fully developed the option: their young are born active.

The reason for the retention of this ancestral birth strategy in marsupials and for its high development in kangaroos is something of a puzzle. Among the proposed explanations are the lack of a true placenta and the lack of a mechanism to extend the active life of the corpora lutea of the ovary: the bodies that secrete progesterone, the hormone that is largely responsible for the preparation and maintenance of the uterus in pregnancy. Neither explanation is acceptable. In some marsupials, notably the rabbit-sized bandicoot, there is both a true placenta and a prolonged active life of the corpora lutea.

Another proposal is that the crucial adaptation leading to prolonged gestation in placental mammals is the mechanism that keeps the pregnant animal from mounting an immune reaction that would reject the "foreign" tissue of the embryo. (The developing embryo is foreign in the sense that it consists of tissues derived not only from the mother but also from the father.) Lillegraven has suggested that in placentals the major barrier between the fetal antigens and the maternal antibodies is an anatomical one, consisting of a special component of the placenta, the trophoblast. The marsupials have not evolved this special barrier and retain the ancestral condition: an inert eggshell (which is, of course, of maternal origin) is maintained until late in pregnancy and apparently blocks the maternal immune response. The retention of the eggshell limits the size of the young because once it breaks down, the embryo has to be born before a full-fledged immunological attack can be launched on it.

The marsupials have retained the basic ancestral therian reproductive pat-



CHANGING GAITS of kangaroos range from pentapedal locomotion (*top*) typical of slow movement to high-speed hopping in excess of 40 kilometers per hour (*bottom*). In pentapedal locomotion the tail of the kangaroo serves as the third leg of a tripod, helping the animal to support itself on its forelimbs while the large hind limbs move forward. In low-speed hopping (*middle*) the tail assists balance but no longer acts as a support. Number of hops per minute is almost constant at speeds between 15 and 35 k.p.h. (about nine to 22 miles per hour); more speed comes from longer strides. High-speed hopping, from 40 to 50 or more k.p.h. (25 to 30 or more m.p.h.), requires still longer strides and faster hopping. These schematic representations are based on silhouettes of red kangaroos prepared by Frank Knight.

tern, but the kangaroos have modified it so that in some respects it even parallels the ungulates' advanced pattern of bearing fully active young. For example, the kangaroo pouch is much larger and more baglike than the average marsupial pouch, and the time the young kangaroo spends in it is markedly extended. In the red kangaroo (*Megaleia rufa*) pregnancy lasts for only 33 days, two days less than the estrous cycle. The offspring weighs only about .8 gram at birth, but it is larger, both relatively and absolutely, than the offspring of any other marsupial. The newborn red kangaroo still has a lot of growing to do after it makes its unaided climb from the birth canal to its mother's pouch. It does not leave the pouch permanently until 235 more days have passed.

When the young kangaroo, or joey, finally does leave the pouch, it can be regarded as having been born again, and it is quite comparable in its development to the newborn young of such placentals as deer and sheep. It continues suckling for a period comparable to the suckling time of ungulate young. During this additional suckling period, however, the nature of the mother's milk changes. The ability of the kangaroo mammary gland to secrete two different types of milk is an adaptation that allows the female to suckle a large "young-at-foot" offspring while simultaneously nursing another in her pouch. The young-at-foot puts its head back into the pouch to suckle from its original teat; the joey in the pouch is attached to another one. The two milks secreted by the separate mammary glands are quite different in volume and composition. How this can be achieved under the same hormonal conditions is an intriguing question.

One peculiarity of the reproduction of most species of kangaroos is the phenomenon known as embryonic diapause. During this period of reproductive quiescence a viable embryo is carried in the uterus with its development entirely arrested at the blastocyst stage, when it consists of from 70 to 100 cells and is about .25 millimeter in diameter. What is the function of this adaptation? To understand it one must consider the kangaroo's reproductive cycle.

Fertilization does not generally alter the course of the kangaroo's estrous cycle, and the cycle is not suppressed during pregnancy. It is halted only by the stimulus of lactation after birth. In most species estrus and mating occur again very soon after birth, that is, at the same time they would have occurred if the female had not become pregnant. The fate of the ovum fertilized at this immediately postbirth mating was first described in 1954 by G. B. Sharman of Macquarie University in Australia.

The ovum develops to the blastocyst stage but then becomes dormant. In the red kangaroo if the newborn young in the pouch survives normally, the blastocyst will remain dormant for about 200 days. Thereafter development resumes, and within 30 days birth takes place; the previous inhabitant of the pouch is evicted shortly before the new offspring enters its shelter. Birth is again followed by estrus and mating, and another blastocyst may result. Since the evicted joey may continue to suckle for another four months, the female red kangaroo may have three offspring in the "pipeline" at any one time: a dormant blastocyst, a small joey nursing and developing in the pouch and a larger young-at-foot still suckling.

The adaptive significance of embryonic diapause has been debated at length because it takes different forms in the macropodid family. For example, among red kangaroos its significance would appear to be associated with a rapid rebuilding of the kangaroo population after a reduction in numbers due to a drought in the arid Australian interior. Unlike the placental embryo, which is plugged into the maternal blood supply for nourishment, the joey is fueled only by milk. During a drought the young in the pouch and even the young-at-foot may perish if the milk supply decreases. When this happens, the blastocyst resumes development, and after the embryo is born the mother mates and produces another blastocyst.

During a drought the stress would be greater on the larger young, and in a prolonged drought the young in the

FEMALE RED KANGAROO rests in the shade of an acacia bush on a hot day with the air temperature 46 degrees Celsius (115 degrees Fahrenheit). Offspring at her side is still nursing but has long since left the pouch. In the female's pouch is another maturing offspring that will soon join the older "young-at-foot." The female's posture, with tail tucked under, helps to reduce radiant-energy input.

pouch die at progressively earlier ages until the female stops breeding altogether. One can thus see the adaptive advantage in an arrangement whereby the female will, except under the severest conditions, harbor a developing offspring. The significance of this adaptation to a widely dispersed population can be further appreciated in view of an additional fact. As A. E. Newsome of the Wildlife Research Division of the Commonwealth Scientific and Industrial Research Organization has shown, male red kangaroos may become temporarily infertile under hot drought conditions.

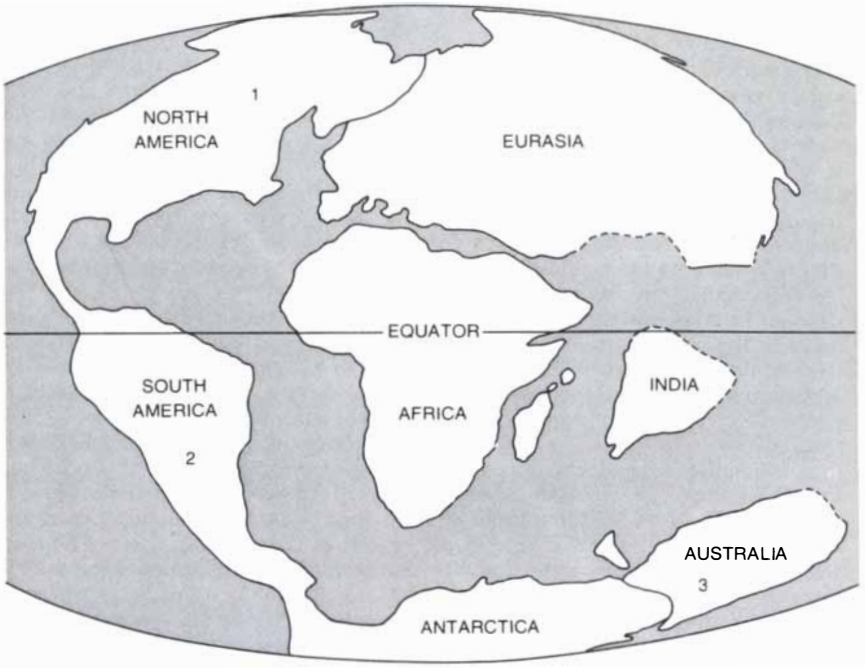
Knowledge of the reproductive physiology of kangaroos has been slowly accumulating for many years. It is only quite recently, however, that the characteristics of kangaroo locomotion have become appreciated. The hopping of kangaroos intrigued the earliest European visitors to Australia. One of the first comments was made by Sir Joseph Banks in the journal of his voyage with Captain James Cook on the *Endeavour* from 1768 to 1771:

"Quadripeds we saw but few, and were able to catch few of them that we did see. The largest was called by the native *Kangaroo*. It is different from any European and indeed any animal I have heard or read of except the Gerbua of Egypt, which is not larger than a rat when this is as large as a middling Lamb; the largest we shot weighed 84 lb. It may however be easily known from all other mammals by the singular property of running or rather hopping upon only its hinder legs carrying its fore bent close to its breast; in this manner however it hops so fast that in the rocky bad ground where it is commonly found it easily beat my grey hound, who tho he was fairly started at several killed only one, and that quite a young one."

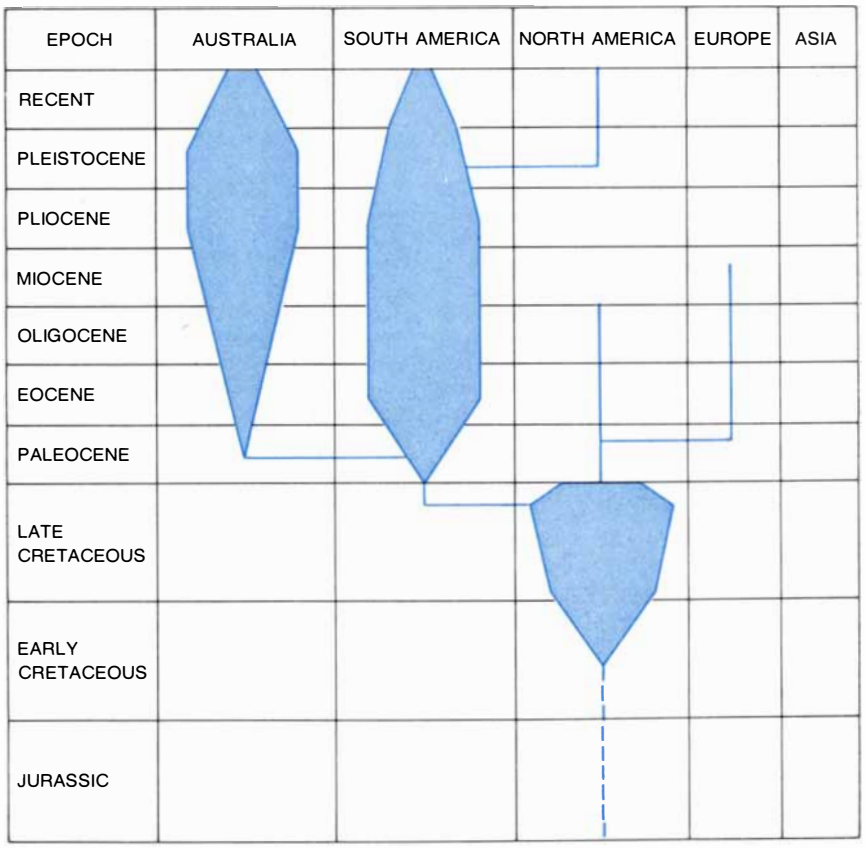
Banks had every reason to be surprised by the sight of a large hopping animal; small hoppers such as his Egyptian "Gerbua" (the jerboa, a hopping rodent) abound in the deserts of the world but large ones were unknown. Indeed, if he had known what we know today he would have been even more surprised: the kangaroo seems to represent the only instance in evolutionary history of any animal weighing more than three to five kilograms having adopted a hopping gait.

It was once believed bipedal dinosaurs hopped, but there is no evidence to support this view. Why, we may ask, is hopping such an unusual method of locomotion in larger animals and why did the kangaroos develop it? As we now begin to understand such matters as the relative energetic cost of various forms of locomotion and the relative agility associated with different gaits, the answers to these questions are becoming more apparent.

An understanding of the characteris-



MARSUPIAL EXODUS to Australia began at the end of the Mesozoic era when some of these successful early mammals left their North American homeland (*1*) for South America (*2*), which was still joined to Antarctica and Australia (*3*) in the early Cenozoic. North and South America were separated for much of the Cenozoic. When they were rejoined, advanced placental mammals from the north competed with the marsupials of South America, leaving Australia the major marsupial reservoir. Marsupials died out in North America by the middle Cenozoic. The opossum is not a Mesozoic leftover but a migrant from South America.



EVOLUTIONARY RECORD of the marsupials is outlined in this diagram, beginning early in the Cretaceous with the divergence between placental and marsupial mammals. The first marsupials flourished in North America and also entered South America at this time. During the Paleocene some marsupials from North America reached Europe; in the Miocene both groups died out. Now the majority of marsupial species are found in Australia and its vicinity.

tics of hopping has been difficult until recently because there was no general framework of information about the energetic cost of locomotion with which the cost of hopping could be compared. C. Richard Taylor of Harvard University, Knut Schmidt-Nielsen of Duke University and their collaborators have now begun to erect such a framework. It is apparent from their studies that for the different types of running there is a regular and predictable energetic cost that depends primarily on the size of the animal and the running speed. The cost of running differs between quadrupeds and bipeds, but it is largely consistent and predictable, at least up to moderate speeds.

A joint interest in several aspects of kangaroo energetics and temperature regulation led to a collaboration between Taylor and me. Four kangaroos were flown from Sydney to Boston. We were able to train two of them to hop on a large treadmill as their oxygen consumption (and consequently the energetic cost of hopping) was measured. These studies showed that the energetic patterns of hopping diverged markedly from those of both bipedal and quadrupedal running. The most significant difference is that a kangaroo traveling at moderate speeds does so more economically than running bipeds or quadrupeds of similar size do. This is obviously advantageous to the kangaroo.

Why, then, is hopping so rare in large animals? The answer probably lies in the locomotor pattern of kangaroos at low speeds. At speeds below six kilometers per hour (about 3.7 miles per hour) kangaroos do not hop. They move in a rather odd way, using their heavy tail as an additional support. This gait has been called pentapedal because the tail acts as a fifth leg that helps the forelimbs to support the animal as the large hind limbs are moved forward together. (Contrary to a widespread belief, kangaroos can move their hind legs independently. They are good swimmers, and as they swim their hind legs alternate.)

The pentapedal gait is clumsy and energetically costly. The reason kangaroos resort to this odd mode of locomotion at low speeds seems to be that it would be even more costly to hop. A simple physical analysis of the various energetic components of hopping supports this view. The same kinds of factors explain why human beings shift from walking to jogging when they want to move faster, and why it is costlier for large animals to climb trees or run up slopes than it is for small animals. Moreover, shifting to quadrupedal locomotion at low speeds, as small hopping animals do, seems to be precluded in kangaroos by the gross anatomical specialization of their hind limbs and the reduction in size of their forelimbs. That these specializations are related to increased size is also indicated by the fact that some small members of the macropodid family do employ quadrupedal locomotion at low speeds.

In short, at low speeds large kangaroos have major locomotion problems but at high speeds their hopping gait has marked advantages over running. Once they start hopping the energetic costs do not change over a wide range of speeds. In the treadmill experiments when the kangaroos hopped at speeds above 15 k.p.h. (about 9 m.p.h.), they traveled more economically than running quadrupeds or bipeds of similar size. The pattern of hopping energy expenditure appears to be explained by the combination of a fixed hopping rate and the elastic storage of energy in tendons. Energy can be stored in such elastic fibrous tissues much as it is stored in the spring of a pogo stick. The treadmill measurements of the energetic costs of hopping were not practical at speeds above 22 to 23 k.p.h. (about 14 m.p.h.), but some overall comments can be made on the basis of knowledge gained from observations of small hopping animals and from field observations of kangaroos.

The proposition can be advanced that unhindered animals normally travel at their most comfortable speed, that is, their economical speed. For kangaroos this appears to be between 20 and 25 k.p.h. Energy costs probably increase above that speed because the elastic storage of energy cannot increase indefinitely. Hopping frequency, however, remains constant up to about 40 k.p.h. (25 m.p.h.); the increase in speed is achieved mainly by increasing the length of the stride. If a kangaroo is pressed, it can maintain this speed for a couple of kilometers. Kangaroos can even increase their speed above 40 k.p.h., but they rarely do so. They hop in such high-speed bursts only in emergencies and then only for a few hundred meters. I have monitored kangaroos traveling at speeds of up to approximately 50 k.p.h., and a speed of 65 k.p.h. (about 40 m.p.h.) has been reported. At these maximum speeds there is a noticeable increase not only in stride length but also in stride frequency. It is certain that an increase in stride frequency is accompanied by a marked increase in energy consumption, but how the energetic cost of hopping at maximum speeds is related to the cost of high-speed quadrupedal locomotion is not yet known.

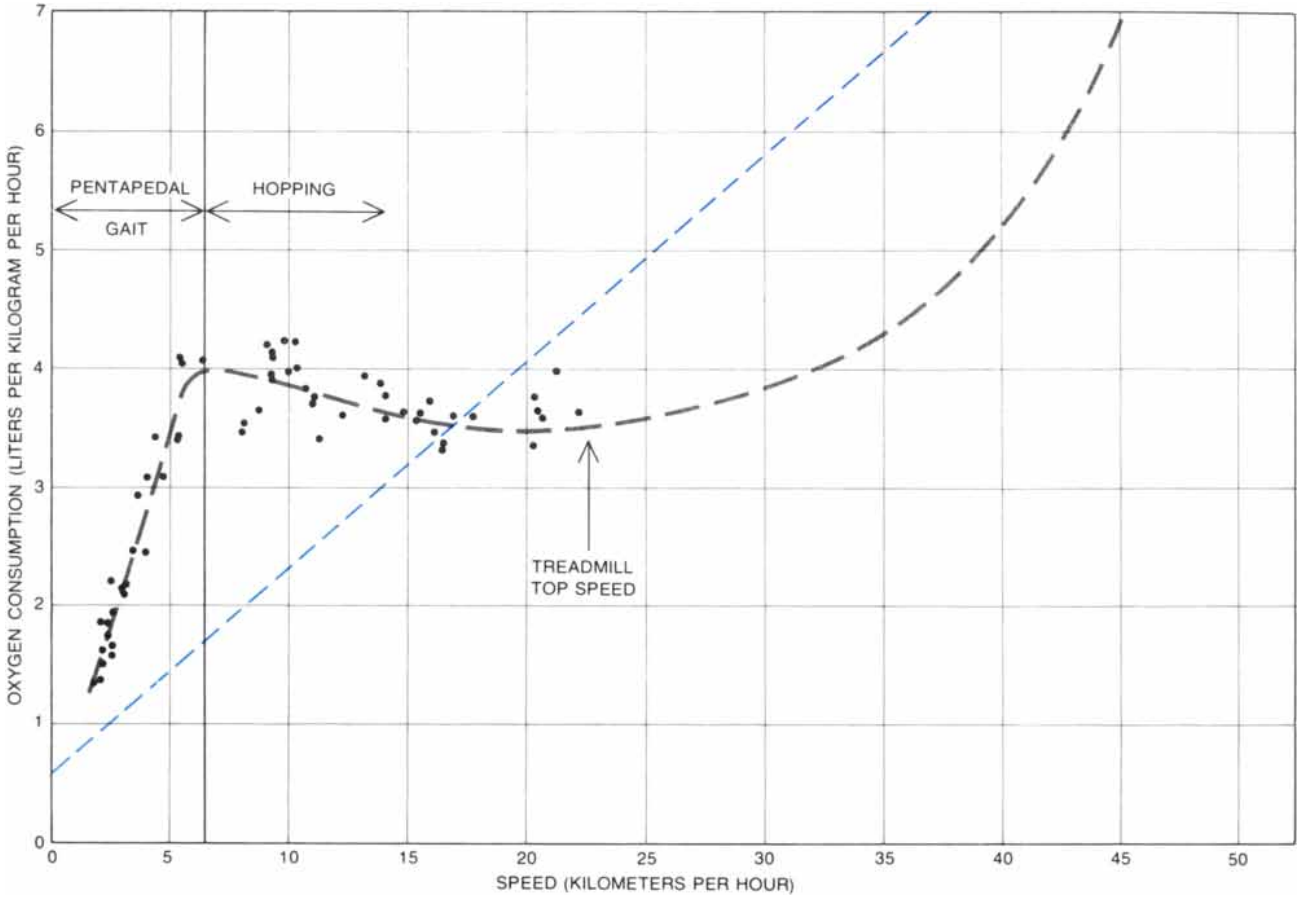
It is clear that the kangaroo mode of locomotion has advantages in spite of its shortcomings at low speed. Kangaroos have radiated widely throughout Australia and have survived the coming not only of aboriginal man and his semidomesticated dog, the dingo, but also of European man and his dogs. A biomechanical analysis of hopping indicates, however, that its advantages diminish significantly with increasing size. It is perhaps pertinent that the largest kangaroos known to have evolved, some of them more than three meters tall, vanished along with many other large marsupials in the Pleistocene epoch.

Why did hopping come into existence at all? The answer may lie in a conservative or perhaps primitive marsupial characteristic. For many years one of the factors that was thought to support the notion that kangaroos are primitive was the reported low metabolic rate and the low and fluctuating body temperature of marsupials in general. The basal metabolism of kangaroos and other marsupials is only about 70 percent of that of advanced placentals. Their body temperature too is slightly lower, 35.5 to 36.5 degrees Celsius (96 to 98 degrees Fahrenheit). They do, however, maintain a constant temperature level.

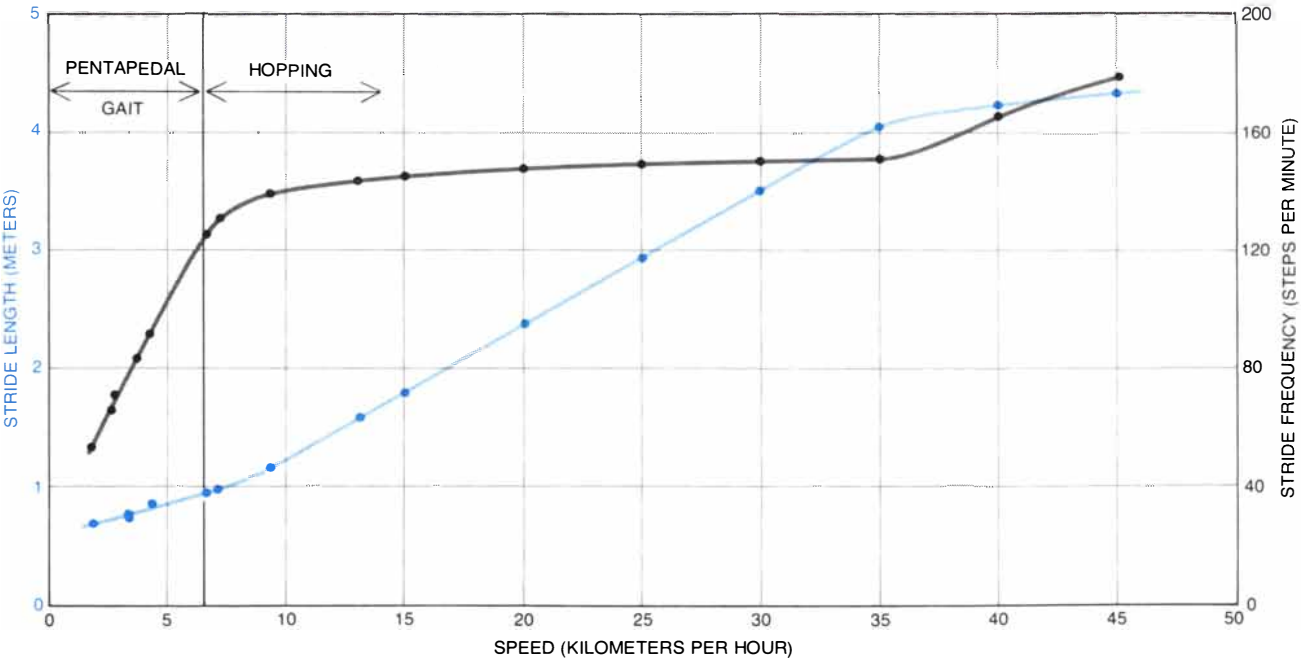
How can a low metabolic rate be reconciled with energetic hopping? Here two aspects of locomotion must be considered. One is maximum sustained speed for escaping danger. The other is economy in long-distance travel. The investigations by Taylor and his co-workers suggest that the costs of quadrupedal locomotion are more or less fixed. Maximum sustained energy consumption may be estimated from maximum oxygen uptake, which in turn appears to be related in some way to basal metabolism. Since marsupials have a low metabolic rate, it is possible that they have a low maximum energy output and consequently a reduced potential for speed of locomotion. The bipedal hopping of kangaroos and other macropodids could be a mechanism that helps to overcome this limitation. Also important for a metabolically limited animal, particularly for a kangaroo ranging over the semiarid interior of Australia, would be economy in long-distance travel.

Past assumptions about the kangaroo's limited ability to regulate its body temperature were based not only on the slightly lower body temperatures of marsupials but also on observations of a kangaroo behavioral trait: thermoregulatory licking. When kangaroos are chased on hot days, they occasionally stop hopping and lick themselves, particularly on the forelegs. Because licking has been generally accepted as a primitive mode of cooling, this behavior helped to sustain the notion that kangaroos cannot maintain a constant body temperature. Such a view would have been difficult to hold if those who held it had ever observed desert kangaroos such as the red kangaroo in their natural habitat.

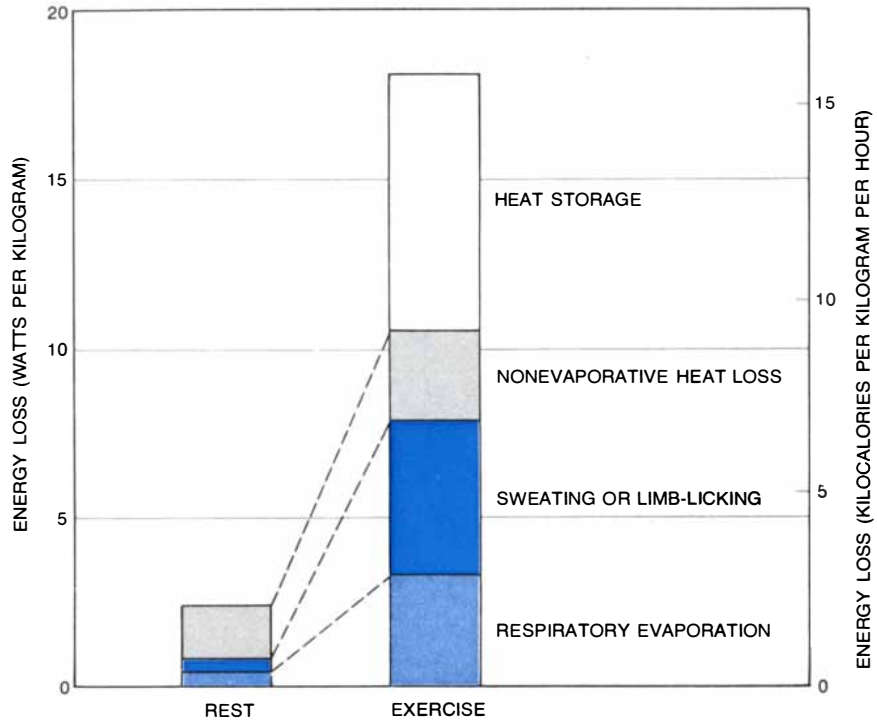
It is hard to conceive of an inefficient temperature regulator spending long summer days in the sparse shade of a small desert tree when the air temperature can be in excess of 45 degrees C. (113 degrees F.) and the radiation heat load makes the effective environmental



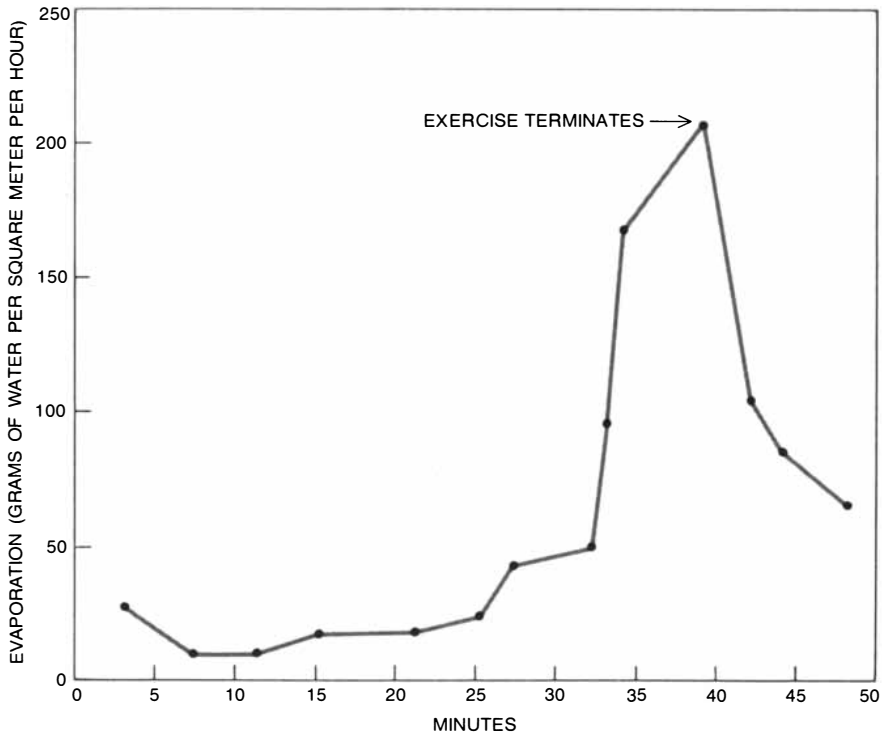
ENERGETIC COST of locomotion, measured in terms of oxygen consumption at various speeds, was determined by placing kangaroos on a treadmill that made it possible to simulate natural gaits at velocities up to 23 k.p.h. (14 m.p.h.). The energetic cost of pentapedal locomotion increased rapidly, but once the animals began to hop the cost leveled off and even decreased. The straight line (*color*) plots the predicted cost increase with speed for a quadruped of the same body weight. Beginning at a speed of about 17 k.p.h. (11 m.p.h.), hopping appears to be more economical of energy than quadrupedal locomotion is. Projecting the curve for the kangaroo (*broken line*) beyond the maximum treadmill speed suggests that this may be true up to the top limit for sustained hopping: approximately 45 k.p.h. (28 m.p.h.).



HOPPING SPEED is a function of frequency and length of stride. The frequency increases rapidly as the speed of pentapedal locomotion rises, reaching a rate of two steps per second before the kangaroo begins to hop. From a hopping speed of 15 k.p.h. (10 m.p.h.) up to a speed of 35 k.p.h. (22 m.p.h.), however, the increase in stride frequency is trivial. In this range the main factor in increasing speed is an increase in length of stride (*color*). At speeds above 35 k.p.h. the kangaroo increases both the length and the frequency of its stride.



HEAT BALANCES of kangaroos at rest and exercising strenuously at the moderate air temperature of 24 degrees C. (75 degrees F.) are compared. Nonevaporative channels of heat loss suffice to maintain a stable balance in the resting kangaroo; skin cooling and respiratory-tract water loss combined are only half as important. The exercising kangaroo cannot significantly increase nonevaporative heat loss, and even sharply increased evaporative heat loss is not sufficient to dissipate the load. The "stored" excess is dissipated when the kangaroo stops exercising.



SWEATING not only is directly correlated with exercise in kangaroos but also requires a warm-up period. The graph records the outbreak of sweat on the face of a kangaroo forced to engage in pentapedal locomotion on a treadmill moving at a rate of 4 k.p.h. (2.4 m.p.h.). Some 30 minutes after starting to exercise the kangaroo began to sweat copiously. When exercise ended some 10 minutes later, sweating also diminished, virtually ceasing in 10 more minutes.

temperature even higher. It has been found that far from being inefficient these kangaroos have possibly the most efficient and best organized thermoregulatory defense against overheating to be found in any mammal. Kangaroos not only lick their forelimbs during heat stress but also pant and sweat. The interrelation of these different methods of evaporative heat dissipation is seen only during exercise. When the kangaroo is resting, it dissipates heat primarily by panting; it does not sweat while resting, but under severe heat load it may lick its forelimbs.

The kangaroo's licking behavior was long a puzzle to my colleagues and me because the forelimb area usually licked is small. It seemed that the overall heat-dissipation benefit would be doubtful in view of certain possible disadvantages. Probably more in frustration than for any other reason we undertook an examination of the blood supply to the forelimbs. Injecting the blood vessels in the forelimbs of a dead kangaroo with liquid latex, we found that there was a dense and intricate network of superficial blood vessels in the region the animals usually lick. Further study revealed that during heat stress the blood flow to the region is greatly increased. The forelimb region is thus a site of significant heat transfer. Indeed, by spreading saliva on their forelimbs kangaroos may well be making the most efficient use possible of an overflow of fluid from their respiratory system, the principal site of evaporative heat dissipation in resting animals.

The pattern of heat loss in exercising kangaroos is very different from that in animals at rest. The reason is that exercising kangaroos sweat. Early in the course of our investigations it was suggested that sweating replaced panting during exercise because panting is not possible when the animal is exercising. We soon realized, however, that evaporation from the respiratory system is still highly significant during exercise because when the animal is exercising, water loss from these surfaces is greatly increased. In fact, the increase raises respiratory evaporation well above the levels the animal achieves by panting when it is at rest. This increased respiratory ventilation is probably primarily in response to the increased demand for oxygen during exercise and is therefore probably limited by the level of oxygen intake. Body sweating, then, appears to bring into play an additional surface area for further evaporative heat loss. Licking is not possible when the animal is hopping, although hopping kangaroos do stop occasionally and spread saliva on their forelimbs. Licking after exercise seems to assist the rapid return to normal body temperature.

An aspect of this picture that sets kangaroos apart from other mammals is that sweating stops as soon as a kangaroo stops exercising, even when its body temperature is elevated and it is panting and licking its forelimbs. Other mammals capable of sustained exercise, such as horses and men, not only sweat to dissipate heat when they are exercising but also sweat to dissipate it when they are at rest. A few mammals, such as cattle, sheep and other bovids, sweat and pant simultaneously. What does the kangaroo achieve by ceasing to sweat when it is at rest and starting to pant? The answer lies in water economy. An animal that lives in an arid environment must husband its body water, and water is better conserved when it is evaporated from the respiratory tract than when it is evaporated from the skin. Evaporation from the skin results in a lowered surface temperature, which leads in turn to a greater inflow of heat from the environment and the utilization of additional water to eliminate the additional heat. Panting, of course, does not lower the body surface temperature.

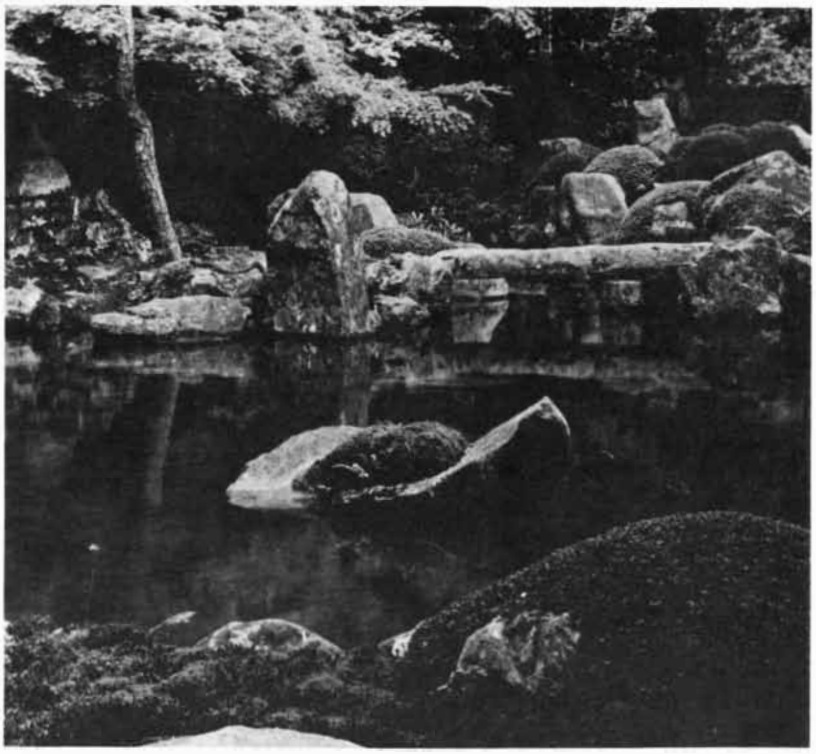
On a very hot summer day the red kangaroo is an example of an animal doing the most it can to both regulate its body temperature and conserve its water. When the temperature is higher than 45 degrees C., the red kangaroo does not lie down in the shade but stands hunched, thus presenting the smallest amount of surface area for the uptake of heat from the environment. The animal's dense fur provides an almost ideal insulation against the ambient heat. To further minimize the exposure of surface area to heat radiation the kangaroo's long, thick tail is pulled between its legs. The tail too has a complex network of superficial blood vessels that are involved in heat dissipation, but when conditions favor heat flow into the body, this system is apparently nonfunctional.

One stratagem for conserving water under arid conditions that has not evolved among kangaroos is the ability to let the body temperature rise during the day. The oryx, a ruminant of Africa and the Middle East, has a special arrangement of blood vessels in the base of the brain that allows the brain to be maintained at a tolerable temperature even when the temperature of the rest of the body rises as high as 46 degrees C. (115 degrees F.). Camels presumably have a similar arrangement. Hence both of these desert animals are able to store heat, which they can then dissipate in the cool of the evening without expending water in evaporative cooling. Kangaroos do not have this vascular arrangement. Under conditions of heat stress and dehydration, however, the kangaroo lowers its temperature overnight, so that it can start the next day with its body temperature from two

Freeman Books on Land and Landscape

Examine them free for 15 days.

Sierra Nevada, photograph by Shelton from *Geology Illustrated*



Rocky Mountain vegetation, and Garden of Rengeji Temple, Kyoto from *Plants in the Landscape*

GEOLOGY ILLUSTRATED

John S. Shelton, Drawings by Hal Shelton

Geologist/pilot Dr. John Shelton has been taking aerial photographs since 1939, and this is his master work. *Geology Illustrated* is a breathtaking view of our home planet: its rocks, minerals, volcanoes, glaciers, lakes, rivers, earthquake faults, mountain ranges, and other features. The nearly 400 photographs and crystal-clear text provide an enjoyable way to understand geology. Shelton addresses the book to "thoughtful people who enjoy the outdoors"—to hikers, rockhounds, pilots (and passengers), environmentalists, and many others. "An artist of aerial photography, Shelton uses some 400 of his finest photos to illuminate a discussion of the whole-earth system. Not a traditional textbook, but a fascinating exploration of problems posed by asking, 'How did that come about?' Worth buying for the photos and book design alone, but you'll probably find yourself becoming interested in geology regardless of your original intentions."
—*Whole Earth Catalog,* Summer 1971

1966, 434 pages, 382 illustrations, $14.50

PLANTS IN THE LANDSCAPE

by Philip L. Carpenter, Theodore D. Walker, and Frederick O. Lanphear, all of Purdue University

Plants in the Landscape is a unique introduction to the principles and practices of ornamental horticulture in landscape architecture. No other single volume covers such a broad range of topics. Here is a wealth of information on subjects ranging from the landscape practices of ancient civilizations to nononsense, dollars-and-cents descriptions of the present-day landscaping industry. The focus throughout is on landscape plants: how to select and maintain them for healthy growth in a wide range of settings.

Plants in the Landscape is a beautiful book. The text is closely integrated with more than 400 illustrations (about 300 photographs). Whether you are building a new home or remodeling an old one, whether you are a professional landscaper or a weekend gardener, you'll find *Plants in the Landscape* a valuable addition to your library.

1975, 481 pages, 534 illustrations, 32 tables, $16.00

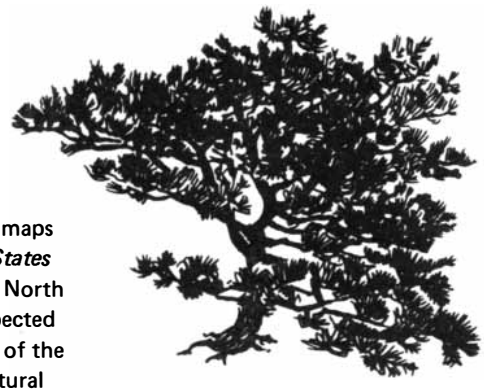
NATURAL REGIONS OF THE UNITED STATES AND CANADA

Charles B. Hunt, The Johns Hopkins University

With its generous use of photographs, line drawings, and maps (558 illustrations in all), *Natural Regions of the United States and Canada* provides a fascinating environmental tour of North America. Charles B. Hunt, a widely experienced and respected scholar, describes and explains the physical environment of the North American continent, Puerto Rico, and Hawaii. Natural regions are discussed in terms of their topography, geology, climate, soil, water, plants, and the interrelationship among all these factors.

"Just open this book to an area you already know. Within pages, the landforms, soils, colors, plants, and importance of your watershed will intensify. An encyclopedia of physiography written in academic style. Lots of maps. Lots of geologic history. The best book to practice mapping North America into your brain."—*Whole Earth Epilog,* 1974

1974, 725 pages, 558 illustrations, 22 tables, $15.95



W. H. FREEMAN AND COMPANY
Dept. SA, 660 Market Street, San Francisco, CA 94104
58 Kings Road, Reading, England RG1 3AA

Remembrance of Things Past

The Long House, the Pow-Wows, the ancient rites of summer, the Jesuit Mission of three centuries past, the villages where Loyalists first cleared the land, tilled, harvested, ground their grain, baked their loaves, dyed their wools, spun and loomed, the log stockades and a great Limestone Citadel that were outposts of Empire, the rattle of rifle-fire and thunder of cannonades, the bustle and bundling of the Fur Trade, the antics of the Voyageur Rendez-vous Days, the gentle rhythms of life as it used to be unfold again every summer. Treat the kids. For more about our living heritage, call COLLECT (416) 965-4008, or write: Ontario Travel, Dept. T.P., Queen's Park, Toronto M7A 2E5. We'll send you our Traveller's Encyclopaedia.

Ontario Canada

to four degrees below the normal level.

As a result of these combined adaptations for water economy the minimum water requirements of kangaroos are similar to those of the desert ungulates of Africa. Field studies indicate that arid-zone kangaroos turn over only a fourth of the water needed by sheep and wild goats. The oryx is completely independent of drinking water; it can get the water it needs from the plants it eats and from the oxidation of foodstuff. The kangaroo, like the camel, does need to drink occasionally under severe heat conditions.

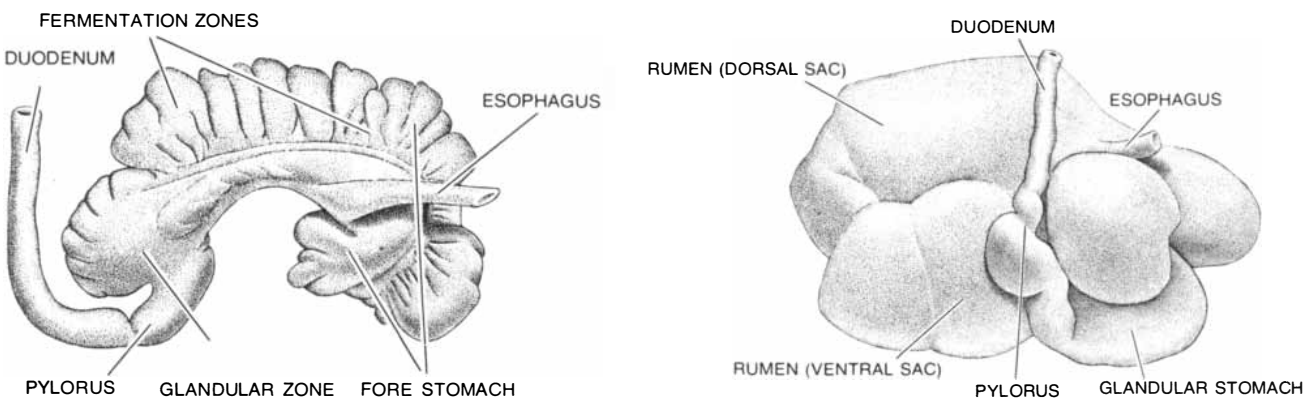
The persistent notion of primitiveness colors many perceptions of the physiology of marsupials. Consider the fact that the kidney of the arid-zone kangaroo has concentrating abilities superior to those of the kidney in comparable placental mammals. This should not really surprise anyone; the fundamental characteristics of the mammalian kidney appear to be very old. Even monotremes have a good kidney of the mammalian type, and it may well be that the reptilian ancestors of the mammals, the therapsids, also had such a kidney. Kangaroos too compare well with placental mammals in other areas of water economy. For example, kangaroos and camels resemble each other in their cardiovascular response to dehydration: when they lose water, they adjust the volume of the vascular system so that the functioning of the heart is not jeopardized.

Adaptations involved in feeding and digestion provide interesting evolutionary comparisons between kangaroos and ruminants. Both lineages have arrived at a similar solution to the problem of living off plant tissues. It has often been suggested that the evolution of the mammals was closely associated with the simultaneous evolution of flowering plants. Early in the Cenozoic era these plants were largely woody types. By the Eocene epoch, however, herbs and grasses had begun to spread, and in the Miocene, beginning some 25 million years ago, large areas of grassland covered the drier interiors of most continents. The evolution of grazing kangaroos appears to be associated with this spread of grassland in Australia, as does the expansion of the artiodactyl group of placental mammals elsewhere in the world.

No mammal can synthesize the enzymes that break down cellulose, the principal carbohydrate of grasses. The utilization of this food source calls for the evolution of a symbiotic association between the mammal and species of bacteria and protozoa that have such enzymes. The microorganisms take up residence in various expanded sections of the mammalian gut. In horses and rabbits they live in parts of the hindgut, such as the caecum. In ruminants they live in sections of the foregut and in specialized enlargements of the esophagus and the stomach. There are marked advantages in allowing the plant material to ferment in such chambers before it reaches the stomach: both the products of fermentation and the microorganisms can be digested and absorbed along the entire length of the small intestine. The greater efficiency of this system may account at least in part for the fact that whereas the dominant herbivores of the Eocene were hindgut fermenters such as horses, these animals were displaced in the Miocene by a massive radiation of fore-stomach fermenters such as ungulates.

Kangaroos are fore-stomach fermenters, so that they share with such modern ruminants as cattle, sheep, goats, deer and camels the benefits of this type of digestion. Those benefits include a lower requirement for protein. Urea, formed by the breakdown of protein in the body, is returned in the saliva to the fore-stomach, where it is resynthesized into nutritive protein by the resident microorganisms instead of being entirely excreted in the urine and lost. This recycling enables kangaroos and ruminants not only to gain nourishment from the dry, low-protein grasses of the arid zone but also to conserve the body water that would otherwise be needed for the excretion of urea. The combination of this efficient mode of digestion and the low rate of marsupial metabolism gives the kangaroo a formidable array of characteristics for the successful exploitation of a dry grassland environment.

The picture of the kangaroo that has now been developed is one of a group of mammals that are not at all primitive but have adapted and radiated rapidly in recent times in response to a new and changing environment. That kangaroos still possess some characteristics that may be considered primitive is true, but it is also true that they have adopted ways of circumventing the limitations of these ancestral traits. Moreover, kangaroos, like other marsupials, have evolved abilities that surpass those of comparable placental mammals. Marsupials and placentals are close relatives; their separate evolution over 130 million years illuminates how there can be different evolutionary solutions to similar problems. It is a large jump from the tiny shrewlike mammals of the early Cretaceous to the hoofed mammals of the Old World and the Americas on the one hand and the kangaroos of Australia on the other. Historical records suggest that there are more kangaroos in Australia now than there were before the Europeans arrived. Thus it may be that, far from traveling a road to extinction, at least the large kangaroos of the arid zone are thriving in spite of competition with man and the ruminant stock he has introduced. This is probably attributable to an extension of a suitable arid habitat. That extension, however, has simultaneously reduced the habitat suited to the kangaroos' smaller relatives, the wallabies and the rat kangaroos. It is these marsupials that have largely disappeared from the interior of Australia.



DIGESTION OF CELLULOSE, made possible by the action of enzymes secreted by stomach bacteria, is a capability independently evolved both by the marsupial kangaroos and by the placental ruminants. At left is the stomach of a red kangaroo, at right the stomach of an ox. The cellulose present in the diet of both grass eaters is largely broken down by fermentation before it passes into the intestine.

The Gas Vacuoles of Blue-Green Algae

The cells of many water-dwelling species of these bacteriumlike organisms contain thousands of tiny cylindrical structures filled with gas. These structures serve to regulate the alga's buoyancy

by A. E. Walsby

Among the many hundreds of species of primitive plants grouped under the term blue-green algae at least 50 are known to contain bright, irregularly shaped structures called gas vacuoles. These structures are found only in prokaryotic microorganisms, that is, bacteria and blue-green algae (which are also known as cyanobacteria). Moreover, they are found only in the free-living prokaryotes that inhabit ponds, lakes and the ocean; they are generally absent from the prokaryotes of streams, the soil and all other terrestrial habitats. This distribution provides a clue to the function of the gas vacuoles, which is to provide buoyancy and to regulate it. For organisms such as blue-green algae, which live in the top five meters or so of the water and depend on sunlight for photosynthesis, the ability to regulate buoyancy is a valuable asset. The details of the structure of the gas vacuoles and of the elegant way these microscopic organelles are regulated by the blue-green algae have now been revealed, after nearly a century of investigation.

The investigation of the function of the gas vacuoles began in 1895, when the German microbiologist Hans Klebahn reported a remarkable experiment he had performed with a suspension of blue-green algae skimmed from the surface of a lake on which an algal "bloom" had formed. He had filled a stout glass bottle to the brim with the suspension, had inserted a cork in such a way that no air space or visible bubbles remained in the bottle and had then struck the cork with a hammer. Three things happened. First, the appearance of the algae changed immediately from a milky green to a dark, translucent green. Second, bubbles of gas collected under the cork. Third, the algae gradually sank to the bottom of the bottle. (Algae that had not been treated in this way floated to the surface.)

Examination of the suspension under the microscope showed that the pressure applied by the blow of the hammer had caused the disappearance of numerous bright, refractile granules from the algal cells. Klebahn concluded that the granules must have been filled with the gas released from the suspension, and he argued that their destruction accounted for the solution's change in turbidity and for the algae's loss of buoyancy. He called the structures gas vacuoles.

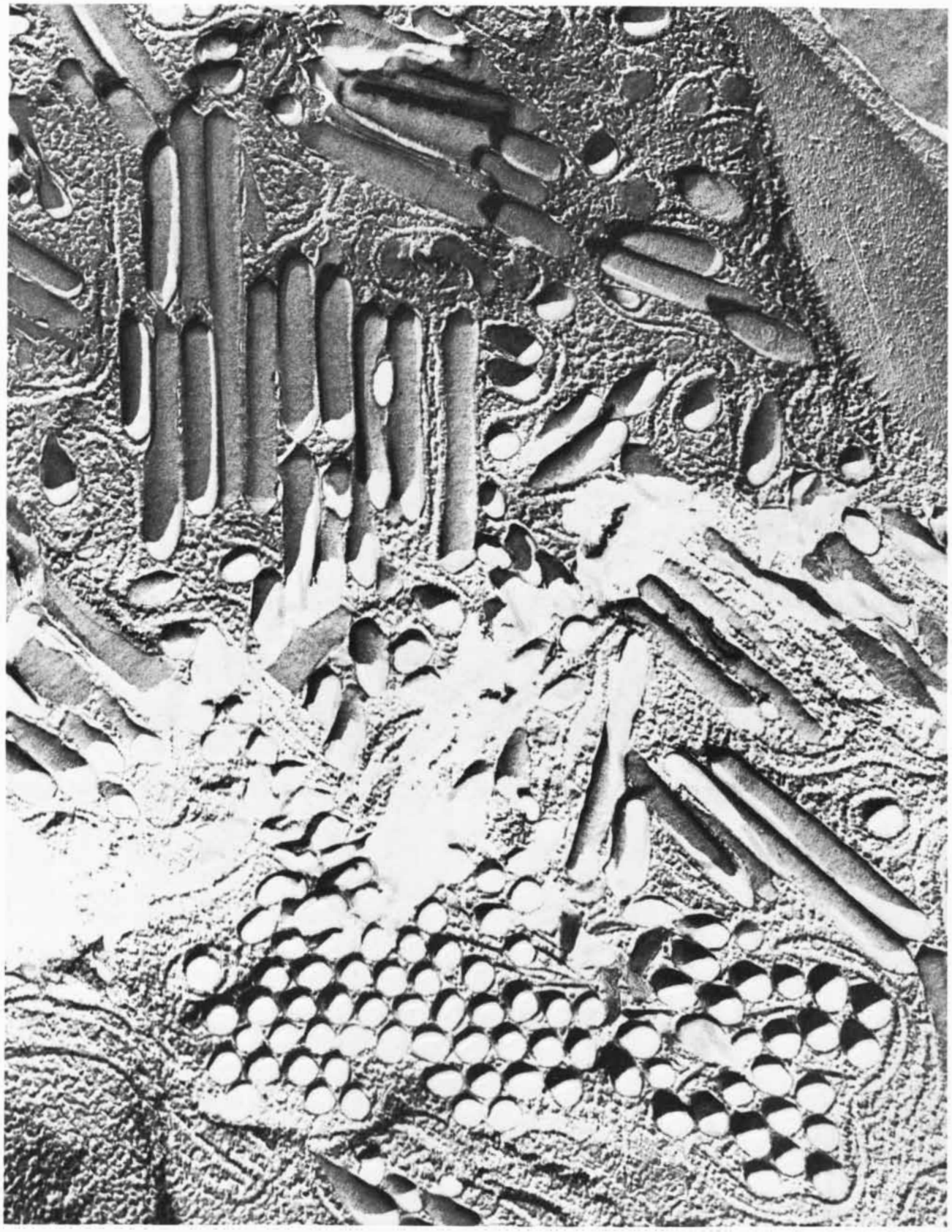
Over the next 30 years Klebahn tried by a number of means to extract the vacuole gas and determine its composition. In this work he assumed that an impermeable membrane surrounded the vacuole and prevented the gas from escaping. He made the assumption to account for the fact that the vacuoles could not be caused to disappear by exposure to a vacuum. (If a solution containing ordinary gas bubbles is placed in a vacuum, the gases diffuse out of the bubbles, which then disappear when the solution is returned to atmospheric pressure.) The mistaken concept of the vacuole as a gastight structure lasted for more than 50 years. It seems to have focused attention on the nature, mode of formation and function of the vacuole gas rather than on the structure of the vacuole itself, with the result that during this period little progress was made in gaining an understanding of the structure.

In 1965 C. C. Bowen and T. E. Jensen of Iowa State University published in *Science* an account of their study of the fine structure of vacuoles in blue-green algae. They found that each algal cell contained a large number of minute hollow cylindrical structures with conical ends. Bowen and Jensen proved that these structures, which they called gas vesicles, were the building blocks of the gas vacuoles by demonstrating that the vesicles disappeared under pressure, leaving membranous sheets in their place. Evidently the hollow structures had collapsed flat.

I heard about this discovery when I was working in G. E. Fogg's department at Westfield College of the University of London, where a number of us were investigating the physiology of blue-green algae. Discussion of the article prompted me to reexamine the problem of the gas in the vacuoles. Fogg provided me with his translations of the early German writings on the topic, which he had reviewed 25 years earlier. There I read Klebahn's conclusion that the gas was mainly nitrogen. With his chemical analysis he had not been able to eliminate the possibility that the inert gas argon was also present. It was important to check this point, because argon is not a product of metabolism; if it turned up, it could only have entered the vacuole by diffusing into it from the surrounding solution.

At that time I was studying the fixation of nitrogen, employing a mass spectrometer to analyze the proportions of various nitrogen isotopes. With this instrument, I thought, it would be easy to distinguish nitrogen from argon in the gas from vacuoles. To obtain the gas I modified one of Klebahn's techniques, first putting a concentrated suspension of algae into a vacuum to remove dissolved gas and then heating the suspension in a sealed vessel to break down the vacuoles. To my dismay I could detect little or no gas.

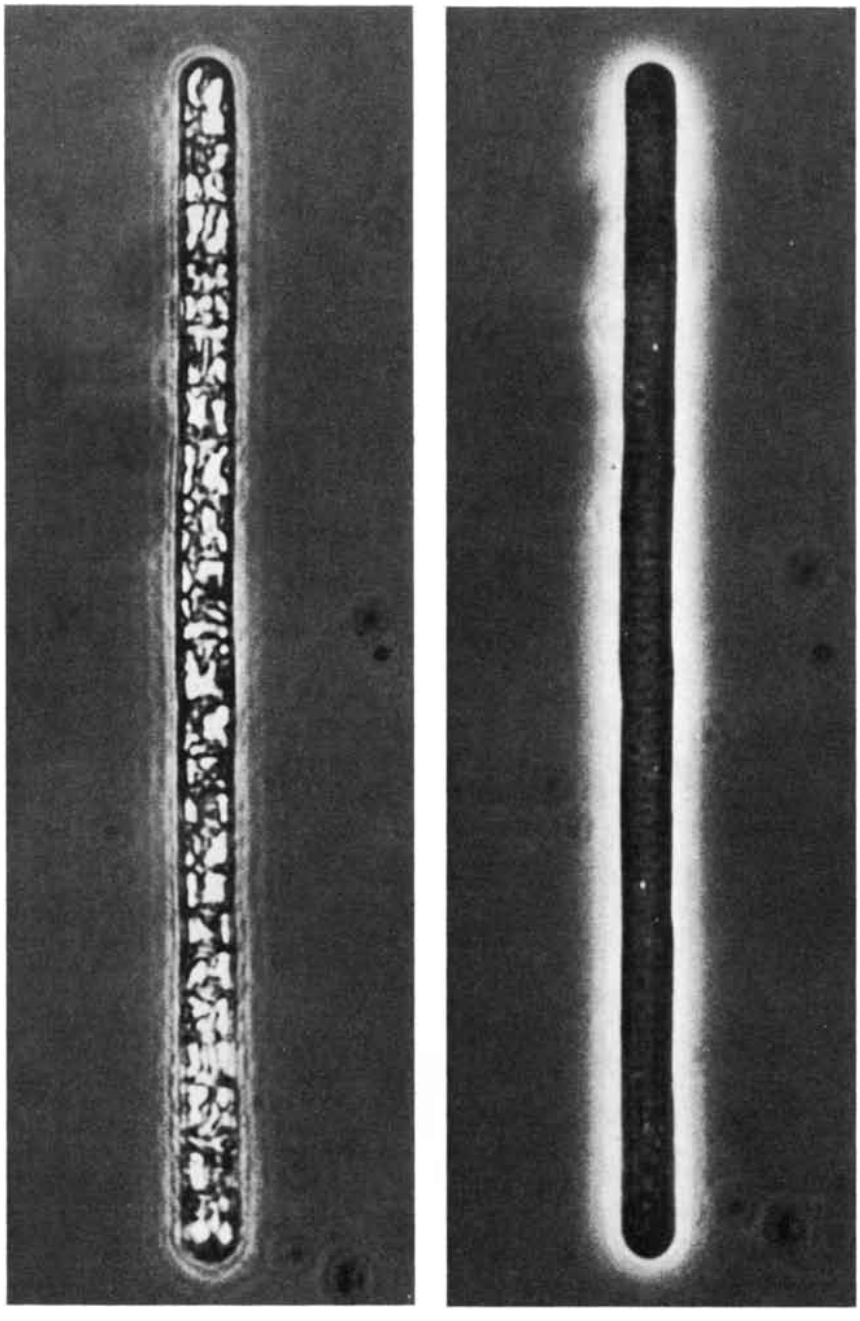
Eventually I realized that the gas in the vacuoles had been removed by putting the algae into a vacuum. In other words, the gas vacuoles (more precisely, their constituent vesicles) were not gastight but were freely permeable to gas. Ironically, I made this discovery not with an instrument of modern technology but with a device from Klebahn's era: the Warburg respirometer. With it I measured the volume of gas that could



GAS VESICLES in the blue-green alga *Anabaena flos-aquae* are enlarged 89,000 diameters in this electron micrograph. The material was frozen and then fractured; in the process some vesicles were broken along their length and some across it. The cigar-shaped structures near the top of the micrograph are vesicles fractured lengthwise, and the small circles near the bottom are vesicles fractured crosswise. Such a group of vesicles constitutes a gas vacuole. The micrograph was made by Daniel Branton, who is now at Harvard University.

be obtained from vacuoles in a concentrated suspension of algae. I found that the volume was always directly proportional to the pressure of the gas dissolved in the water in which the cells were suspended. (The pressure of the dissolved gas is known to be the same as the pressure of the overlying gas when the system has been equilibrated by thorough shaking.) By suitably modifying the Warburg apparatus I was able to follow the exchange of gas across the gas-vesicle membranes. If the pressure of the overlying gas was raised, more gas dissolved in the water and diffused through the porous membrane of the gas vesicle into the space inside it; if the pressure was lowered, the flow was reversed. In this way I was able to demonstrate the permeability of the membrane to a number of gases.

This discovery showed that it was pointless to try to extract the gas from vacuoles, since it would always be of the same pressure and composition as the gases dissolved in the surrounding solution. The finding also presented a number of new questions. What prevented the gas vesicle from collapsing under moderate pressure? How did the hollow space form? Why did it not fill up with water? Since the vesicle could not store particular gases, what was its function? To find the answers I have had to delve into many branches of biology, from biochemistry to ecology.

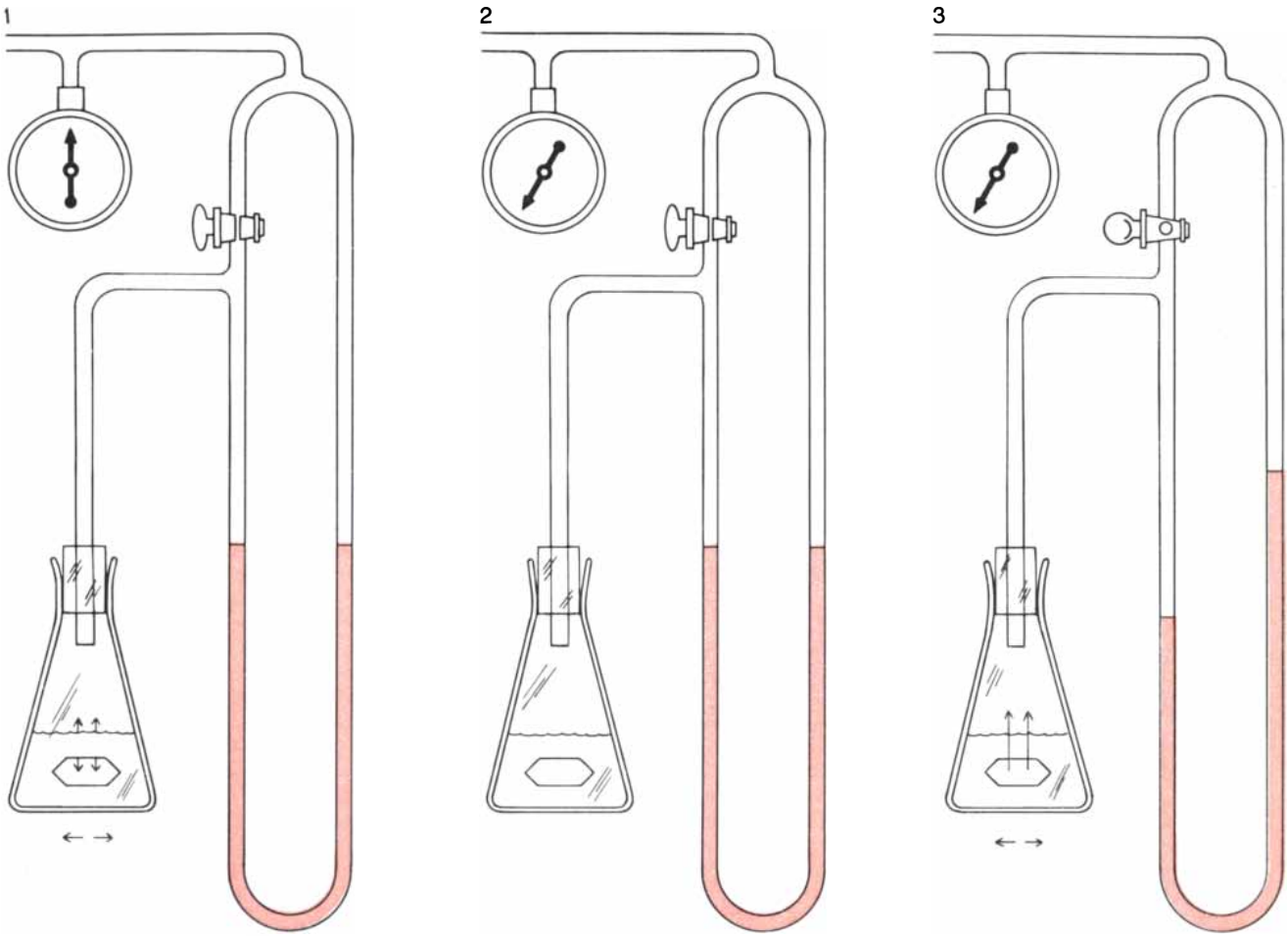


FILAMENT OF ALGAE appears with and without gas vacuoles in these phase-contrast micrographs. A filament is an assemblage of several algal cells. In the micrograph at the left the vacuoles are visible as bright granules, each of which is made up of several hundred hollow cylindrical vesicles. Although the vesicles are too small to be seen under the light microscope, they are made visible because the gas-filled spaces they contain scatter light. The micrograph at the right shows the same filament after it was subjected to a pressure of 12 atmospheres. The pressure caused the vesicles to collapse, so that they are no longer visible. The filament (blue-green algae of the species *Oscillatoria agardhii*) is enlarged approximately 1,400 diameters.

Klebahn had argued correctly that the gas in the vacuole must be enclosed in a rigid membrane because the gas could be at a pressure lower than the hydrostatic pressure in the algal cell. (The sugars, amino acids, ions and so on dissolved in the water inside the cell generate an osmotic pressure, so that water from the outside tends to diffuse into the cell and the cell tends to swell up. The swelling is resisted by the cell wall, which generates a back pressure: the hydrostatic or turgor pressure.) Moreover, under the light microscope Klebahn could detect no change in the size of a gas vacuole when he applied a moderate pressure. It does not necessarily follow that the volume of the constituent vesicles is not altered, but I have confirmed that they show little elastic deformation with changing pressure. For one thing, algal filaments (strings of algal cells) or colonies that are buoyant with gas vacuoles do not rise any more rapidly when the suspension is put into a vacuum. For another, the amount of light scattered by a suspension of intact, isolated vesicles shows virtually no reversible change when the pressure on them is increased or decreased.

An important feature in the rigidity of a gas vesicle must be its cylindrical shape, which is similar to the shape of diving submersibles or the cylinders made for storing commercial gases. Those structures are of course specifically designed to withstand pressure. At Westfield College, Henry Eichelberger and I found that gas vesicles maintain their form when they are isolated from cells, proving that the vesicles do not rely for support on the protoplasm that normally surrounds them.

An everyday analogy may clarify the nature of the vesicle. It is not like a balloon, whose size is determined by the excess pressure of the gas in it. Rather, it resembles a rigid porous pipe with closed ends. The pressure of the gas inside is determined by the partial pres-



WARBURG RESPIROMETER was the apparatus with which the author determined that the membrane of the vesicle was permeable rather than gastight. The instrument consists of a small flask attached to a glass manometer (a *U*-shaped tube containing colored water) by a gastight joint. Isolated gas vacuoles from blue-green algae were suspended in water in the flask at atmospheric pressure (*1*). The flask was shaken (*bottom arrows*) to bring the gas in the water and in the vacuoles to equilibrium (*upper arrows*). The apparatus was stopped and a vacuum was rapidly applied (*2*) to minimize the loss of gas from the suspension by diffusion. Finally, the tap to the manometer was closed (*3*), so that gas evolved from the suspension in the flask pressed down on the fluid in the left side of the manometer tube. The change in pressure is calculated from the difference in the height of the fluid in the two sides of the manometer. A difference of one centimeter is equivalent to a pressure of .001 atmosphere. If the vesicles had been gastight, no gas would have been detected in manometer.

sures of the gases dissolved in the surrounding water and is independent of the hydrostatic pressure acting on the outside of the structure. The difference between these two pressures is borne by the vesicle wall, which is relatively strong but brittle, so that under sufficient pressure it collapses. The structure remains stable in a vacuum even though the gas diffuses out of it.

I decided to see how much pressure was required to collapse gas vesicles, a measurement that surprisingly had not been made previously. To a suspension of the gas-vacuolate alga *Anabaena flos-aquae* I applied pressure in increments of .5 atmosphere, measuring the turbidity of the suspension after each step. I found that at each increment above about one atmosphere there was a decrease in turbidity that could be accounted for by the collapse of a proportion of the refractile gas vesicles in each cell. At about 4.5 atmospheres no intact vesicles remained.

A vesicle would collapse instantly at a certain pressure, its critical pressure, but at lower pressures it would remain intact indefinitely. Each cell contained several thousand gas vesicles. The pressure at which they collapsed was distributed normally around a mean value of approximately 2.5 atmospheres.

From the outset I knew that the critical pressures I was measuring would underestimate the true strength of the vesicles because they were already sustaining the turgor pressure inside the algal cells. By suspending the cells in a concentrated solution of sucrose, which reversed the normal osmotic relation and caused water to flow out of the cells, I could remove the turgor pressure and measure the true critical-pressure distribution of the vesicles. It turned out to range from about 4.5 to 7.5 atmospheres.

A consequence of performing such measurements was that I was able to make the first reliable estimates of turgor pressure in prokaryotic cells. The calculation involved only a subtraction of the apparent critical pressure of gas vesicles from the true pressure. With this method I surveyed a number of organisms with gas vesicles and found an interesting correlation: in nearly every case the range of critical pressures was about one atmosphere above the turgor pressure of the cell. The difference provided an adequate margin of safety to protect the vesicles from collapsing under the hydrostatic pressure where the organism lives. I have recently come across an interesting exception in the marine blue-green alga *Trichodesmium*. It has extremely strong gas vesicles, which are able to withstand the pressure the alga experiences when it is carried to considerable depths in the ocean.

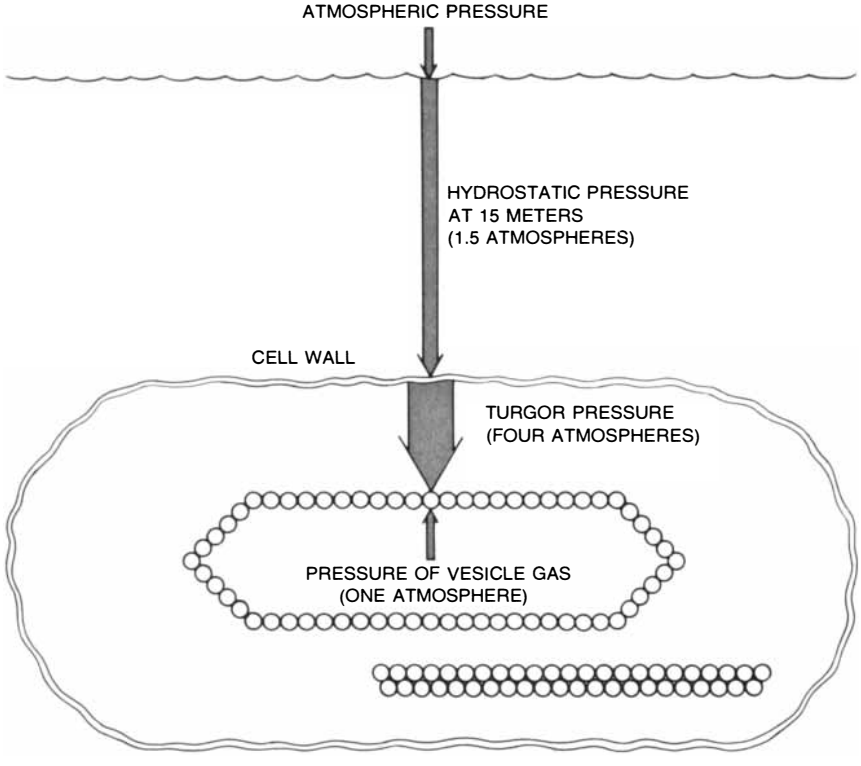
While I was making the measure-

ments of turgor pressure on *Anabaena flos-aquae,* I got an unexpected bonus that opened up a new field of gas-vacuole research for me. I noted that in the concentrated sucrose solutions the algal cells burst open, apparently because their shrinkage in volume caused the cell walls to buckle. From each cell I saw myriad vesicles, each one a tiny point of light under the phase-contrast microscope, come jostling out. This observation formed a basis for isolating intact gas vesicles from the cells. (The traditional methods of breaking cells generate forces that cause the fragile vesicles to collapse.)

I learned from Pål Falkenberg, who was working in Helge Larsen's laboratory at the Technical University of Norway on gas vesicles in halobacteria (bacteria that thrive in a salty environment), that the vesicles could be separated from the other components of the cell by centrifugation. When the vesicles were spun in the centrifuge, they would float to the surface, where they could be creamed off. The principal problem was that centrifugation generates a pressure that could collapse the vesicles. They would then sink. Since I could determine the minimum critical pressure of the gas vesicles in my preparation, however, I could make sure that this pressure was not exceeded in the centrifuge and so could get a good yield.

By centrifuging the vesicles and then putting them through a membrane filter Barbara Buckland (a graduate student) and I obtained highly purified preparations of gas vesicles. We then set about analyzing their chemical composition, with surprising results. Larsen's group in Norway and Walter Stoeckenius' group at the University of California School of Medicine at San Francisco, working with partially purified vesicles from halobacteria, had reported that protein was an important component of the vesicle. We could not find anything but protein. Clearly the vesicle's membrane, if it can be called that, was different from typical membranes in the cell, which often contain as much as 50 percent by weight of lipid, or fat.

While we were doing this work, Daniel D. Jones and Michael Jost of Michigan State University were isolating gas vesicles from the blue-green alga *Microcystis aeruginosa.* Soon after we had given a brief report on our work in *Nature* they published an account of their more thorough analyses, which not only confirmed that protein is the sole constituent of the vesicle but also showed that only a single type of protein molecule is present. It appeared to be relatively small (14,000 daltons). By careful measurements with the electron microscope these workers produced evidence that the wall or membrane of the vesicle is only one protein molecule thick, the molecules being arranged in rows along the ribs that form the cylindrical structure. From these studies emerged a picture of the gas vesicle as a biological structure of unparalleled simplicity. X-ray-diffraction studies by Allen Blaurock, working with W. Wober at the University of California School of Medicine at San Francisco and with me at the University of London, have revealed details of the arrangement of the protein molecules in the membrane.

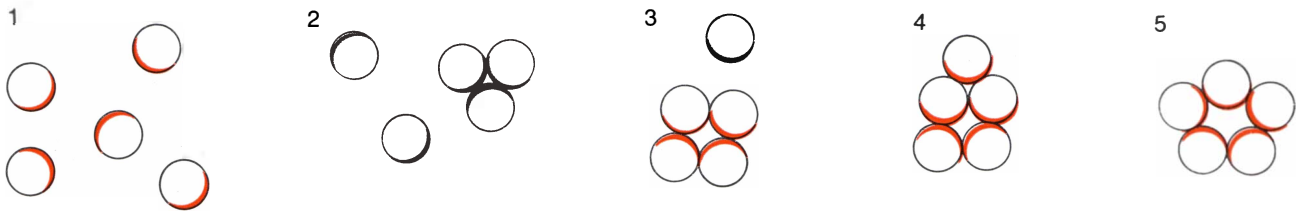


PRESSURE RELATIONS of a vesicle in an algal cell are depicted. The total pressure on the outside of the vesicle is 6.5 atmospheres: one atmosphere from the normal pressure at the surface of the water, 1.5 atmospheres from the water pressure at a depth of 15 meters and four atmospheres from turgor pressure, which is generated by the resistance of the cell wall to the tendency of the cell to swell on taking up water. The pressure of the gas in the vesicle is about one atmosphere, so that a pressure difference of about 5.5 atmospheres is borne by the vesicle wall. For each vesicle there is a point (the critical pressure) at which the difference exceeds the strength of the wall, causing the vesicle to collapse. In the intact vesicle in this schematic cell the critical pressure is six atmospheres; in the vesicle below it the pressure was five atmospheres.

This knowledge has made it possible to answer, at least in part, the questions I raised above about the properties, formation and function of the gas vesicles. Let us consider why water does not get inside the vesicle. It is not necessary to postulate that the protein wall is impermeable to water molecules; all that is needed is for the inner surface of the vesicle to be hydrophobic. That property would prevent water droplets from forming inside the structure and would generate sufficient surface tension to prevent liquid water from creeping through any pores there might be in the membrane.

Can a protein provide such a hydrophobic surface? Yes, by having its aliphatic amino acids (amino acids with a fatty side chain) appropriately exposed on that side. X-ray-diffraction studies by Blaurock indicate that such an arrangement may exist in the protein of the gas vesicle. Recent work by David Worcester of the Atomic Energy Research Establishment at Harwell in England has confirmed this arrangement. At Blaurock's suggestion he allowed preparations of my collapsed *Anabaena* vesicles, which had been dried on a flat surface, to absorb heavy water: water in which the hydrogen is not the common isotope of the element but the heavy isotope deuterium. The preparations were then exposed to a beam of neutrons, which could distinguish between the deuterium in the heavy water and the ordinary hydrogen in the surrounding material. It could be deduced from the neutron-scattering patterns that the heavy water did not penetrate between the two opposing halves of the collapsed membrane, that is, the surfaces that had faced the gas space in the intact structure. Ordinary water would behave the same way.

What of the outer surface of the gas



FORMATION OF VESICLE begins when molecules of protein, depicted here as spheres, come together in the algal cell. The protein has one hydrophobic (water-repelling) surface (*color*) and one hydrophilic (water-attracting) surface. Originally the molecules are suspended individually in the cell (*1*). As aggregation begins (*2 and 3*) the hydrophobic surfaces face inward, so that the hollow space that eventually forms (*4 and 5*) tends to exclude water. Once the hollow space has formed, gas diffuses into the vesicle from the surrounding water.

vesicle, which faces the water? It is more likely to be hydrophilic, thereby minimizing the pressure generated by surface tension, which would tend to collapse the structure. Several observations indicate that the outer surface of the gas vesicle is wettable. For example, if isolated gas vesicles are shaken with olive oil and water (something I did at Fogg's suggestion), they partition completely into the water.

How does the hollow space inside the gas vesicle form? When I realized that the permeability of the membrane ruled out the possibility that it was inflated by gas, I proposed that the vesicle must be a self-erecting structure. One can visualize its formation as follows. The starting point would be a small, solid cluster of molecules of the vesicle's membrane protein. The addition of more molecules would then push apart the first ones to create a hollow space. Gases could diffuse into the space from the surrounding solution. The energy necessary for the erection of the structure comes from the decrease in free energy of the protein molecules, and the mutual attraction of the molecules prevents them from falling into the hollow space. J. Robert Waaland and Daniel Branton of the University of California at Berkeley obtained evidence of de novo assembly by following stages in the growth of gas vesicles induced in the blue-green alga *Nostoc muscorum*. The formation began with a small biconical structure that, when it reached the appropriate diameter, grew by the creation and elongation of the central cylindrical portion.

The protein of the gas vesicle appears to have a purely structural role: creating and maintaining a hollow space. This is the space that performs the functions attributed to gas vacuoles. Several possible functions have been suggested by the events accompanying the destruction of vacuoles in the hammer, cork and bottle experiment.

The formation of gas bubbles in the experiment suggested that the function of the vacuoles was gas storage, but this possibility must now be discounted in the light of the discovery that the vacuoles are far too permeable to store any gas. The changed appearance of the algae in the experiment suggested a light-shielding function, in which the vacuoles would protect light-sensitive substances in the algal cells. Such a function might be of considerable value to blue-green algae floating at the water surface under the summer sun. The case for light-shielding, however, is not clear. If gas vesicles are efficient at shading, why are they not found in the algae that inhabit the surfaces of rock and soil? My view is that if the vacuoles have a light-shielding role, it is secondary to their principal function, which is to provide buoyancy.

The hammer, cork and bottle experiment clearly demonstrates that gas vacuoles do provide buoyancy. A consideration of their density shows that they do so with great efficiency. The density of the gas vesicle is about .1 gram per cubic centimeter. Because water fills the space between the many vesicles that make up a vacuole, the density of the entire vacuole is about .2 gram per cubic centimeter, still much lower than the lightest liquid or solid found in living cells.

The distribution of gas vacuoles among prokaryotic organisms provides strong circumstantial evidence that the buoyancy they provide is valuable. The vacuoles are found almost exclusively in aquatic planktonic (free-floating) microorganisms. The point is emphasized by the fact that, whereas bacteria with gas vacuoles are virtually unknown among the thousands of nonaquatic strains that are maintained in laboratories around the world, I found nearly 30 different species of bacteria with vacuoles in a single pond in Minnesota.

The commonest habitats for gas-vacuole bacteria are relatively stable lakes that become thermally stratified, as Alison Clark (a graduate student working with me) has recently confirmed in an extensive survey of bodies of water in Britain. During the summer, populations of these bacteria develop from the bottom mud and become suspended in the hypolimnion (the layer of cooler water at the bottom of a lake). The population of some bacteria peaks at a particular depth, where presumably the conditions most favor their growth. For example, photosynthetic bacteria require conditions of dim light and the absence of oxygen, which in such a lake they may find just below the thermocline (the boundary between the warm surface layer and the cooler water below it). On the other hand, planktonic blue-green algae may form stable populations toward the bottom of the epilimnion (the warm surface layer).

Such phenomena suggest that gas vacuoles may be doing more than simply lessening the density of these organisms; they may also be involved in actively regulating buoyancy, so that the cells can poise themselves at particular stations in the vertical column of water. The planktonic algae I have kept in culture for studying gas vacuoles gave me my first clue to how the regulation of buoyancy might be achieved.

I noticed that if the alga was grown under a low intensity of light, it was invariably buoyant, whereas under a high intensity of light it sank. I realized that if the alga behaved the same way in a lake, it would stratify at a depth with an intermediate intensity of light. On investigating further I found that the alga lost its buoyancy quite fast (within an hour or so) when I transferred it from a low light intensity to a high one.

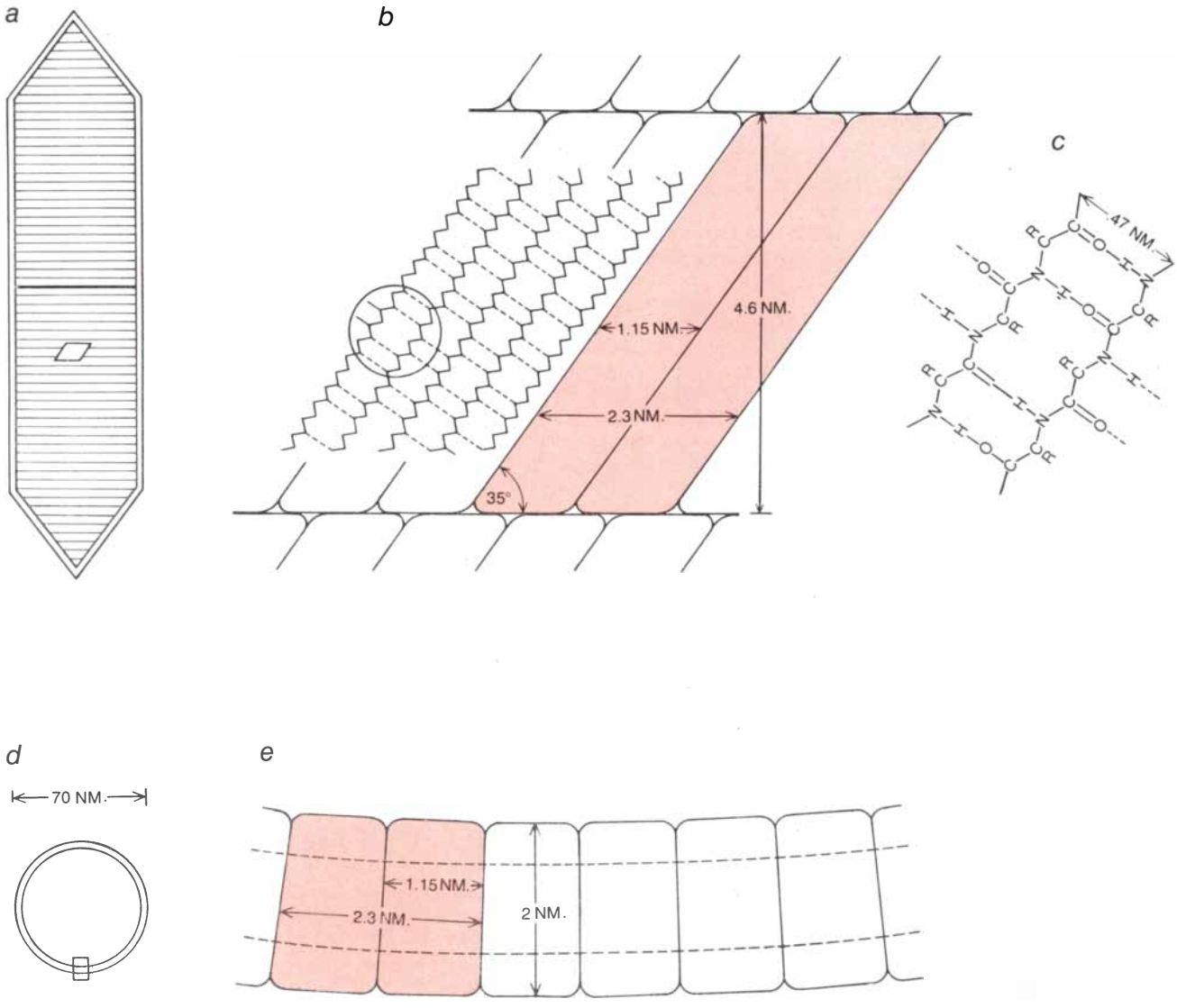
The next step was fortuitous. As I was trying to establish the range of turgor pressure in this alga by my gas-vacuole method I happened to notice that turgor was higher in algal cells left under high illumination than in cells left under low illumination. It occurred to me that in algae transferred to higher intensities of light the turgor pressure might rise enough to collapse the weaker vesicles, thus explaining the loss of buoyancy. At Westfield College, M. T. Dinsdale and I subsequently obtained direct evidence for the suspected rise in turgor and demonstrated that it was dependent on photosynthesis. We had begun to unravel the chain of events involved in this behavioral response.

During the past few summers I have studied blue-green algae in lakes to see if they regulate their buoyancy in this way. The evidence is that they do. For exam-

ple, Andrew Klemer of the University of Minnesota and I examined a lake in northern Minnesota that contained a permanent population of the blue-green algal species *Oscillatoria agardhii,* which during the summer showed a sharp peak at a depth of five meters, near the bottom of the euphotic zone (the zone where there is enough light to support the growth of green plants). Measurements with a pressure turbidimeter (a portable apparatus I designed for measuring gas vacuoles and the pressure at which they collapse) showed that the algae at that depth had just enough gas vacuoles to be neutrally buoyant. Was this accidental, or did the alga regulate its buoyancy?

We performed a simple experiment to show that the alga was indeed regulating its buoyancy. Drawing samples of the lake water from the five-meter depth, where the algal population was about 1,000 times more concentrated than it was at the surface, we put them in glass bottles that we stoppered, tied to a line and then suspended from a floating platform at various depths. After a day we took up the bottles and found that the algae in the bottle from five meters had remained evenly suspended. In the bottles tied near the surface the algae had sunk, and in the bottles from below five meters they had tended to rise. The differences persisted when the bottles were left at a constant temperature, showing that there had been a permanent change in the density of the algae. Subsequent observations suggested that the change in density resulted from a change in gas vacuolation, which in turn occurred in response to the decreasing intensity of light with depth. The gas vacuolation of the alga was apparently regulated by the turgor-pressure mechanism I have described.

The same mechanism is also involved in the diurnal vertical migrations of the alga *Aphanizomenon flos-aquae* in Lake Mendota in Wisconsin. I investigated the migrations with Allan Konopka of T. D. Brock's laboratory at the University of Wisconsin. Lake Mendota is too turbulent to become permanently strati-



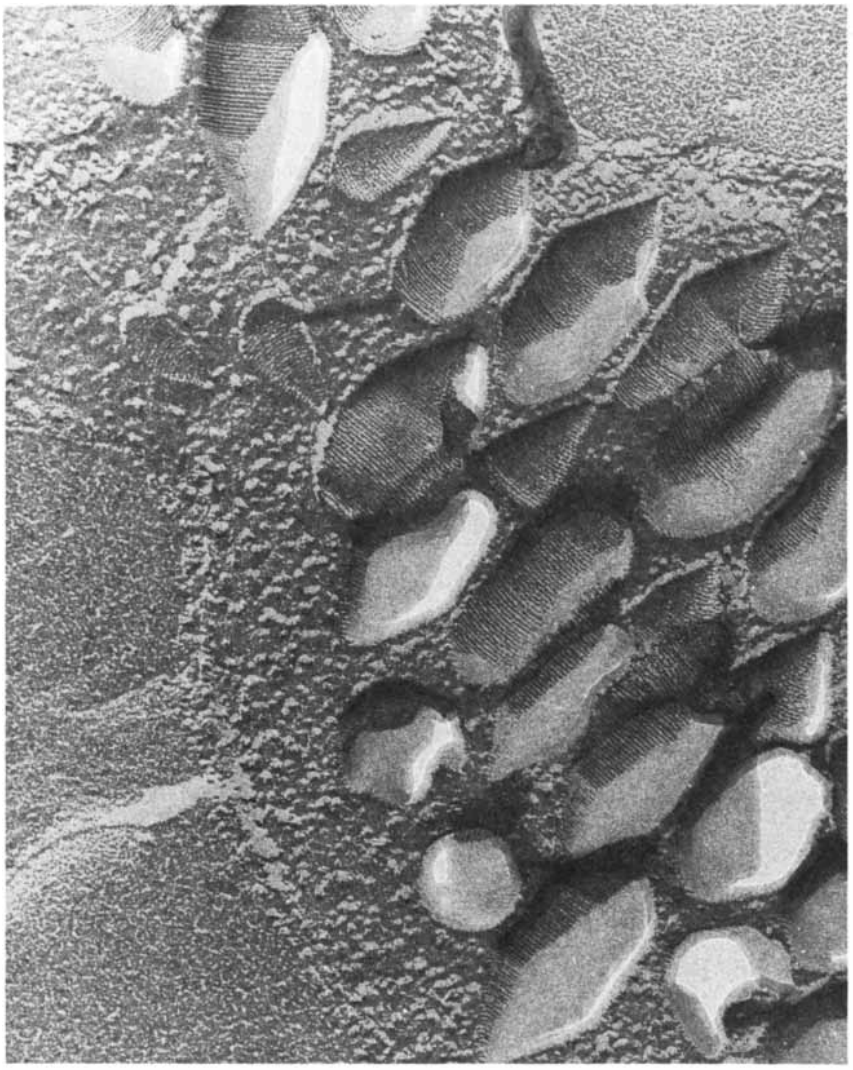
PROPOSED STRUCTURE of a gas vesicle in a blue-green alga is depicted on the basis of chemical analysis and X-ray-diffraction studies. At *a* the vesicle appears in longitudinal section. The walls are two nanometers thick and each rib is 4.6 nanometers wide. The rib area enclosed by the box is enlarged at *b,* where one sees that the rib is made up of unit cells 1.15 nanometers wide and sloping at an angle of 35 degrees. Each protein molecule probably incorporates two such unit cells (*color*). The fine structure of part of the molecule is shown at the left; its polypeptide chain is wound into eight parallel lengths in two layers, only one of which appears here. The circled area is shown in greater detail at *c,* where the chemical bonding of the molecule is depicted. The entire vesicle appears in cross section at *d,* and the area of wall enclosed by the box is reproduced at a larger scale at *e.* The two broken lines represent the two layers of the polypeptide chain. Area in color shows the putative protein molecule, which is 2.3 nanometers wide and consists of two unit cells each 1.15 nanometers wide.

fied, and the alga is usually suspended evenly within the epilimnion. During calm periods, however, the alga responds to the changing light intensity by floating up at night and sinking in the daytime. *Aphanizomenon* is able to do this because it forms large colonies that (for hydromechanical reasons) move much faster through the water than the tiny filaments of *Oscillatoria*. Another colonial blue-green alga, *Microcystis aeruginosa,* performs similar migrations, but perhaps by a different mechanism. The gas vacuolation may decrease at high light intensities because new cell material is formed faster than new vesicles are made.

These studies have led to an understanding of algal stratification and vertical migration. A third phenomenon, however, appears to run counter to these observations. It is the formation of water blooms, the thick accumulations of blue-green algae that appear at the surface of many lakes during the summer. Colin Reynolds of the Freshwater Biological Association in Britain has demonstrated that blooms do not form by the growth of algae at the surface; growth occurs earlier, while the algae are suspended in the epilimnion. The bloom forms during periods of calm weather when the algae, overbuoyant with excess gas vacuoles, float to the surface. The formation of the bloom is paradoxical in that the algae often die at the surface because of the depletion of mineral nutrients and the exposure to intense sunlight.

Reynolds and I, reviewing published material on algal blooms, concluded that the formation results from the failure of the algae to decrease their gas vacuolation. This failure suggests a failure to sustain high rates of photosynthesis at the surface, notwithstanding the high level of illumination there, since both methods of gas-vacuole regulation (turgor pressure and cell growth) depend on photosynthetic activity. We concluded that there are several causes for the failure of photosynthesis.

One of them has been suggested by experiments Michael Booker and I did in a large laboratory tank that, with overhead illumination and cooling at the bottom, is intended to model a stratified lake. We found that when a planktonic blue-green alga is grown in the tank under conditions where nutrients are limited, it stratifies at a depth where the light intensity is low. If the water is enriched with nutrients, particularly phosphate, the algal population increases and by its photosynthetic activity depletes the carbon dioxide dissolved in the water. Since photosynthesis requires carbon dioxide, further photosynthesis is limited, and the algae, becoming highly vacuolate, rise to the surface, forming a bloom. It is not known whether a similar explanation holds for natural waters, but it is significant that the lakes where blue-green algae stratify best are oligotrophic (having low levels of nutrients such as phosphate), whereas blooms tend to form in waters that are eutrophic (having high levels of nutrients).



GAS VESICLES in a bacterium, *Prosthecomicrobium pneumaticum,* are enlarged 139,000 diameters in this freeze-fracture electron micrograph made by Branton. A few water-dwelling species of bacteria have vesicles that are quite similar to the ones found in certain species of blue-green algae. One sees both the shape of the vesicle (cylindrical with pointed ends) and the ribs, only 4.6 nanometers wide, that the protein molecules of the vesicle wall form. The vesicles are 120 nanometers wide and up to 300 nanometers long; blue-green algal vesicles have a similar shape but are narrower (70 nanometers) and longer (up to 1,000 nanometers).

Water blooms can be a considerable nuisance. They interfere with water supplies by clogging filters and releasing substances that produce unpleasant tastes. They affect freshwater fisheries adversely. Some bloom-forming blue-green algae synthesize potent nerve toxins that can cause the death of livestock and wildlife. Blooms also spoil recreational waters.

In many instances these problems are of human making, since the eutrophic conditions that support water blooms result from the enrichment of lakes and reservoirs in nutrients from sewage and from fertilizers applied in excess to agricultural land. It is possible that a better understanding of the gas vacuole and its role in the biology of planktonic blue-green algae will lead to a means of controlling water blooms. My somewhat facetious suggestion that the development of such blooms might be checked by collapsing the gas vacuoles with explosives has been shown by D. C. Menday and A. A. Buck of the University of Wales to be economically viable, but it is of course unacceptable for other reasons. My own feeling is that the problem of water blooms can be solved only by reversing the eutrophication of the waters where they appear.

The History of the Airflow Car

As the first streamlined car, the Chrysler Airflow represented a milestone in automotive development. Although its sales were disappointing, it had a profound influence on automobile design

by Howard S. Irwin

The evolution of the automobile from the time of the "horseless carriage" to the present has proceeded in a series of steps that have been gradual and predictable, with the conspicuous exception of the Airflow, which was conceived by the automotive engineer Carl Breer and his associates and marketed by the Chrysler Corporation in the model years 1934 through 1937. The Chrysler Airflow and its sister car the DeSoto Airflow were utter failures in the showroom, but the impact of the design on the shape of the modern automobile has been profound.

The Airflow was Breer's attempt to achieve in automobile design what had been achieved in the same era by the DC-3 and the China Clipper in aircraft and by the Pullman Company's M-10,000 diesel locomotive and Henry Dreyfuss' steam-powered trains that ran as the 20th-Century Limited. In each case the designer's aim was to unify previously disparate design elements, not merely to please the eye by creating an illusion of forward motion but primarily to reduce the resistance of air to a speeding vehicle.

By 1927, when Breer and his associates started their experiments with streamlining, a number of milestones had been passed in the development of the automobile. One of them, the electric starter, made large motors with eight or more cylinders practical. With large motors the cars became heavier, and since weight conferred a gentler ride it came to be preferred in spite of its drawbacks in the form of higher operating cost and harder handling for the driver. Achieving the appearance of weight became a goal of design, epitomized by the first Chrysler designed by Breer. It was the 1924 model, featuring a low profile, balloon tires, thick wheel spokes, crowned fenders and rounded body angles. With the introduction of hydraulic brakes the speed potential of the Chrysler high-compression engine could be safely exploited. The form of the car, however, needed drastic revision. It had a boxlike configuration, and its various parts did not really form a harmonious whole. Moreover, the dynamics of its ride resembled the rocking of a hobbyhorse. It was to these problems that Breer addressed himself in 1927. The basic problem was to find the true form of least resistance and adapt it to the automobile.

Streamlining as a principle of design had its roots in hydrodynamics and aerodynamics. In the 19th century the Scottish physicist William J. M. Rankine determined that the motion of fluids has two forms: laminar flow and turbulent flow. Laminar flow can be visualized as a series of parallel layers in a moving fluid, each with its own velocity and direction without disturbance in its forward motion. Turbulent flow is characterized by eddies or vortexes and can be visualized as a tumbling of the fluid caused by a solid body. The turbulence creates a partial vacuum behind the body, which in turn causes drag and impedes the body's forward progress. When a body immersed in a flow does not cause turbulence, it is said to be streamlined.

In 1804 Sir George Cayley had proposed for the dirigible balloon "a form approaching to that of a very long spheroid." Examining natural forms, he took the measurements of the trout and the porpoise. He concluded that the spindle shape of these creatures would not only serve the dirigible well but also, if it were split lengthwise, produce two ideal ship hulls. Francis Wenham refined these early proposals in the wind tunnel. Ludwig Mach tested the laminar and turbulent airflow of different objects by means of silk threads, cigarette smoke and glowing particles of iron (recorded on a photographic plate). Étienne Jules Marey analyzed the aerodynamics of birds in flight.

In his classic treatise of 1917 on growth and form D'Arcy Wentworth Thompson employed the term "streamlining" to describe organic structures that offer the least resistance when they are in motion. To solve the problem of how such forms came into being he mathematically analyzed the shape of a bird's egg by applying the principle of least action, which states that a fluid medium tends to impress its "stream lines" on a deformable body until the body yields and offers a minimum of resistance. Similar examples are the contours of snowdrifts, sand dunes and lamp flames, which illustrate eddy curves that have been imposed by moving air facilitating its own flow. The same principle, Thompson concluded, must have come into play in the evolution of the body form of a fish or a bird.

Breer was one of the first "automobile nuts" and became one of the industry's outstanding engineers. In 1901, as a 17-year-old in his native California, he painstakingly built a steam-powered car. Following his graduation from Stanford University with a degree in engineering, he worked for several automobile manufacturers until 1921, when he organized a firm of consulting engineers with Fred M. Zeder and Owen R. Skelton. Two years later the firm transferred its activities to the Maxwell Motor Car Corporation, where Breer became executive engineer. In 1925, when Walter P. Chrysler bought out Maxwell, Breer was named director of research of the Chrysler Corporation, a position he held until his retirement in 1949.

Breer saw everything in automobiles from the engineering point of view, but he remained alert to the basic role of the car as a servant of people. For example, in order to improve the way cars rode he had to consider the spring suspension, and so he first analyzed the gait at which the human body seemed most comfortable. He found that the least tiring stride was from 80 to 100 steps per minute, equivalent to a speed of from 2.5 to three miles per hour. Applying this finding to the design of the automobile suspension, Breer concluded that the up-and-down motion of a car's body should

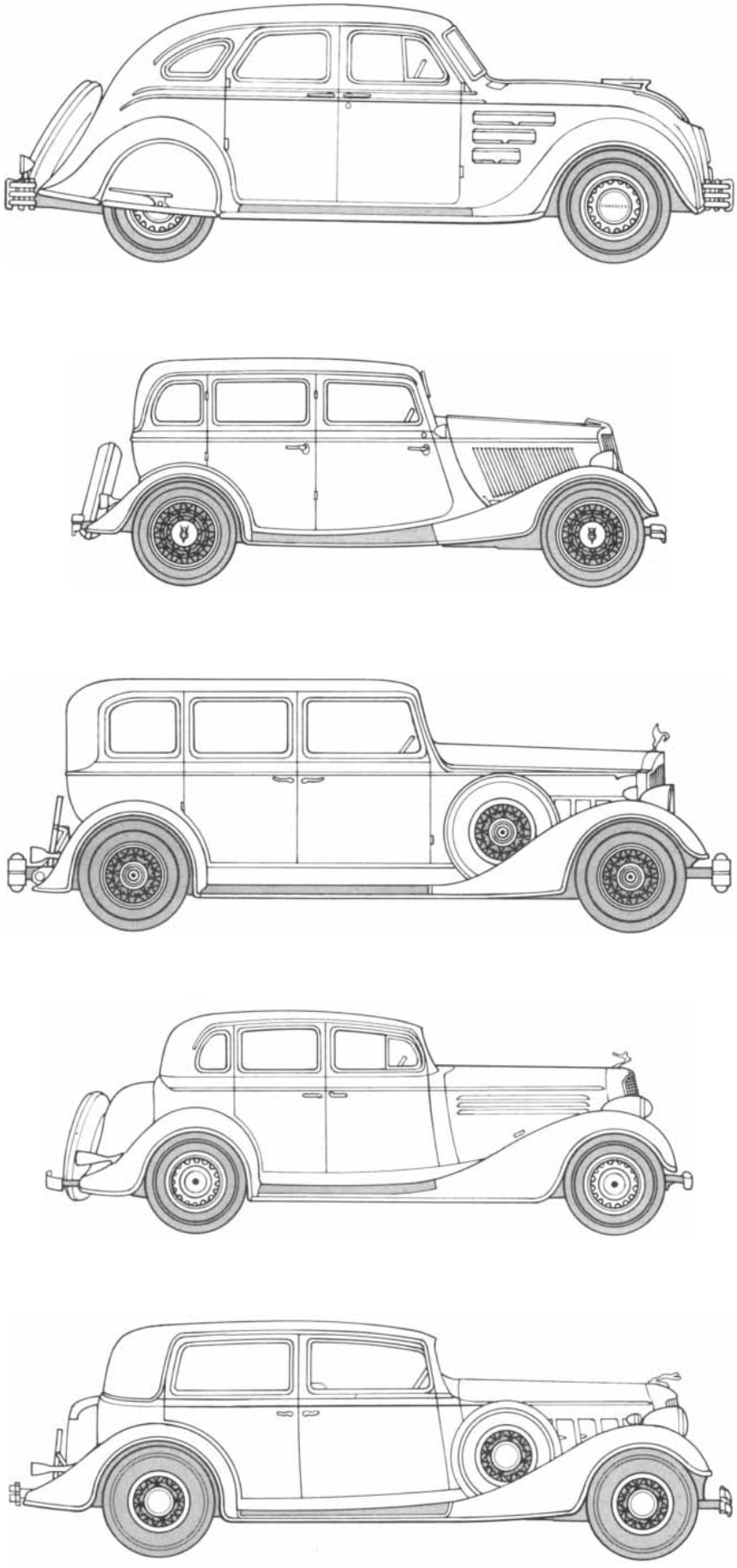
be kept within the same range if maximum comfort was to be achieved for the passengers. The stiffly sprung bodies of the automobiles of the 1920's subjected passengers to much uncomfortable tossing and pitching (at different rates fore and aft). Breer determined that a main goal of engineering had to be the improvement of the ride. To accomplish that, however, it would be necessary to reproportion the weight borne by the front and rear wheels. A new body design was required.

Breer's desire to create a new car was also influenced by new trends in industrial design, which embodied an increasing awareness of the fact that functional objects such as automobiles and refrigerators could also be attractive in appearance. The work of designers such as Norman Bel Geddes, Raymond Loewy, Russel Wright and many others found expression in objects as diverse as cameras and ocean liners. R. Buckminster Fuller carried the trend a step further with his "Dymaxion" concept, in which structural efficiency was emphasized and was brought to realization in the now well-known geodesic dome. Fuller's experimental three-wheeled Dymaxion automobile—light in weight, ovoid in shape, steered by the single rear wheel—markedly resembled the fuselage of an airplane.

It was against this background that Breer began to think seriously about a new body design for the automobile. At that time the ungainly shape of the automobile represented little more than a series of unrelated compromises and had emerged essentially as a big box for people behind a little box for the motor. Breer, believing the speed potential of the high-compression engines developed by Chrysler could not be realized in the bodies then on the drawing boards, asked William Earnshaw, an outside engineer, to determine how seriously the boxy shape of contemporary automobile bodies impeded movement through the air. Earnshaw found that the resistance was high in the forward direction and less if the car went backward. Earnshaw then went to Orville Wright, who advised him to set up a small wind tunnel to test the resistance of differently shaped blocks of wood. Intrigued by the results, Walter Chrysler had a larger tunnel built so that Breer, Zeder and Skelton could carry the investigation further. Breer hoped to find a form that would employ the phenomenon of lift in reverse and thereby press the car more firmly against the road at high speeds.

The wind-tunnel tests continued into

CHRYSLER AIRFLOW of 1934 is owned by the author. The Airflow design was marketed by the Chrysler Corporation under the Chrysler and DeSoto lines in the model years 1934 through 1936 and as a Chrysler only in 1937. The streamlined shape of the car represented a distinct break with previous automobile designs, which had been basically boxlike. The author's Airflow, which is still roadworthy after 43 years, has the second of two grille designs offered in 1934. The modification was made in an effort to meet criticism of the original design, which had lighter, more numerous bars that were more curved. Original grille appears in top illustration on page 102.



1931. By then the Reo Royale of 1931 and the Graham Eight of 1932, each incorporating innovations in design, were being acclaimed. Breer, having found that the teardrop shape generated the least resistance, decided that it would be ideal for a new car. First he would have to rearrange the engine and the passengers within the form. He was disappointed to find that when any of the available engines were mounted at the rear of an experimental model, the car was tail-heavy and also handled poorly. Moreover, the long and vulnerable appendage at the rear would inevitably impede maneuvering. The teardrop would have to be modified.

After mocking up several trial models in wood the three engineers produced in December, 1932, a steel-bodied prototype that they named the Trifon Special. It was a sedan with a short, wide, rounded hood, buried headlights, a steeply sloping windshield and a gently undulating rear deck trailing off to the bumper. Walter Chrysler liked the car, but Breer was not yet satisfied.

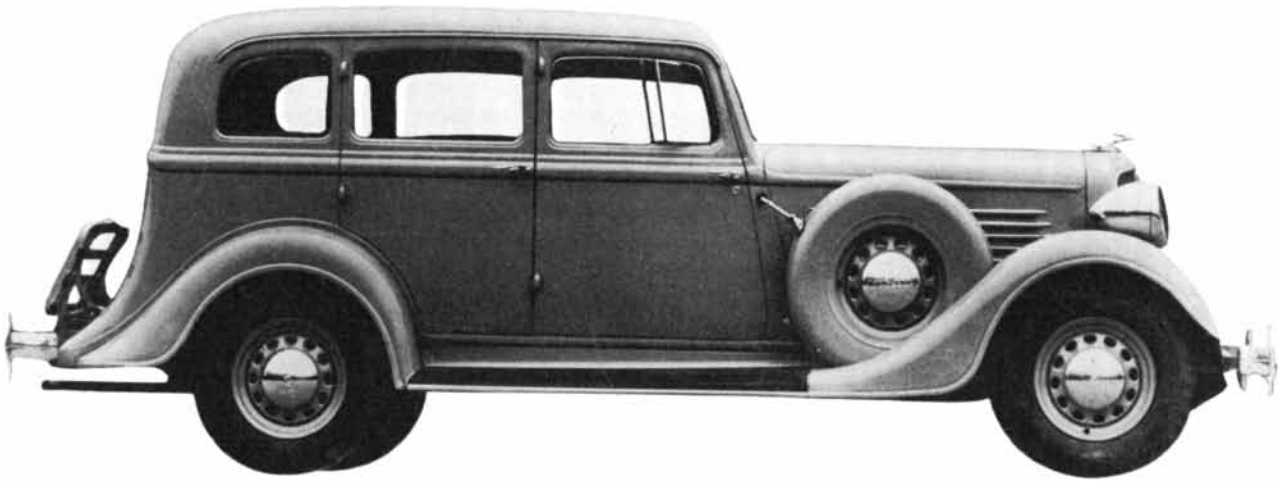
The Trifon's rounded hood was a problem, since it departed sharply from the distinctive long, tapered form, diverging gradually behind a highly stylized radiator shell, that imparted to the automobiles of the day so much of their visual character. Breer and his colleagues mocked up somewhat more conventional hood designs as late as May, 1933, but they finally decided that if the new design was to be faithful to the aerodynamic principles guiding their work, the rounded nose would have to be part of it.

After some 50 experimental designs had been tested tools and production were set to manufacture the Airflow for 1934. It was produced in five body sizes, for which an interesting array of interchangeable components were developed. The bodies were put on wheelbases ranging in length from 115.5 to 146.5 inches.

The smallest version, powered by a six-cylinder engine, was marketed under the DeSoto name. The other four were all Chryslers and had eight-cylinder engines. The largest model was an immense limousine that had a curved one-piece windshield. Both Chrysler and DeSoto offered the Airflow as a sedan with two or four doors and as a two-door coupé. The coupé was undoubtedly the purest expression of the Airflow design.

The Airflow was by far the most streamlined car to have been put on the

FIVE 1934 AUTOMOBILES are portrayed in side views to show the contrast between the Airflow and the standard designs of the time. From the top the cars are the Chrysler Airflow, the Ford, the Packard, the Buick and the Lincoln. All five cars are at the same scale.



CONVENTIONAL 1934 CHRYSLER was a six-cylinder model the Chrysler Corporation offered while it was also marketing the Airflow. Public response to the Airflow was so unfavorable that the company would not have been able to maintain its sales position if the conventional model had not sold well. The four Chrysler Airflow models of 1934 all had eight cylinders; the DeSoto Airflow had six.

market up to that time. It looked so utterly unconventional that a reviewer of the 1934 automobile show in New York reported that it took two or three days to become accustomed to it. Rarely had a major manufacturer brought forth a design so little bound by past traditions. The financial risk was tremendous, particularly since Chrysler's sales had slumped in the preceding two years.

The proportions of the Airflow differed from those of all the other automobiles then made. Passengers were moved 20 inches forward with respect to the rear axle, with the result that the engine was pushed out over the front axle and the radiator went even farther forward. The passengers were therefore seated nearer the middle of the car and so gained considerably in riding comfort. On the other hand, the same changes gave the car its nose-heavy appearance. Indeed, 55 percent of the Airflow's weight was borne by the front tires; most of the other cars had about 40 percent of their weight there. The shift in weight facilitated the improved riding characteristics that Breer sought by enabling him to put more nearly similar springs on the front and back. The shorter front springs had more leaves, and their flexing was limited to the 80 to 100 cycles per minute that Breer had determined to be the most comfortable for human beings. The Airflow's ride was in fact a great improvement, contrasting sharply with the rocking and pitching that was common in all but the heaviest cars of the day.

The passengers were not only shifted forward but also given more room laterally. The body and windshield of the Airflow were 10 inches wider than in the preceding Chrysler models. This result was achieved by spreading the body out over the fender wells, making it possible for the first time to carry a total of five or six passengers on the two seats.

One outstanding structural innovation of the Airflow was in the construction of its body shell. Instead of consisting of a series of separate steel panels attached to a light wood frame, which had been the general practice, the entire body except for the fenders, the doors and the hood was a single stressed-steel member that, in combination with the very light frame, had 40 times more torsional rigidity than the body of the preceding Chrysler model.

Overall the body of the Airflow resembled a section through the wing of an airplane: it was a parabolic curve rising from the front bumper and gently trailing off to the rear. The form was stepped to accommodate the sloping, divided windshield, which could be opened outward from the bottom. The headlights and front fenders were absorbed into the single rising curve of the front façade. A slight upturn and thinning of the trailing edges of the fenders reinforced the teardrop motif.

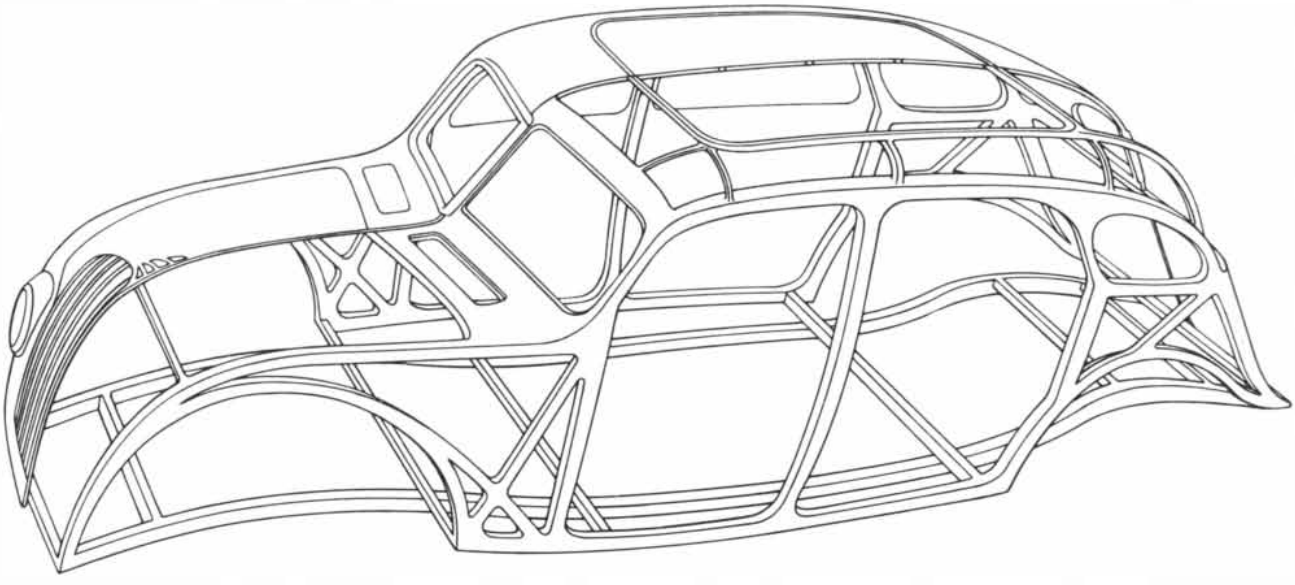
"Art deco" themes appeared in much of the ornamentation, such as the three-tiered bumpers, the three horizontal hinged vents on each side of the hood and the parallel belt moldings along the body. Several unfamiliar materials were used inside the car: molded, washable roof panels; tubular-steel, chromium-plated seat frames and marbled rubber floor mats.

Did all of this result in a beautiful car? The sloping rear deck was widely admired, particularly in the coupé, which carried the spare tire inside so that nothing interrupted the clean sweep down the back. The raked *V* windshield and the full skirts of the rear fenders had met with approval on earlier cars. The broad, rounded front end, however, remained an unsolved problem. The stubby hood, with the headlights, radiator grille and fenders all blended into a massive cascade of metal, contrasted totally with the delicate, aquiline, prowlike grille, the free-standing headlights and the graceful outrigger front fenders of cars such as the widely acclaimed LaSalle. The Airflow evoked strong reactions; people either admired it intensely or detested it.

In the end the opinion of the detractors prevailed. Although the company had a good backlog of orders early in 1934, the summer sales figures told the disappointing truth: DeSoto's sales had slipped 47 percent below those of the previous year, and Chrysler's sales position (10th) was maintained only because the company simultaneously offered an array of conventional six-cylinder models. The round blandness of the Airflow's front end was described by critics variously as being bug-eyed, having a look of rhinocerine ungainliness and resembling a human face covered with a stocking. The car soon became the object of ridicule.

Nothing could be done to alter the basic shape of the Airflow. Walter Chrysler was not, however, ready to abandon it. Stylists were rushed in, first to strengthen the face of the 1934 Chrysler Airflow by replacing the curious compound-curve waterfall grille with fewer, straighter, stronger bars and then for the three succeeding model years to graft on a more conventional prowlike radiator grille and to simplify other front-end ornamentation. The rakish tilt of the steering wheel was reduced, the headlights were toned down, the anachronistic freestanding taillights were drawn into the fenders and broad bumpers replaced the easily damaged three-tiered ones. A bustle trunk mounted on the sedans brought the spare tire inside. Wool broadcloth and cut pile replaced

MODIFICATION OF GRILLE was the chief means whereby the Chrysler Corporation tried to make its Airflow car more acceptable to the public. The heavy-looking rounded front of the car was a focus of criticism from the outset. Here one sees at the top left the original grille design of 1934, which was modified during the year, and, reading from left to right, the grille designs of 1935, 1936 and 1937.



BODY FRAME OF AIRFLOW incorporated several innovations. It was light and made of steel, whereas in most cars up to that time it had been made of wood. The body attached to the Airflow frame was, except for the fenders, doors and hood, a single piece of stressed steel, whereas other cars had separate steel panels attached to the wood frame. The Airflow body had uncommonly high torsional rigidity.

the rubber mats, molded roof panels and chromium-plated tubing inside.

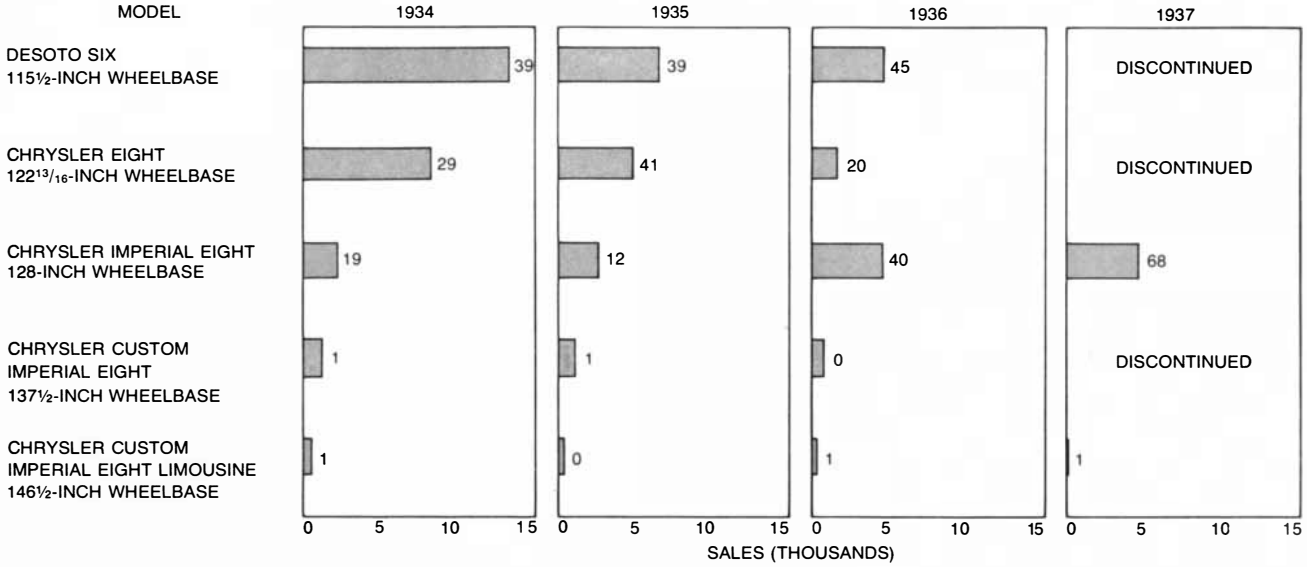
Notwithstanding these and other largely cosmetic changes, the sales of the Airflow diminished year by year. After years of work and millions of dollars spent it was a bitter disappointment in the salesroom, much to the dismay of Breer, Zeder and Skelton and particularly Walter Chrysler. Although the company lost money only in 1934, it was embarrassed and frustrated by its magnificent failure.

The fate of the Airflow dampened the enthusiasm of the advocates of streamlining. The Reo and the Graham a few years earlier had been striking at first look, but neither of them had flouted earlier standards of beauty. The Airflow, however, not only jumped ahead of its predecessors but also incorporated such an array of novel features that the buying public was turned away rather than attracted. In effect the Airflow posed a question: Is honest functional design in an automobile beautiful, even when it breaks completely with tradition? The answer in the summer of 1934 was emphatically no.

All the same the impact of the Airflow on design was soon evident, although the evolution of the design was more successfully accomplished by other manufacturers. Certainly the Lincoln Zephyr, introduced in 1936, was Ford's answer to the Airflow. In basic body engineering the two designs had much in common: a short hood, a sweeping tail, passenger seating in the middle. The Zephyr's shell was even wider than the Airflow's, and its unitized body trusswork was even further developed. It also had a graceful appearance. Its hood was flat, behind a *V*-shaped, prowlike grille, and it was flanked by distinct fenders, each one carrying a partly buried headlight, all relieving the heaviness disliked in the Airflow. In spite of its unexciting engine, drive train and suspension the Zephyr, unlike the Airflow, was widely praised in the automotive press. Even so, its sales too were slow at first.

Ironically, just seven years after the debacle of the first Airflows the General Motors Corporation introduced its "fastback" design, with a profile almost identical with that of the Airflow coupé except for a vertical grille and a long, flat hood. The design was marketed through all the company's automobile divisions until well into the 1950's and was strong in sales.

Equally ironic is the great success of the Volkswagen "Beetle." Its designer, Ferdinand Porsche, was so impressed by the engineering logic reflected in the Airflow that he immediately began work on a scaled-down version that would incorporate the rear-engine feature Breer had abandoned. The car came on the market in 1936 and before



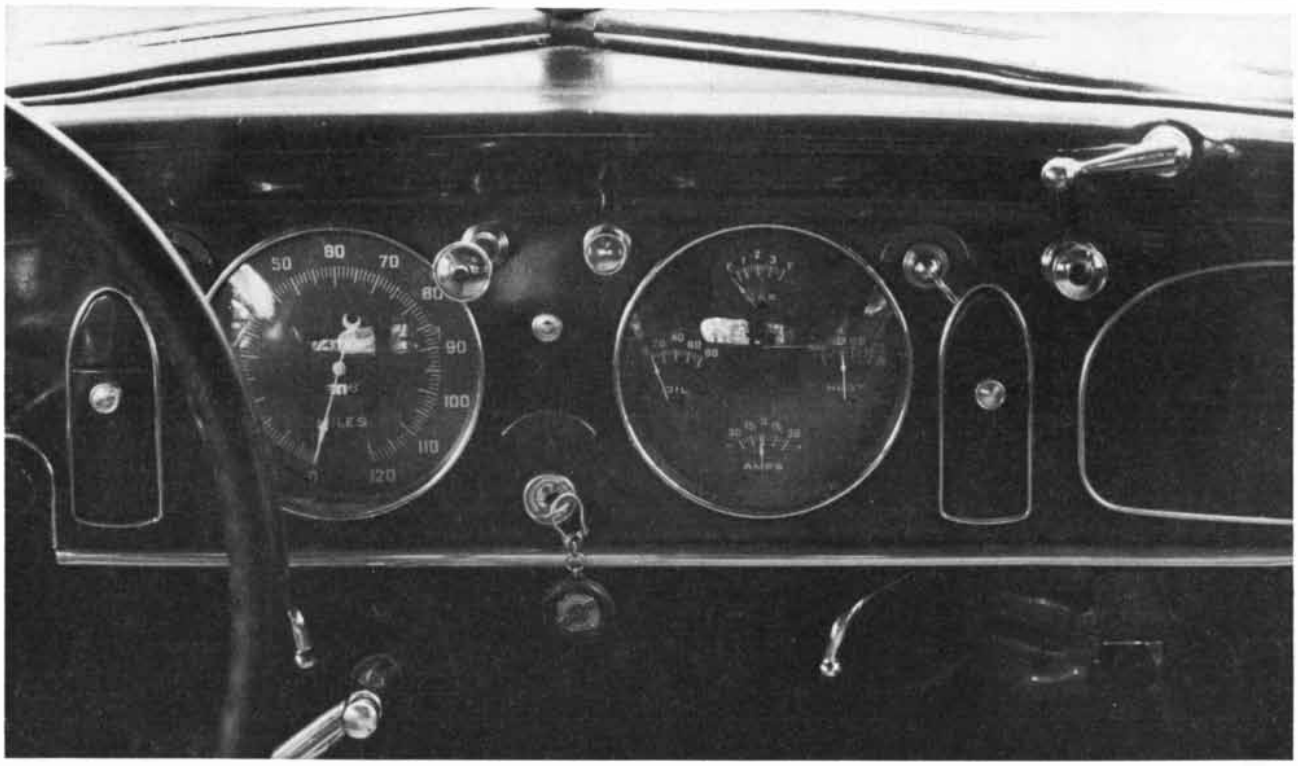
SALES AND SURVIVAL of the Chrysler and DeSoto Airflows are charted. For each model year the bars show the number of cars manufactured, and the numeral at the end of each bar shows how many are still in running condition according to the Airflow Club of America.

long was widely bought in many countries.

Thus one cannot call the Airflow a failure even though it sold poorly. In it Breer introduced a number of important technical innovations, gave high priority to the comfort of the passengers and conditioned the public to accept the streamlining principles that were to be evident in all automobiles by the end of the decade.

I recently bought a 1934 Chrysler Airflow sedan, the third Airflow I have owned. My daughter's first remark on seeing it was, "It looks like a big Beetle!" Friends and colleagues regard it as an antique curiosity and a shrewd investment, but only the most perceptive see in it the milestone of automotive history it represents.

After 43 years and undetermined but undoubtedly high mileage the body remains free of rattles, bearing silent testimony to the soundness of one of Breer's engineering tenets. Although the profile of the body was low for the 1930's, it is



SYMMETRICAL DASHBOARD of the Chrysler Airflow was photographed in the author's 1934 model. The instrument panel at the left incorporates the speedometer and two odometers (the lower one reading up to 99.9 miles for the measurement of trips). In the panel at the right are, clockwise from the top, gauges for fuel, engine temperature, amperage and oil pressure. The handle at the upper right serves to open and close the right half of the windshield from the bottom; similar handle for the left half is barely visible in the picture.

SCIENCE/SCOPE

Survival odds have increased for US Army tank crews, thanks to a remarkable new fire-suppression system to be incorporated in the US Army's new XM1 main battle tank . . . scheduled for use during the 1980's. The system will detect and suppress a fuel explosion inside the tank, extinguishing it within 1/10 of a second which is less than half the time it takes to blink an eye.

Similar devices in the past have triggered false alarms due to signals as common as the energy reflected from colored shirts. So combat crews simply turned them off. In contrast, the Hughes dual-spectrum sensor can detect a mini-explosion even in direct sunlight. Yet, it will not false-alarm even when pointed directly at the sun or other light sources such as gunfire, rockets, lightning, matches or other vehicles. Infrared sensors and related electronics for the fire-protection system will be built by Santa Barbara Research Center, a Hughes subsidiary, for the XM1's prime contractor, Chrysler Corporation.

Improved forecasting of major crop yields, a step in the battle against world famine, is among the benefits predicted from an advanced space instrument called Thematic Mapper. Scheduled for launch in early 1981, it will be installed on Landsat-D, fourth in the NASA Goddard Space Flight Center's series of Earth Resources Technology Satellites. Hughes, with its Santa Barbara Research Center subsidiary, will design and develop the instrument's basic structure, telescope, calibrating system, detector arrays and processing electronics.

The new sensing instrument, with a ground resolution expectation 2½ times greater than present sensors, should also contribute to improvements in agricultural land use, forest and water resource management, land use mapping and mineral exploration. Launched in a low-altitude north-south orbit that carries them over both poles, Landsat satellites provide full coverage of the earth's surface.

Hughes Aircraft Company has openings for experienced engineers and scientists in its Culver City, California laboratories. Radar Systems Group needs are for electronic circuit engineers to design microwave transmitters, power supplies, and RF, IF and video frequency analog circuits; mechanical engineers to design airborne microwave antenna structures; engineers to develop hybrid microcircuitry; systems analysts to apply modern control theory to tracking systems, and signal theory for radar system detection and acquisition functions; signature technology engineers to analyze and simulate target return digital processing for target identification.

Use of the laser as a high-speed cutting tool has been envisioned since the first laser was demonstrated at Hughes more than 15 years ago. Since 1971, when the first lasercutter was used to cut cloth in apparel manufacture, its advantages in speed, accuracy, material saving, response to design change, and adaptability to computer control have been applied to other industries. Systems have been used in the US and abroad for such applications as cutting patterns for footwear and cutting sheets of boron epoxy broadgoods used in aircraft part manufacture.

Now, the first, large-scale, metal-lasercutting system is being built for Garrett AiResearch to cut corrugated steel. The system eliminates the need for costly dies, provides for more efficient use of materials, and permits quick, economical response to special engineering changes.

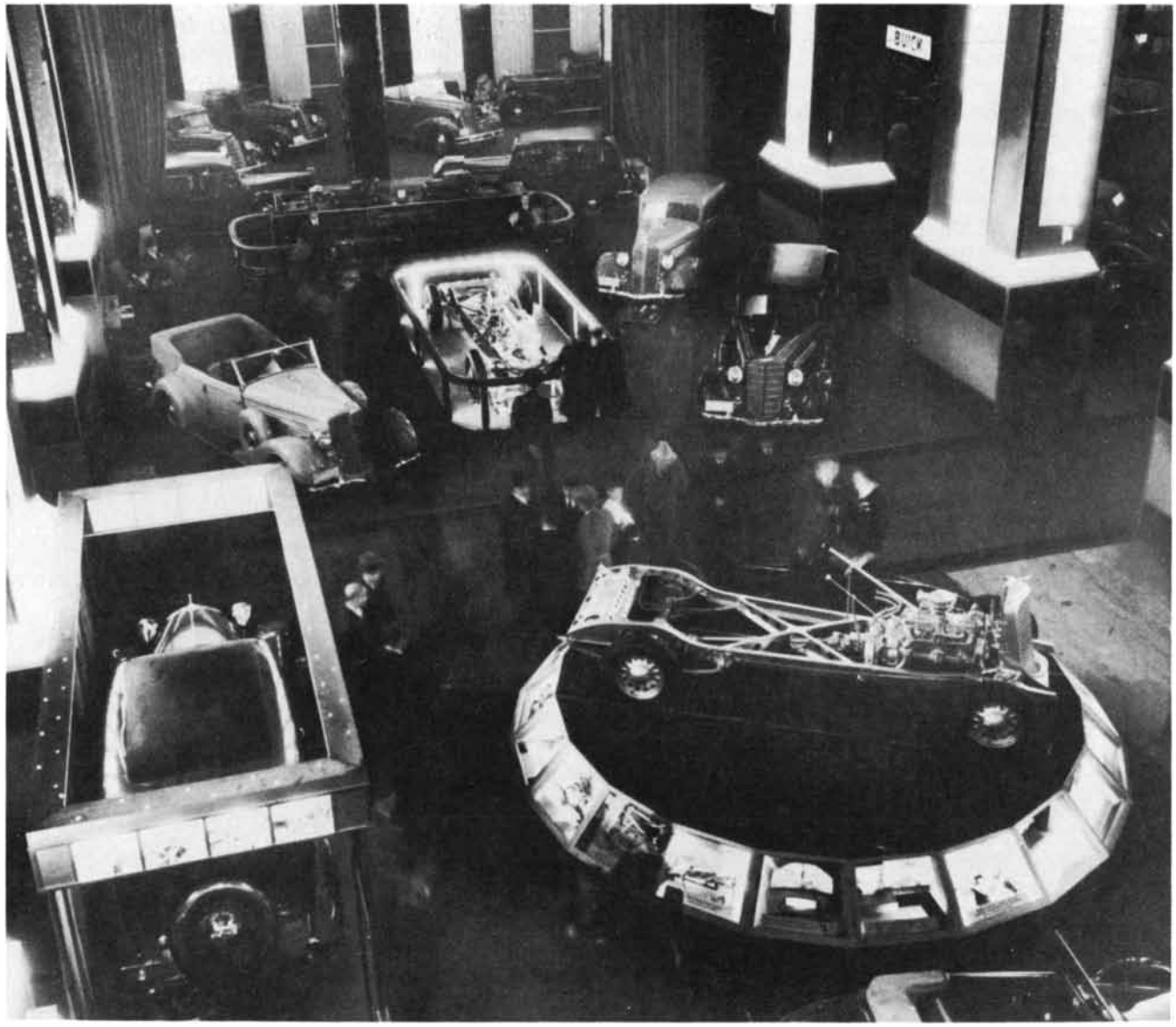
68.5 inches high, so that one has to step up by means of the running board to get into the car. Seating is high by today's standards, but visibility is limited, partly by the modest amount of glass in the car (particularly in the rear) and partly by the thick pillars associated with the doors and the windshield.

The 112-horsepower, straight-eight, flat-head, side-valve engine is very quiet, as is the transmission. Acceleration is modest by today's standards, but fuel consumption (from 18 to 22 miles per gallon with an overdrive at high speeds) is respectable for a car that weighs 3,900 pounds. The unassisted drum brakes are balanced and positive in action but require heavy foot pressure. Some larger Airflow models had vacuum boosters to relieve this problem.

The ride is best described as heavy but nimble. Steering is manageable as long as the car is moving, but without any hydraulic assistance it is heavy for parking. Cornering is remarkably well controlled for such a softly sprung car, and excessive leaning is avoided by a rear-suspension stabilizer bar (moved to the front suspension in later models).

Because of the six-volt electrical system the cranking done by the starter is slower than today's cars provide with 12-volt systems, and the feeble bulb-and-reflector lights become exceedingly dim when at low engine speeds they are not assisted by the generator. The lighting was never satisfactory in the Airflows, even when they were new. Separate gauges for amperage, oil pressure, water temperature and fuel are welcome, but their location in a circular cluster on the right side of the symmetrical dashboard is distracting. The grille and the hood top open as a rear-hinged unit but offer only limited access to the long engine, which is mounted high over the front axle. Lateral access to the engine, which is required to adjust the valve tappets, requires removal of the right front wheel and the detachable inner body panel.

Today the Airflow has largely passed from the public consciousness. Several hundred of the cars survive (perhaps 1,000 or more if badly deteriorated ones are counted), many of them restored, pampered and registered on the rolls of the Airflow Club of America, a doughty band of devotees whose activities did much to create the current interest in the preservation and restoration of old automobiles. What is important about the Airflow is that the clamor it caused in 1934 pushed the automobile manufacturers into making a start toward adopting new principles of engineering and design. Once that process had begun there was no turning back.



NEW YORK AUTOMOBILE SHOW of 1934 provided the occasion for the public introduction of the Airflow line. This photograph, made from a balcony overlooking the displays of the Ford Motor Company and the General Motors Corporation, indicates how sharply the Airflows differed from the other cars offered in 1934. The two cars in the foreground are Lincolns; behind them is a group of Buicks.

INGENIUM PRO BONO HUMANITATIS

The Marconi International Fellowship awarded annually to honor noble achievement in the service of mankind

The Marconi International Fellowship commemorates Guglielmo Marconi's creative contributions to scientific discovery, engineering, and technology. Established in 1974 on the centenary of the inventor's birth, the Fellowship recognizes and inspires achievements of living men and women whose efforts in the fields of communications sciences and technologies, like Marconi's, are characterized by a profound commitment to human betterment.

The Fellowship, a $25,000 grant, enables the recipient to undertake or complete a project or study which has as its ultimate objective the well-being of mankind.

The first three Fellowships have been awarded in the fields of communications and electronics to Dr. James R. Killian, Jr., Honorary Chairman of the Massachusetts Institute of Technology, in association with Lord Briggs, Worcester College, Oxford University; to Professor Hiroshi Inose, University of Tokyo; and to Arthur L. Schawlow, Stanford University.

Learned societies, academies and universities from all countries as well as individuals in industry and public life are invited to propose candidates for the Fellowship. All nominations must be received by October 15, 1977. Please address all inquiries to:

The Marconi International Fellowship Council
Aspen Institute for Humanistic Studies
1919 Fourteenth Street, No. 811
Boulder, Colorado 80302 USA

"Second Messengers" in the Brain

Nerve cells communicate by secreting neurotransmitters. These chemical messages are translated by second messengers within the cell into transient and longer-lasting physiological actions

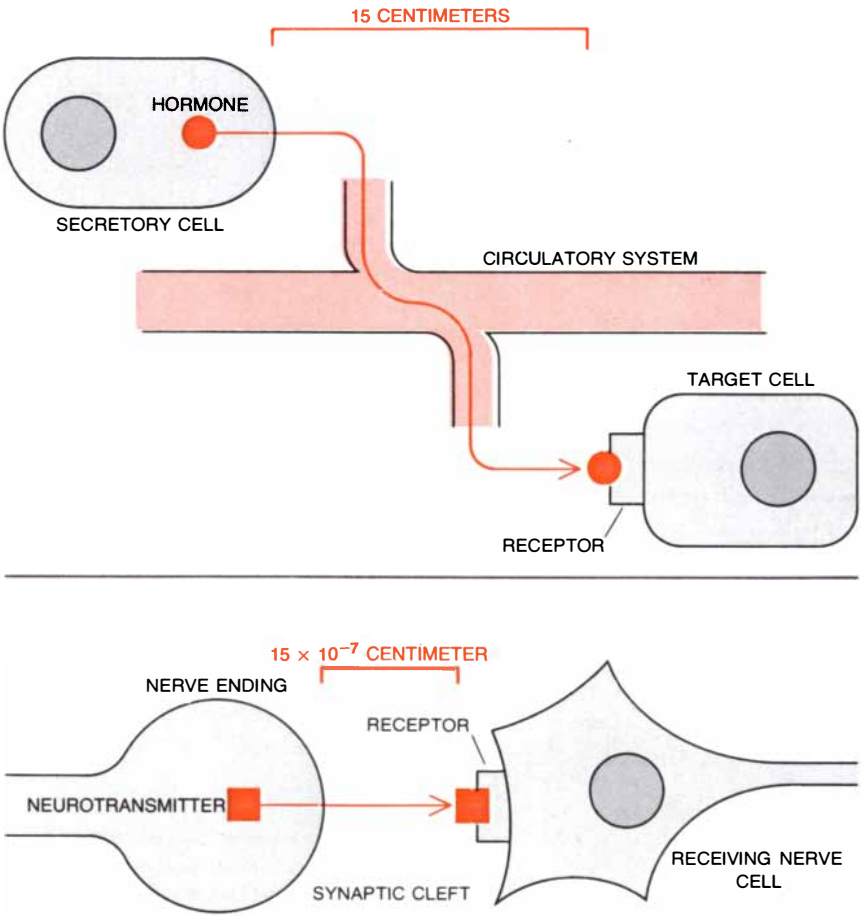
by James A. Nathanson and Paul Greengard

For a multicellular organism to survive and function effectively its component cells must act in a coordinated fashion. Such coordination necessitates the transfer of information between cells in widely separated parts of the organism, and in most higher animals there are two major pathways of intercellular communication: the endocrine system and the nervous system. In the endocrine system specialized cells secrete hormones, which are carried in the bloodstream to distant parts of the body and influence the activity of specifically responsive target cells. In the nervous system a network of nerve cells with elongated processes communicate with one another by secreting neurotransmitters, which traverse the tiny gap between two nerve cells and produce a change in the electrical activity of the receiving cell.

Both systems involve messenger molecules that are released from one cell, travel a certain distance and interact with the surface of a second cell to modify its activity. In view of this similarity and assuming that nature functions economically one can postulate that some types of chemical transmission between nerve cells might be mediated by mechanisms similar to those mediating the physiological effects of certain hormones. This point of view has led to new perceptions of the biochemical organization of the brain and the mechanism of action of many drugs that affect behavior.

To provide background for a discussion of the molecular mechanisms of nerve transmission, let us first review some current concepts of the mechanism of hormone action. Hormones regulate an enormous range of biochemical processes in their target cells by influencing the rate at which enzymes and other proteins are manufactured, by affecting the activity of enzymes in key metabolic pathways or by altering the permeability of cell membranes. Because these processes operate in the interior of the target cell, the hormone itself or the information conveyed by it must somehow be made available inside the cell.

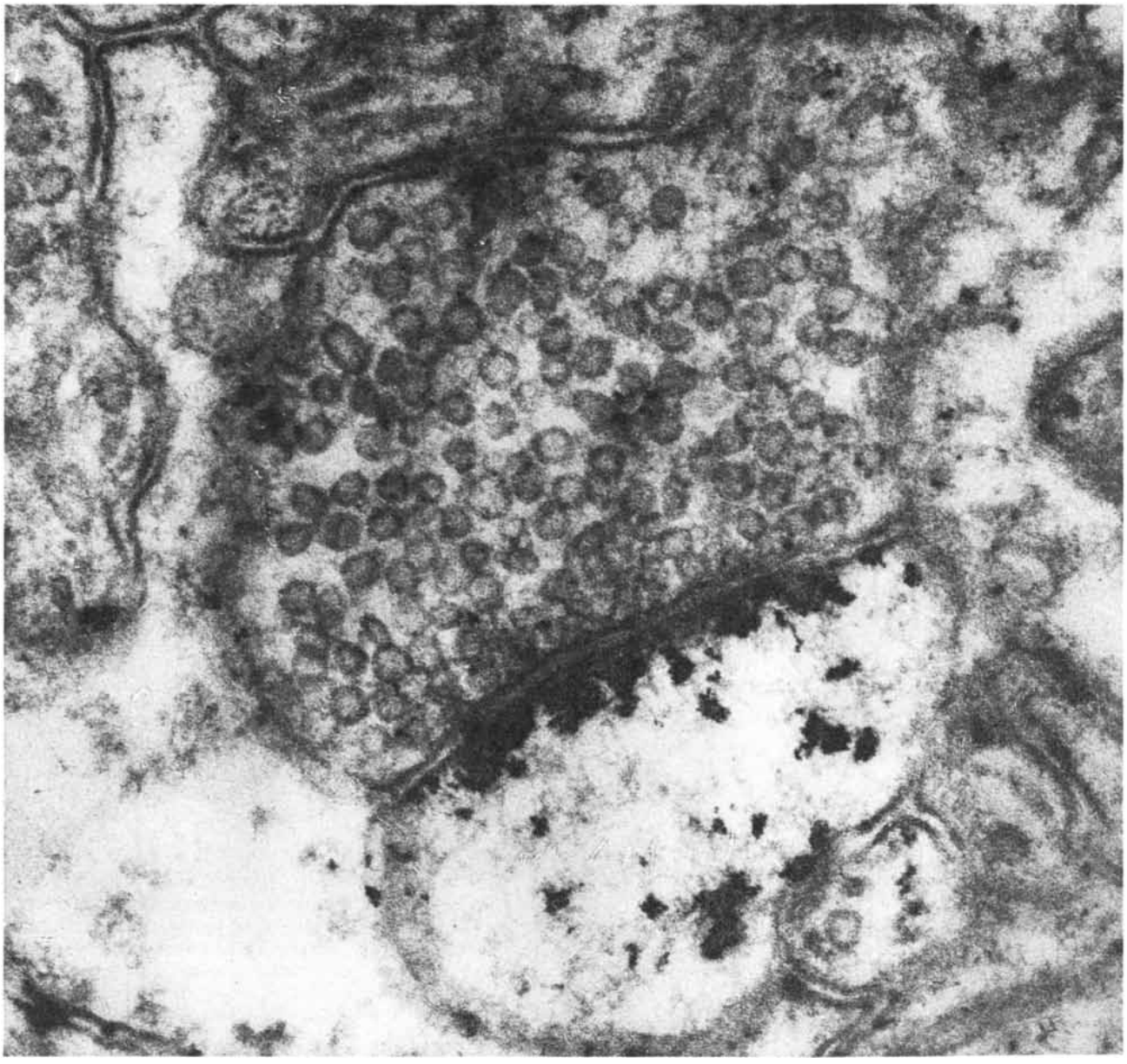
The delivery of a hormonal message to the interior of the cell appears to be achieved in one of two ways. Steroid hormones, which are derived from cholesterol, are easily soluble in fat and hence can pass through the fatty outer membrane of the target cell and directly influence processes inside it. Hormones of this type include cortisone and the sex hormones estradiol and testosterone. Peptide and amino-acid-derived hormones such as insulin and adrenalin (epinephrine), however, are not able to penetrate the cell membrane because of their size or molecular structure. Instead they attach themselves to specialized receptor sites on the surface of the cell and influence the biochemical machinery from the outside.



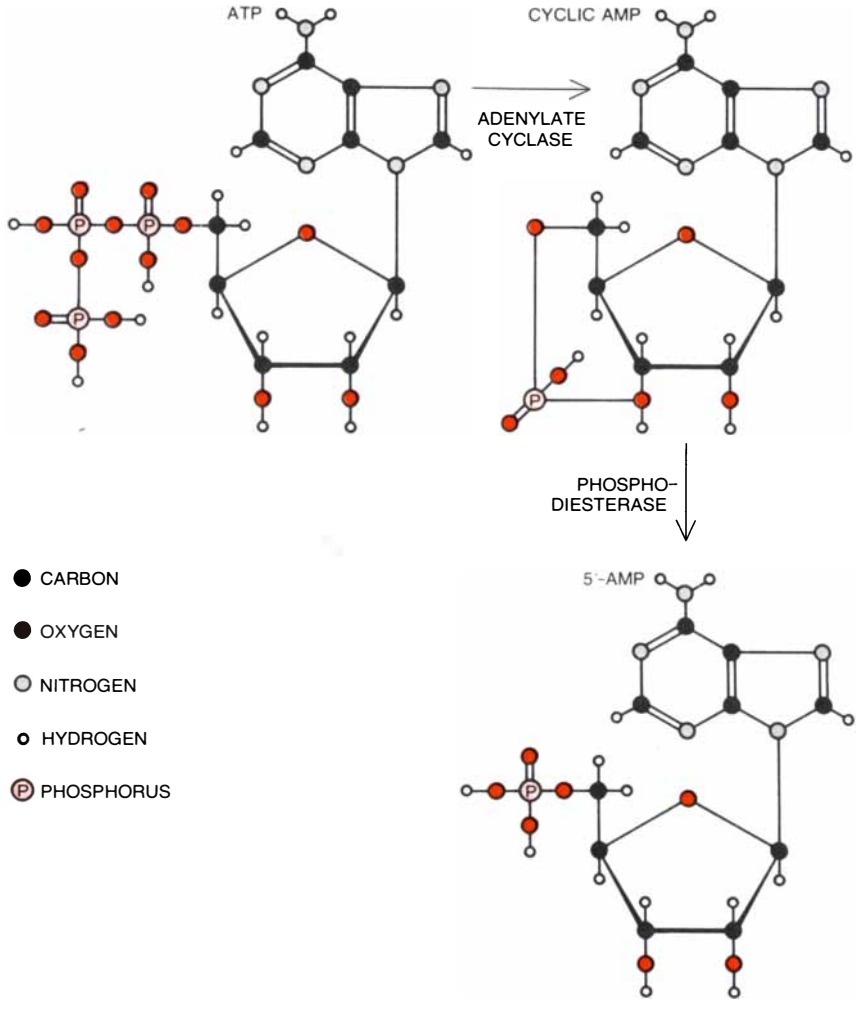
ANALOGOUS MECHANISMS underlie communication in the endocrine system (*top*) and the nervous system (*bottom*). In both chemical messengers (respectively hormones and neurotransmitters) are released from one cell, travel through the extracellular medium and bind to receptors on the surface of a second cell to modify its activity. The intracellular effects of many hormones and neurotransmitters appear to be mediated by the second-messenger cyclic AMP.

The hormone receptors—large protein molecules embedded in the cell membrane—are quite selective in their ability to bind the hormone for which they were designed, presumably because the molecular configuration of the receptor allows the hormone molecule to fit the receptor quite precisely. The forces that hold the hormone to the receptor are not the covalent ones that bind the atoms in a molecule; they are weaker forces that soon release the hormone, leaving the receptor free to accept additional molecules of the same hormone. Thus the extent to which a hormone influences its target cell depends on the concentration of the hormone in the fluid outside the cell and the affinity of the hormone for the membrane receptor.

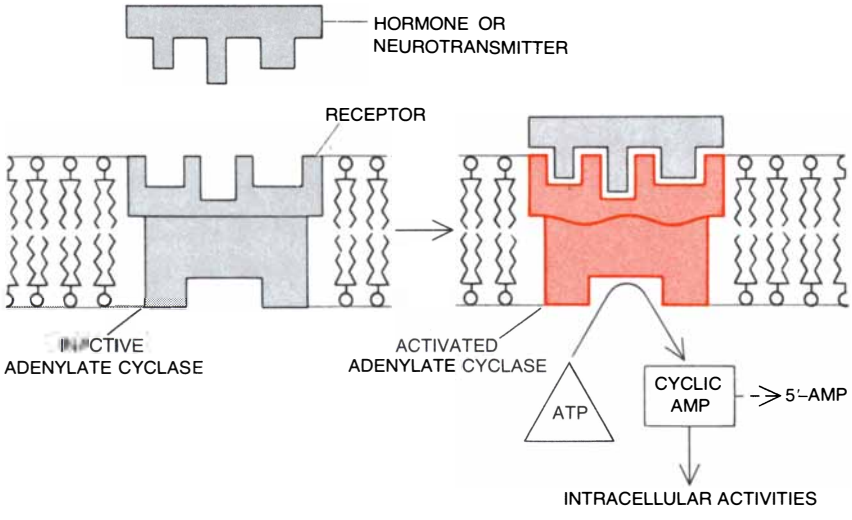
Once a hormone has bound to a receptor, how does it convey its message to the interior of the cell? Earl W. Sutherland and his colleagues at Case Western Reserve University first addressed this question some 20 years ago. At that time they were studying the mechanism by which the hormone adrenalin causes liver cells to release the sugar glucose into the bloodstream when the body is under stress. The released glucose comes from the breakdown of glycogen (animal starch), which is held in reserve by the liver. Sutherland and his colleague Theodore W. Rall found that when they exposed to adrenalin cell membranes isolated from liver cells, an unidentified factor was produced that, when it was mixed with the cytoplasm of the liver cells, mimicked the action of adrenalin in causing the conversion of glycogen to glucose. Hence it seemed that the response to the hormone had at least two stages: the interaction of the



SECOND-MESSENGER FUNCTION of cyclic AMP in nerve-cell communication is suggested by the localization of phosphodiesterase, the enzyme that degrades cyclic AMP, at the synaptic junction between two nerve cells where chemical transmission occurs. Noel T. Florendo and Russell J. Barrnett, in collaboration with the authors at the Yale University School of Medicine, localized the enzyme by applying chemicals to brain tissue that reacted with the enzyme's product to form a dense precipitate. They then examined the tissue in the electron microscope. In the micrograph of a single synaptic junction shown here, the precipitate marking the location of phosphodiesterase is associated with the region of the receiving cell membrane possessing neurotransmitter receptors. This finding supports the hypothesis that the neurotransmitter molecules released from the circular vesicles in the nerve terminal (*top*) bind to the receptors and induce the manufacture of cyclic AMP, which then mediates a variety of actions inside the cell. Magnification is 138,000 diameters.



SYNTHESIS AND DEGRADATION of cyclic AMP are accomplished by two enzymes associated with the outer membrane of certain hormone-sensitive or neurotransmitter-sensitive cells. Adenylate cyclase converts the energy-carrier adenosine triphosphate (ATP) into cyclic AMP by removing two phosphate groups from the molecule and joining the remaining phosphate with the molecule's carbon backbone to form a ring. Phosphodiesterase inactivates cyclic AMP by opening the phosphate ring, converting the molecule into an inert form of AMP.



ADENYLATE CYCLASE IS ACTIVATED by the binding of a hormone or a neurotransmitter to its receptor on the cell membrane. The enzyme then converts some of the ATP in the cytoplasm into cyclic AMP, which relays the signal from the membrane to interior of the cell.

hormone with the membrane to form the unidentified factor, followed by the factor's activation of the biochemical mechanism in the cytoplasm.

Subsequent experiments identified the factor as cyclic adenosine monophosphate (cyclic AMP), one of the class of small molecules called nucleotides. It is related in structure to adenosine triphosphate (ATP), the universal currency of chemical energy in the cell. The "cyclic" in cyclic AMP refers to the fact that the single phosphate group (PO_4) in the molecule forms a ring with the carbon atoms to which it is attached.

Sutherland and his co-workers soon found that the membranes of liver cells (and many other cells) contain an enzyme, adenylate cyclase, that converts ATP into cyclic AMP. Because ATP is located almost exclusively in the cytoplasm, Sutherland and his colleagues G. Alan Robison and Reginald W. Butcher reasoned that at least part of the adenylate cyclase molecule must face inward and that the cyclic AMP manufactured by the enzyme is released into the interior of the cell. When they exposed liver cells to adrenalin, the rate at which adenylate cyclase converted ATP into cyclic AMP was substantially increased, indicating that there was a functional link between the binding of hormone to the receptor on the outside of the membrane and the activation of adenylate cyclase on the inside. Such a connection was subsequently shown to exist for a wide variety of hormones that bind to membrane receptors, although since different kinds of cells possess receptors for different hormones, a given hormone will increase the level of cyclic AMP in its target cells but not in other cells.

It is now generally accepted that the cyclic AMP generated by adenylate cyclase in response to the binding of a hormone to the receptor acts as a "second messenger" to relay the message of the hormone (the first messenger) from the membrane to the cell's biochemical machinery. In this way the low-level signal represented by the hormone can be amplified many thousands of times by the manufacture of cyclic AMP.

Up to this point we have discussed the interactions among hormones, receptors and cyclic AMP in cells outside the nervous system. Can analogous mechanisms help to explain how nerve cells communicate? The nerve cell, or neuron, is the basic structural and functional unit of the central nervous system; behavior is the net result of the complex interaction of many nerve cells. Information is conveyed along the elongated fiber of a neuron in the form of an electrochemical impulse. The impulse stops, however, when it reaches the tiny synapse, or gap, that separates the fiber terminal from its receiving neuron. To bridge the synapse and reinitiate

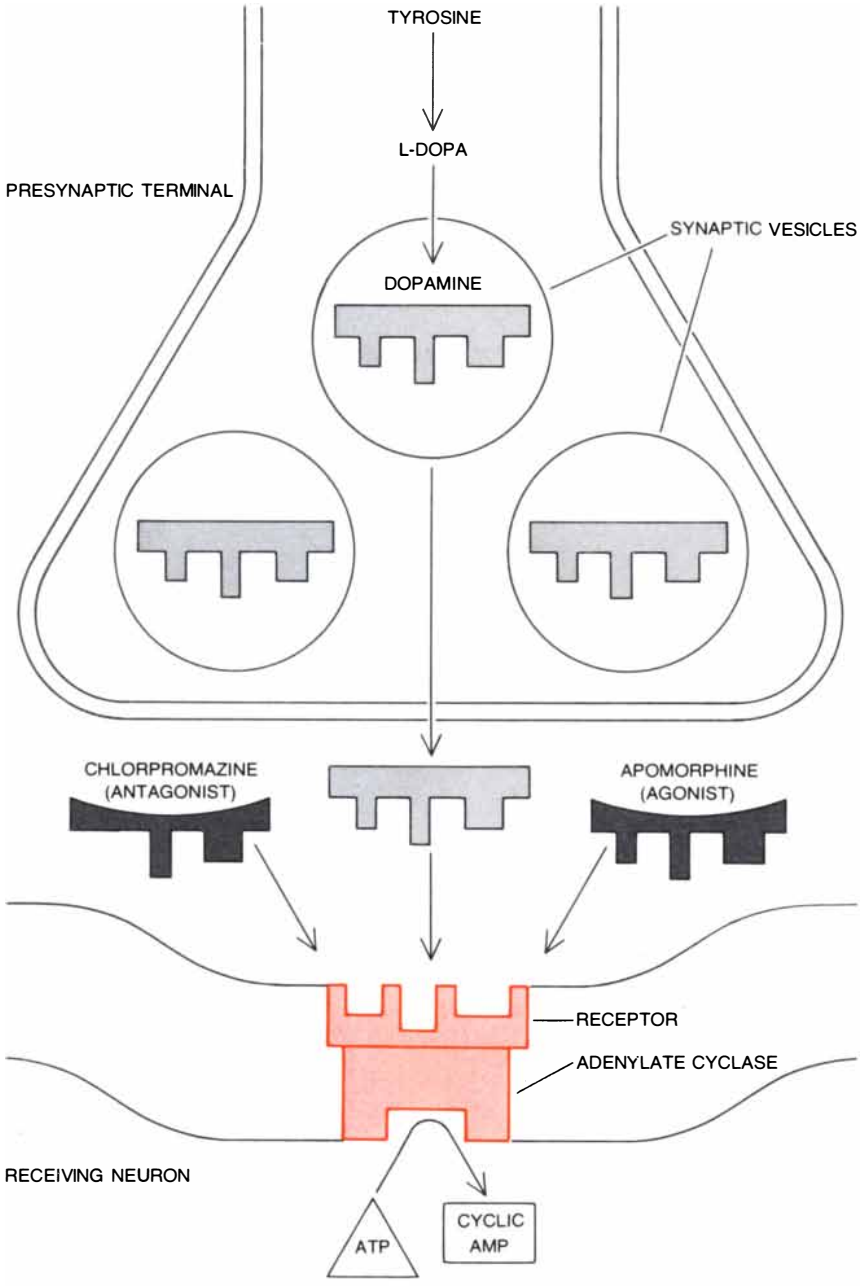
electrochemical transmission, neurotransmitter is released and travels across the gap between one cell and the next. Like the hormones we have been discussing, neurotransmitters do not enter the receiving cell but interact with receptors on the outside of the cell membrane, leading to a change in the electric potential across the membrane.

The first connection between cyclic AMP and brain function was made when Sutherland and his co-workers found a large amount of the enzyme adenylate cyclase in the brain of vertebrate animals, indicating that cyclic AMP was being actively manufactured there. In 1967 Eduardo De Robertis, working in collaboration with Butcher and Sutherland, found that when brain tissue was disrupted by homogenization and the resulting subcellular components were separated according to their density by spinning them in a centrifuge, those fractions containing the largest amount of pinched-off nerve endings also showed the highest level of adenylate cyclase activity. This finding was interesting because many of the nerve-ending particles include fragments of membrane from both sides of the synaptic junction and hence represent the precise areas of the brain where nerve cells communicate. The fact that adenylate cyclase was specifically associated with those areas suggested that cyclic AMP might somehow be involved in synaptic transmission.

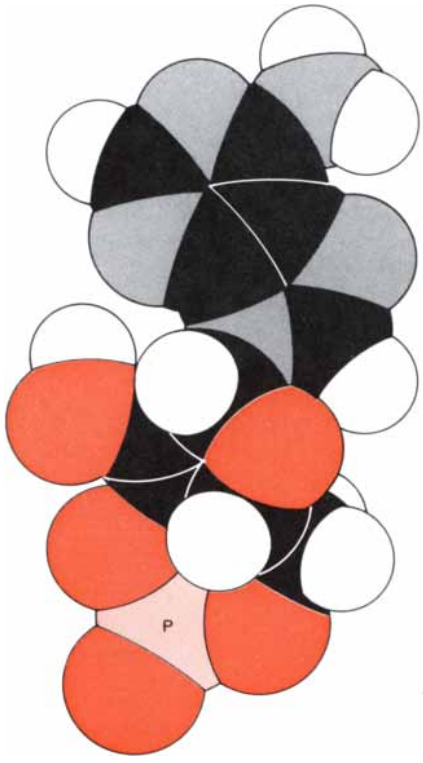
In the same experiments it was found that the nerve-ending particles also contained high levels of phosphodiesterase, an enzyme that degrades cyclic AMP into a physiologically inactive form of adenosine monophosphate. Somewhat later Noel T. Florendo and Russell J. Barrnett, working in collaboration with our laboratory at the Yale University School of Medicine, developed a cytochemical procedure to determine the location of phosphodiesterase within individual cells with the aid of the electron microscope. Using this technique, they demonstrated that the phosphodiesterase activity in the synaptic region was localized in the part of the membrane of the receiving cell thought to possess neurotransmitter receptors. This localization further suggested a role for cyclic AMP in synaptic transmission, since phosphodiesterase at the postsynaptic site would have access to, and hence be able to degrade, cyclic AMP that had been synthesized as a result of the stimulation of adenylate cyclase by a neurotransmitter.

At about this time several experiments demonstrated that electrical or neurotransmitter-induced stimulation of nerve tissue could result in large increases in cyclic AMP. In our laboratory Donald A. McAfee and Michel Schorderet, working with a sympathetic-nervous-system ganglion from the neck of rabbits and cattle, showed that electrical stimulation of the nerves innervating the ganglion, resulting in synaptic transmission, was associated with an elevation in cyclic AMP levels. McAfee and John W. Kebabian then showed that the application of the neurotransmitter dopamine to these ganglia mimicked the effect of electrical stimulation in elevating cyclic AMP levels, and conversely that the application of cyclic AMP could reproduce some of the electrophysiological effects of dopamine. Meanwhile Shiro Kakiuchi and Rall at Case Western Reserve, and somewhat later John W. Daly and his co-workers at the National Institute of Arthritis, Metabolism, and Digestive Diseases, reported that chopped pieces of brain tissue show large increases in cyclic AMP content when they are exposed to solutions containing various neurotransmitters such as norepinephrine or histamine.

Although all these experiments indi-



DOPAMINE-SENSITIVE ADENYLATE CYCLASE mediates the postsynaptic effects of the neurotransmitter dopamine and is the site of action of certain drugs that affect behavior. In this illustration dopamine released by the activity of presynaptic nerve terminals traverses the synaptic space and binds to a postsynaptic dopamine receptor, resulting in the activation of adenylate cyclase and the synthesis of cyclic AMP within the postsynaptic cell. The activity of the dopamine-sensitive adenylate cyclase is influenced by drugs known to bind specifically to dopamine receptors. Antagonist drugs such as the antischizophrenic agent chlorpromazine (Thorazine) block the receptor and prevent its activation by dopamine. Agonist drugs such as apomorphine, on the other hand, mimic the action of dopamine by activating the receptor.



SPACE-FILLING MODEL of cyclic AMP shows that the molecule is compact. It consists of a five-carbon sugar to which are attached an adenine ring (*top*) and a phosphate group.

cated that there was some kind of association between neurotransmitter receptors and cyclic AMP levels, they left unanswered the question of whether the effects of any particular neurotransmitter were coupled directly to the activation of adenylate cyclase. In order to prove a functional link between the binding of a particular neurotransmitter to its receptor and the subsequent synthesis of cyclic AMP it was necessary to demonstrate the presence of a specific neurotransmitter-sensitive adenylate cyclase, that is, an enzyme whose activity is largely dependent on the presence of a particular neurotransmitter. Before describing this enzyme in more detail it is worth discussing briefly how a neurotransmitter receptor is linked to the action of drugs that affect behavior.

Since nerve cells communicate primarily by releasing neurotransmitters at synaptic junctions, it follows that anything that interferes with the binding of a neurotransmitter to its receptor will disrupt normal nerve-cell communication and thereby alter behavior. Such interference can be produced by foreign substances, for example drugs, that have been introduced into the bloodstream. If a particular drug has a molecular configuration similar to that of a native neurotransmitter, it may be able to bind to the membrane receptor for that neurotransmitter and mimic the neurotransmitter's action. Drugs of this type are called receptor agonists. If the drug has a configuration similar to that of a neurotransmitter but not quite as similar as that of a receptor agonist, it may be able to bind to the receptor without activating it. In that case the drug will prevent the activation of the receptor by a neurotransmitter. Drugs of this type are called receptor antagonists. Other drugs that influence behavior are neither agonists nor antagonists. Some may have a mixed action by binding to the neurotransmitter receptor and causing partial activation; some may affect the receptor indirectly by altering the amount of neurotransmitter available for binding. Over the past several years this relatively simple concept of the relations among drugs, receptors and behavior has led to substantial advances in our understanding and treatment of disorders of the nervous system and has focused attention on the role cyclic AMP may play in both normal and abnormal brain function.

For example, it is generally accepted that the symptoms of Parkinson's disease, or shaking palsy, result from the degeneration of a group of neurons whose fibers project to the basal ganglia near the center of the brain from a region at the base of the brain known as the substantia nigra. These neurons secrete the neurotransmitter dopamine at their fiber terminals, and their degeneration reduces the amount of dopamine available to interact with the receptors on the receiving cells in the basal ganglia. As a result the receiving cells begin to function abnormally and cause the symptoms characteristic of Parkinson's disease: tremor, rigidity and a delay in the initiation of movement. Although dopamine is depleted in the brains of patients with the disease, the dopamine receptors in the basal ganglia appear to be undamaged.

These facts, set forth in large part by Arvid Carlsson of the University of Göteborg, Oleh Hornykiewicz of the University of Vienna and the late George C. Cotzias of the Brookhaven National Laboratory, have led to a dramatic new treatment for Parkinson's disease: the administration of the drug *levo*-dihydroxyphenylalanine (L-DOPA), the amino acid precursor of dopamine. When L-DOPA is given orally, it enters the bloodstream and travels to the brain, where it is taken up and converted to dopamine. (The neurotransmitter itself cannot enter the brain from the circulation.) The newly manufactured dopamine can then act as an agonist: it stimulates the dopamine receptors in the basal ganglia. By making up for the lack of native dopamine in this way, L-DOPA is able to reverse some of the symptoms of the disease.

Drugs that act as antagonists of the dopamine receptor are also of therapeutic value. Such drugs include the phenothiazine tranquilizer chlorpromazine (Thorazine), which is widely used in the treatment of schizophrenia. Psychotic patients treated with chlorpromazine often show a dramatic improvement in their mental symptoms, but the side effects of the drug limit its clinical usefulness. For example, after prolonged treatment patients may begin to manifest tremors and other abnormal movements similar to those seen in Parkinson's disease. When administration of the tranquilizer is stopped, the abnormal movements usually disappear.

NEUROTRANSMITTER	TYPE OF RECEPTOR
DOPAMINE	dopamine
NOREPINEPHRINE	alpha-adrenergic
	beta-adrenergic
SEROTONIN	serotonin
HISTAMINE	H_1
	H_2
ACETYLCHOLINE	muscarinic (slow)
	nicotinic (fast)
ENKEPHALIN	opiate

DRUGS AFFECTING BEHAVIOR can be conveniently separated into two categories: those that interact directly with the neurotransmitter receptor and those that affect the

Hence it seems that chlorpromazine brings on a drug-induced Parkinson's disease by blocking dopamine receptors in the basal ganglia, thereby mimicking the symptoms of dopamine depletion even though normal amounts of the neurotransmitter may be present.

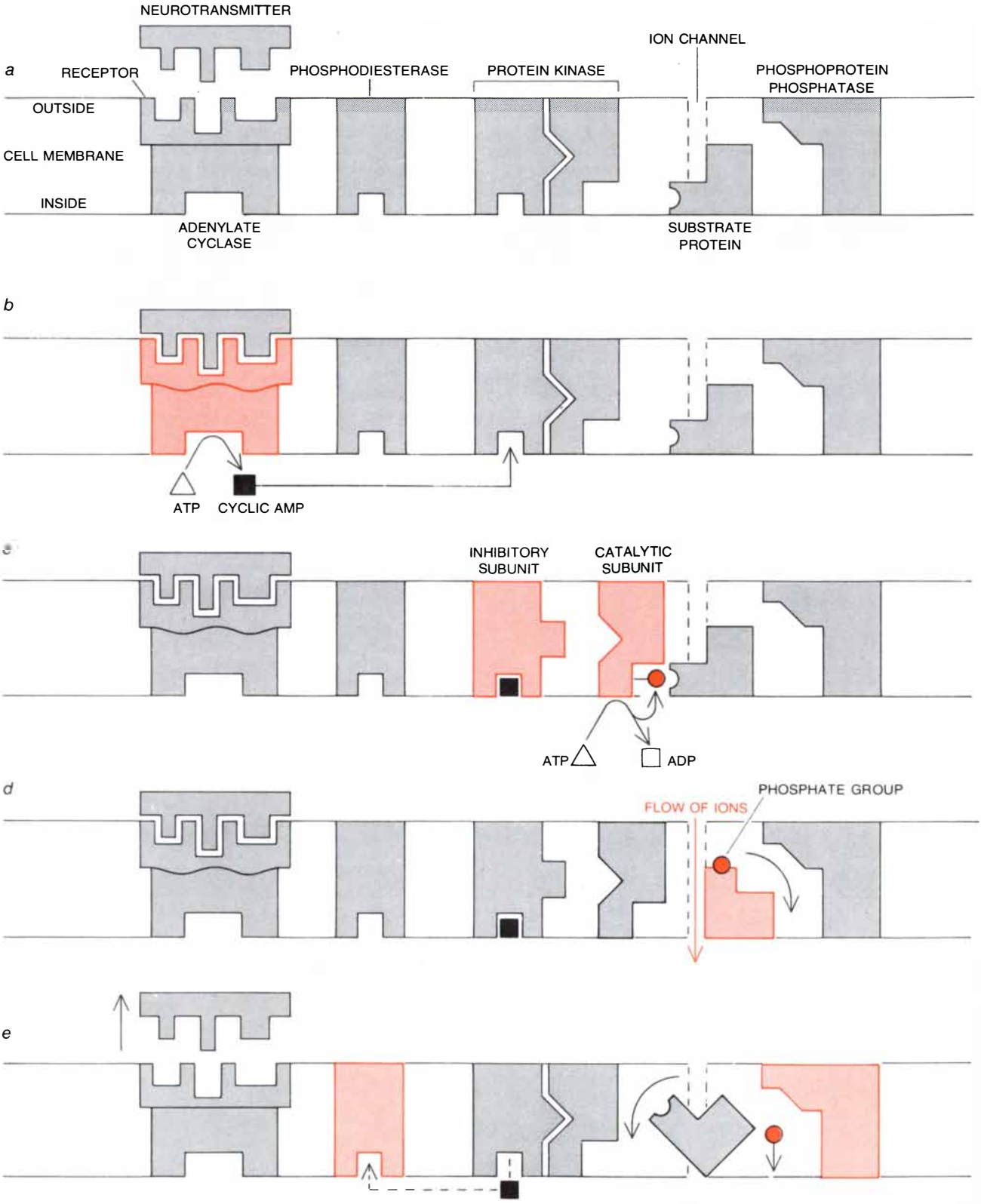
Besides being a troublesome side effect of antipsychotic drug treatment, drug-induced Parkinson's disease has shed some light on the biochemical abnormalities that may underlie schizophrenia. If drugs that appear to block dopamine receptors alleviate the symptoms of schizophrenia, then perhaps schizophrenia results from an overactivity of dopamine-containing neurons in certain parts of the brain. The overactivity of these cells would cause an excessive release of dopamine from their nerve endings, leading to an overstimulation of postsynaptic dopamine receptors. By blocking these receptors chlorpromazine apparently prevents this overstimulation and so diminishes the symptoms of schizophrenia.

These two findings—that dopamine-receptor agonists are useful agents against Parkinson's disease and that dopamine-receptor antagonists are effective agents against schizophrenia—made it highly desirable to further understand the biochemical nature of the dopamine receptor in the brain of mammals. The work of Kebabian, McAfee and Schorderet described above had suggested that cyclic AMP might play an important role in mediating neuronal responses to dopamine. In order to show that neurotransmitter binding and cyclic AMP synthesis were functionally linked, however, it was necessary to identify an adenylate cyclase that would synthesize cyclic AMP from ATP in the presence of dopamine.

In 1972 Kebabian, together with Gary L. Petzold of our laboratory, obtained experimental evidence demonstrating the presence of a dopamine-sensitive adenylate cyclase in the caudate nucleus, one of the basal ganglia of the brain and a region rich in dopamine receptors. The enzyme was located in synaptic membranes and bore a remarkable similarity to the dopamine receptor. Its activity was stimulated by very low concentrations of dopamine and was strongly inhibited by two classes of antischizophrenic drugs known to block the dopamine receptor. These results strongly suggested that the dopamine receptor in the caudate nucleus and in certain other areas of the mammalian brain might in fact be a component of the dopamine-sensitive adenylate cyclase, and that cy-

EVIDENCE FOR MEDIATION BY A CYCLIC NUCLEOTIDE		DRUGS ACTING AT THE NEUROTRANSMITTER RECEPTOR		DRUGS AFFECTING THE LEVEL OF NEUROTRANSMITTER AVAILABLE TO THE RECEPTOR	
CYCLIC AMP	CYCLIC GMP	AGONISTS (ACTIVATE RECEPTOR)	ANTAGONISTS (BLOCK RECEPTOR)	INCREASE LEVELS	DECREASE LEVELS
yes	no	apomorphine alpha-bromcryptine	antischizophrenics (Thorazine, Haldol)	*levo*-dihydroxyphenylalanine (L-DOPA) amantadine amphetamines (Dexedrine) methylphenidate (Ritalin)	alpha-methyl-para-tyrosine
no	yes (?)	phenylephrine (Neosynephrine)	phentolamine (Regitine)	tricyclic antidepressants (Elavil, Tofranil) MAO inhibitors (Parnate) amphetamines cocaine	reserpine (Serpasil) alpha-methyl-dopa (Aldomet)
yes	no	epinephrine (adrenalin) isoproterenol (Isuprel)	propranolol (Inderal)		
yes	no	5-methoxy-N,N-dimethyltryptamine	lysergic acid diethylamide (LSD) methysergide (Sansert)	tricyclic antidepressants (Elavil, Anafranil) tryptophan	para-chloro-phenylalanine
no	yes (?)	2-methyl-histamine	diphenhydramine (Benadryl) dimenhydrinate (Dramamine)	histidine amodiaquine	alpha-hydrazino-histidine brocresine
yes	no	betazole (Histalog)	metiamide cimetidine		
no	yes (?)	pilocarpine carbachol bethanechol (Urecholine)	scopolamine atropine (belladonna) propantheline (Pro-Banthine)	insecticides (Parathion, Malathion) nerve gas (DFP) pyridostigmine (Mestinon)	botulinus toxin
no	no	nicotine	d-tubocurarine (curare) succinylcholine (Anectine)		
yes	?	morphine heroin methadone meperidine (Demerol)	naloxone (Narcan)		

receptor indirectly by altering the amount of neurotransmitter available for binding. Drugs affecting neurotransmitter levels work in various ways, such as increasing the synthesis of a neurotransmitter (L-DOPA), inducing the release of preexisting neurotransmitter (amphetamine) or blocking the breakdown or sequestration of neurotransmitter (certain antidepressants). Cyclic AMP and a related compound, cyclic GMP, appear to mediate the actions of several neurotransmitters and to play a role in the action of many behavioral drugs.



THEORETICAL MODEL explains how the increase in the level of cyclic AMP inside a receiving nerve cell can result in a transient change in the permeability of the cell membrane to ions, thereby altering the electrical excitability of the cell. In panel *a* the several molecular components that have been implicated in the process are shown in highly schematic form in their resting state. When an electrical impulse reaches the presynaptic terminal, neurotransmitter is released. It then crosses the synaptic space and binds to receptors on the receiving cell, activating adenylate cyclase to convert ATP into cyclic AMP (*b*). The cyclic AMP generated in this manner binds to the inhibitory subunit of protein kinase, causing it to dissociate from the catalytic subunit. Now activated, the catalytic subunit transfers a phosphate group from ATP to a substrate protein (*c*). The addition of phosphate changes the shape or position of the substrate protein, allowing certain ions to flow through existing pores in the membrane and altering the electrical excitability of the cell (*d*). Termination of the reaction begins when the neurotransmitter dissociates from the receptor, stopping further synthesis of cyclic AMP. Phosphodiesterase rapidly inactivates the remaining supply. Finally, phosphoprotein phosphatase removes phosphate group from the substrate protein (*e*).

clic AMP might therefore mediate the intracellular action of dopamine at certain synapses.

Further investigations by Yvonne Clement-Cormier and Kebabian in our laboratory and by Leslie L. Iversen and his co-workers at the University of Cambridge have largely confirmed that suggestion: among a large number of substances there is a remarkable parallel between the ability to act as an agonist or antagonist of the dopamine receptor and the ability to activate or inhibit the dopamine-sensitive adenylate cyclase. This correlation has since led to a new methodology for the rapid screening and development of drugs as potential dopamine receptor agonists (agents against Parkinson's disease) or antagonists (agents against schizophrenia). Whereas the traditional methods of measuring dopamine receptor activity involved behavioral tests that were time-consuming and often imprecise, the determination of the activity of the dopamine-sensitive adenylate cyclase in the presence of the drug under investigation provides a rapid and quantitative way of evaluating the drug's ability to either block or activate the dopamine receptor.

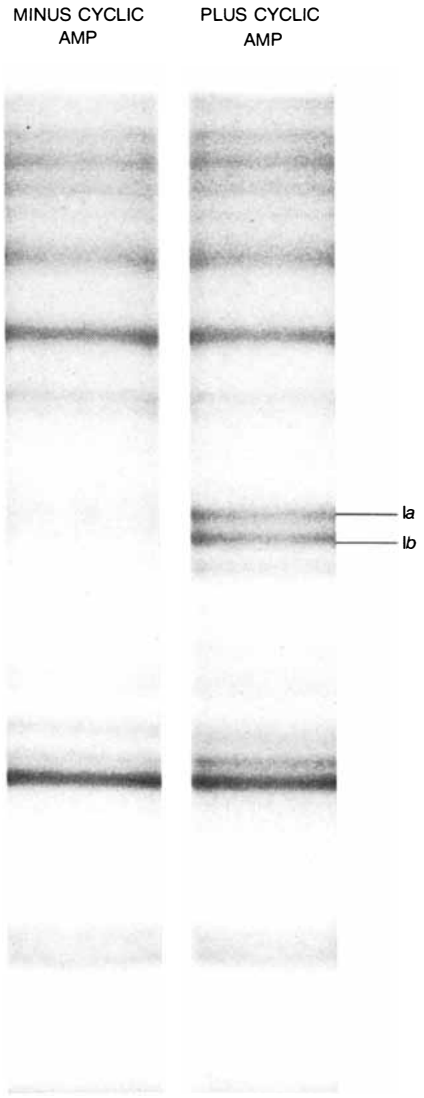
A link between cyclic AMP and synaptic transmission was also indicated when Floyd E. Bloom, Barry Hoffer and George Siggins, working at the National Institute of Mental Health, found physiological evidence that some of the actions of the neurotransmitter norepinephrine are mediated by cyclic AMP. At that time Bloom and his colleagues were studying the regulation by norepinephrine of neuronal activity in the cerebellum, which controls many of the automatic movements of the body, such as walking. The key elements that regulate such movements are large, elaborately branching neurons called Purkinje cells found in the cortex, or outer layer, of the cerebellum. Extending earlier anatomical work by Tomas Hokfelt and Kjell Fuxe at the Karolinska Institute in Stockholm, Bloom and his colleagues found that the locus coeruleus, a small cluster of norepinephrine-containing neurons buried deep in the brain stem, sends fine connections to the cortex of the cerebellum, and that stimulation of this pathway causes the firing rate of the Purkinje cells to decrease markedly. When they administered measured amounts of norepinephrine or cyclic AMP to individual Purkinje cells with a micropipette, they observed that the application of either compound slowed the firing rate of these cells in the same way that stimulation of the locus coeruleus did. Moreover, by means of a fluorescent-labeling technique that stains selectively for cyclic AMP they were able to show that either stimulation of the locus coeruleus or the direct application of norepinephrine causes a dramatic increase in the cyclic AMP content of the Purkinje cells. These and other results have suggested that the effects of norepinephrine on Purkinje-cell firing are mediated through the stimulation of a norepinephrine-sensitive adenylate cyclase and the resulting synthesis of cyclic AMP inside the cell.

Recent experiments in our laboratory have indicated that the role of cyclic AMP in the functioning of neurotransmitter receptors is by no means limited to the nervous system of vertebrates. For example, we have found in the thoracic ganglia of insects an enzyme that may mediate the effects of the neurotransmitter serotonin. This serotonin-sensitive adenylate cyclase is activated by very low concentrations of serotonin, and the effect of serotonin on the activity of the enzyme is specifically inhibited by drugs that are known to block serotonin receptors. Interestingly enough, one of the most potent of these blocking agents is the hallucinogen lysergic acid diethylamide (LSD). Our findings suggest that the serotonin receptor of neural tissue is intimately associated with a serotonin-sensitive adenylate cyclase and that some of the physiological effects of LSD may result from the inhibition of the enzyme.

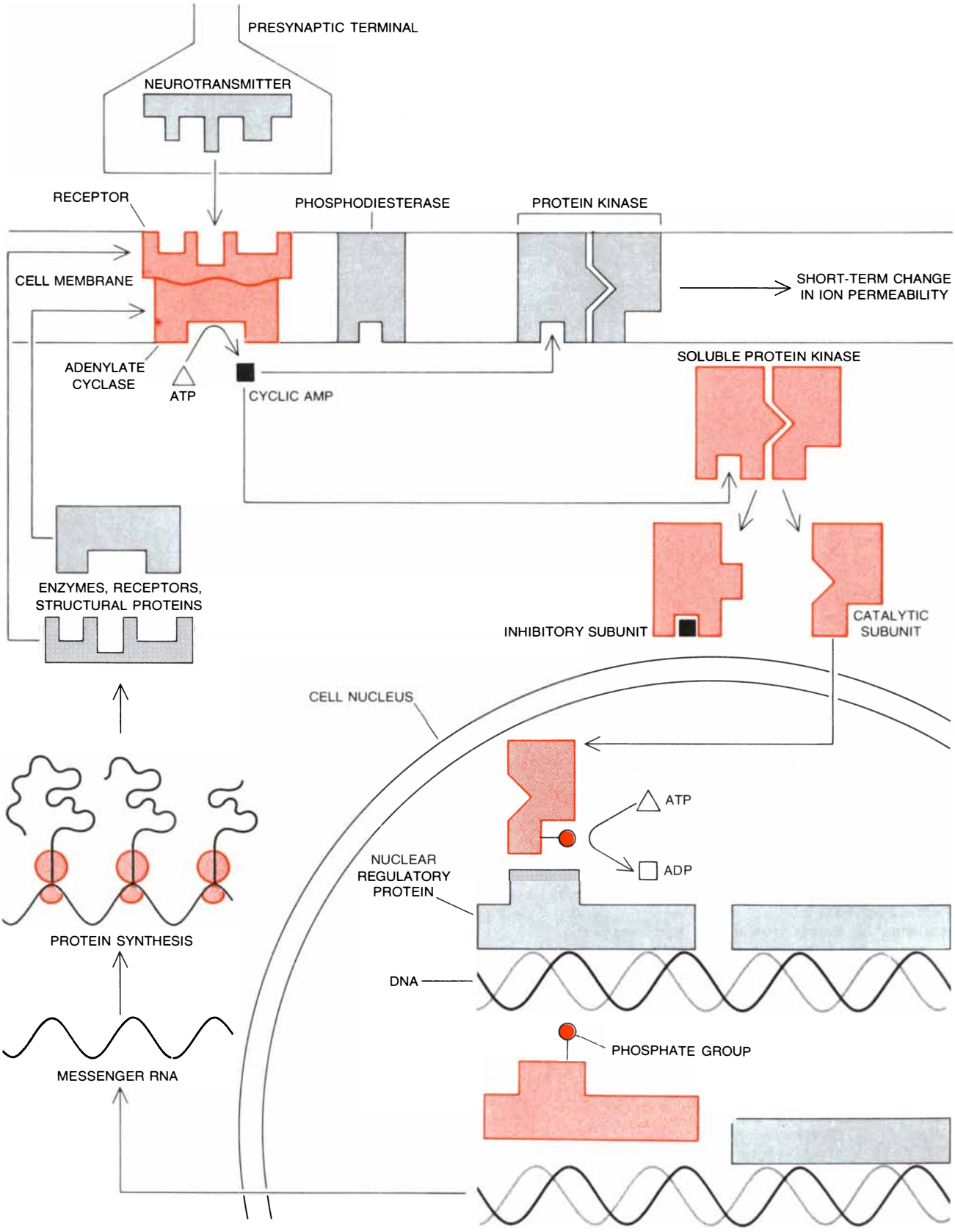
At least five neurotransmitters have now been shown to stimulate specific neurotransmitter-sensitive adenylate cyclases: dopamine, norepinephrine, serotonin, histamine and octopamine. Moreover, recently acquired evidence suggests that a second cyclic nucleotide, cyclic guanosine monophosphate (cyclic GMP), may be involved in the effects of the neurotransmitter acetylcholine at certain synapses through the activation of guanylate cyclase, an enzyme that converts guanosine triphosphate (GTP) into cyclic GMP. Cyclic GMP may also be involved in mediating the effects of norepinephrine and histamine at certain receptors that are distinct from those associated with the cyclic AMP system.

It is worth noting that many of the behavioral drugs that affect neurotransmitter receptors produce effects resembling those that occur naturally in patients with mental or neurological diseases. For example, miners exposed to manganese poisoning or patients given antischizophrenic drugs exhibit a syndrome indistinguishable from naturally occurring Parkinson's disease, and LSD can produce hallucinations similar to those experienced by schizophrenics. These and other clinical observations raise the possibility that certain neurological and mental diseases may result from abnormalities in specific receptor-adenylate-cyclase systems. Such abnormalities might arise genetically through nonlethal mutations or be acquired through exposure to high levels of environmental toxins such as manganese.

Not all neurotransmitter receptors,



PHOSPHORYLATION of synaptic-membrane proteins in response to cyclic AMP is highly specific, as is shown in this experiment performed by Bruce K. Krueger, Javier Forn and Tetsufumi Ueda in the authors' laboratory. Synaptic membranes prepared from rat brain were incubated in a solution containing radioactive ATP in the presence or absence of added cyclic AMP, enabling molecules of protein kinase in the membrane to transfer a radioactive phosphate group from ATP to various substrate proteins. The phosphorylated proteins were then removed from the membrane with a detergent and separated from one another by placing a mixture of them at the top of a column of the gel-like substance polyacrylamide. Application of a high-voltage current caused the proteins to migrate downward and separate into a series of thin bands according to their size. By next placing the gels in a darkroom on top of X-ray film it was possible to detect which bands exposed the film and hence had incorporated radioactive phosphate. The print of an exposed film shown here reveals that although many proteins were phosphorylated by enzymes whose activity was independent of cyclic AMP, only two proteins were phosphorylated by a protein kinase whose activity was dependent on presence of cyclic AMP. The proteins are designated 1*a* (molecular weight of 86,000 daltons) and 1*b* (80,000 daltons).



LONG-TERM EFFECTS of cyclic AMP on nerve-cell function are postulated to occur through the activation of a soluble protein kinase located in the cytoplasm. Once activated, this enzyme enters the cell nucleus, where it transfers a phosphate group from ATP to one of the nuclear regulatory proteins closely associated with the DNA of inactive genes. Phosphorylation alters the shape or binding characteristics of the regulatory protein in such a way that it dissociates from the DNA double helix, exposing the underlying stretch of DNA so that protein synthesis can occur. Proteins whose synthesis might be induced in this manner include neurotransmitter receptors, adenylate cyclases, structural proteins and enzymes involved in neurotransmitter synthesis or breakdown. In this way the cyclic AMP generated during the short-term process of synaptic transmission could produce long-term alterations in the electrical properties of the receiving cell.

however, have been linked to the synthesis of a cyclic nucleotide. For example, one of the best-studied synaptic mechanisms—the stimulation of the contraction of voluntary muscle through the release of acetylcholine at the junction between a nerve fiber and a muscle—appears to operate independently of the synthesis of a cyclic nucleotide [see "The Response to Acetylcholine," by Henry A. Lester; SCIENTIFIC AMERICAN, February]. This makes sense in view of the fact that the mediation of synaptic transmission by cyclic nucleotides involves a complex series of steps, which we shall discuss, that makes the process relatively long-lasting on the time scale of neuronal events. Those types of transmission that place a premium on speed, such as the contraction of voluntary muscle, therefore rely on faster receptor mechanisms.

So far we have described how certain neurotransmitters, by binding to a specific receptor, stimulate the production of cyclic AMP inside the receiving cell. The question remains: How does an increase in the level of cyclic AMP translate the message of the neurotransmitter into physiological action?

We know from the work of many neurobiologists over the past 20 years that a nerve cell responds to synaptic stimulation with a brief change in the permeability of the synaptic membrane to one or more kinds of ion. This change in permeability allows ions to flow across the membrane, generating an electric current and thereby altering the membrane's electric potential, or voltage. Depending on the ions that move and the direction of their movement, the change in electric potential induced by a neurotransmitter will make it either more or less likely that the cell will reach the threshold of excitation necessary for it to generate an impulse. In this respect a neuron acts much like an analogue computer, integrating the many hundreds of inhibitory and excitatory chemical messages that impinge on its surface at any given moment before deciding whether or not to fire.

In order to explore the possible role of cyclic AMP in mediating some of the changes in membrane permeability induced by neurotransmitters, we began several years ago a detailed investigation of the chain of biochemical events that occur inside a neuron following the binding of a neurotransmitter to its receptor and the activation of adenylate cyclase. Earlier experiments by Edwin G. Krebs and his colleagues at the University of Washington had provided some important clues. In their investigation of the mechanism by which adrenalin induces the conversion of glycogen into glucose in muscle cells, they found that the cyclic AMP generated inside the cells by the hormone-sensitive adenylate cyclase caused the transfer of a phosphate group from ATP to the enzyme responsible for the initiation of glycogen breakdown. Somewhat later they discovered that the actual transfer of the phosphate group, the process called phosphorylation, was accomplished by the enzyme protein kinase. In biochemical nomenclature, kinase (more properly, phosphokinase) is a term reserved for enzymes that transfer a phosphate group from ATP to another molecule; for a protein kinase the other molecule is always a protein.

Cyclic AMP is thought to activate protein kinase by binding to an inhibitory subunit of the kinase in such a way as to change its shape and cause it to dissociate from the rest of the enzyme. The loss of the inhibitory component enhances the activity of the kinase, which then readily transfers a phosphate group to the enzyme controlling the pathway of glycogen breakdown. The addition of a phosphate group to this enzyme causes it to trigger a "cascade" of enzymatic reactions, leading ultimately to the breakdown of glycogen. Protein kinase therefore acts as the link between the cyclic AMP generated by the hormone and the activation of the biochemical pathway that accounts for the hormone's metabolic effects.

On the basis of this scheme we began looking for an analogous system in brain cells. Although glycogen breakdown is not a prominent effect of cyclic AMP in nervous tissue, we felt that a similar protein kinase could translate the change in cyclic AMP levels brought about by a neurotransmitter into a change in the permeability of the postsynaptic membrane to ions.

Within a relatively short time J.-F. Kuo and Eishichi Miyamoto, working in our laboratory, isolated a cyclic-AMP-dependent protein kinase from cattle brain. The enzyme was present in relatively large amounts in the brain tissue compared with most other tissues, and its activity was notably dependent on the presence of cyclic AMP. Furthermore, the concentration of cyclic AMP required to stimulate the enzyme to half its maximum activity was similar to the concentration of cyclic AMP normally found in brain tissue. This finding indicated that the increases in cyclic AMP levels known to be induced by various neurotransmitters would cause large changes in the activity of the cyclic-AMP-dependent protein kinase.

Other experiments demonstrated that the protein kinase could transfer a phosphate group to proteins other than those associated with the breakdown of glycogen. In fact, studies by Hiroo Maeno and Edward Johnson in our laboratory indicated that when brain tissue was homogenized, the proteins phosphorylated to the greatest extent were in the fractions of the homogenized tissue with the largest amounts of synaptic-membrane fragments, and that these fractions also had the greatest concentration of protein kinase.

Thus it became evident that the synaptic membrane contains all the major elements for the neurotransmitter-induced transfer of a phosphate group to a substrate protein in the membrane: (1) a neurotransmitter-sensitive adenylate cyclase that generates cyclic AMP in response to a specific neurotransmitter, (2) a cyclic-AMP-dependent protein kinase that phosphorylates a substrate protein in the presence of cyclic AMP and (3) the membrane substrate protein that is phosphorylated by the protein kinase. The transfer of a phosphate group to such a protein could conceivably change the permeability of the membrane to ions, either directly by changing the configuration of the protein in order to open a channel that ions can flow through, or indirectly, for example by affecting the activity of an enzyme "pump" that physically transports ions across the membrane.

Further investigation showed that the synaptic-membrane fractions also contain the enzymatic machinery necessary for stopping the cyclic-AMP-dependent phosphorylation process and restoring the membrane to its resting state. These enzymes include not only phosphodiesterase, which degrades cyclic AMP to an inactive form of AMP, but also phosphoprotein phosphatase, which removes the phosphate group from proteins that have previously been phosphorylated by the cyclic-AMP-dependent protein kinase. Several of the components of this membrane-bound enzyme system have been shown to exist as a complex, thereby reducing the distance reaction products must travel from one enzyme to the next and enabling a phosphate group to be rapidly added to and removed from the substrate protein.

Following the discovery that fractions of homogenized brain tissue containing the largest amounts of synaptic membrane serve as excellent substrates for the cyclic-AMP-dependent protein kinase, Tetsufumi Ueda, Bruce K. Krueger and Javier Forn of our laboratory, together with Johnson and Maeno, attempted to isolate the specific protein or proteins that were being phosphorylated. Up to this point it had not been possible to tell whether a large percentage or only a very few of the dozens of proteins known to be present in the synaptic membrane were acting as substrates for phosphorylation. When Ueda and his co-workers incubated synaptic-membrane fractions in a solution containing radioactively labeled ATP in the presence or absence of cyclic AMP, they found that the phosphorylation of only two or three of the several dozen proteins in the membrane fraction was markedly increased by cyclic AMP. This result was gratifying, since it indi-

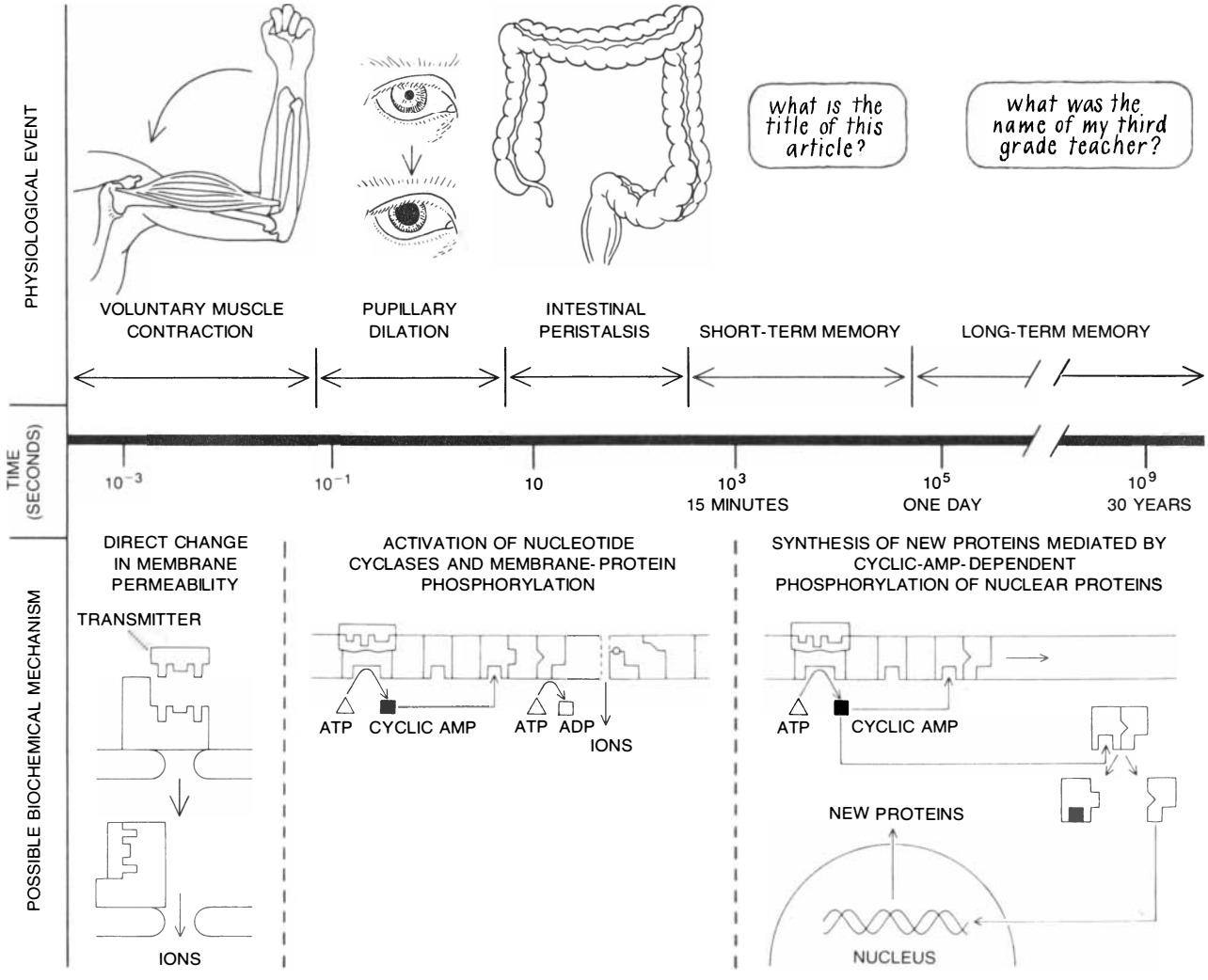
cated that the effects of cyclic AMP on the phosphorylation of synaptic-membrane protein were specific and did not simply involve a general stimulation of cellular metabolism.

One of the synaptic-membrane proteins specifically phosphorylated by the cyclic-AMP-dependent protein kinase, designated Protein I, has recently been investigated in detail in our laboratory. Protein I, which has two molecular subunits, I*a* and I*b*, is localized almost exclusively in the synaptic region of nerve tissue and is apparently absent from those organelles of nerve cells (such as nuclei, mitochondria and ribosomes) that are not directly associated with synaptic transmission. Moreover, Protein I has not been found at all in any nonneural tissue (such as heart or liver) yet examined. It is also absent from the brain of fetal rats at stages of development before the formation of connections between nerve cells, and it appears for the first time at the stage when synaptic complexes form.

The phosphorylation of Protein I by protein kinase in the synaptic membrane is extremely fast. In fact, in the shortest period that has yet been accurately studied—five seconds—the phosphorylation is already at the maximum level. The speed of this reaction is virtually a prerequisite to seriously considering the possibility that Protein I might be involved in the generation of the very brief changes in the permeability of the synaptic membrane to ions (lasting for a few hundred milliseconds or less) that give rise to postsynaptic potentials.

If protein phosphorylation mediated by cyclic AMP is indeed responsible for changes in the permeability of the synaptic membrane to ions, then there should be a correlation between the state of phosphorylation of a particular membrane protein or proteins and the state of ion permeability of the cell membrane. The short duration of changes in the permeability of membranes in the nervous system has made it difficult to establish such a correlation within the limitations of present methodology. For this reason we have turned to some non-neuronal model systems, including the red blood cell of the turkey and the frog, to study such permeability changes in detail. Under appropriate conditions these cells respond to hormones such as adrenalin with an increase in the movement of sodium and potassium ions across the cell membrane. This increase, which is due to a change in membrane permeability, appears to be mediated by cyclic AMP.

Recently Stephen A. Rudolph of our laboratory has found that a single large membrane protein becomes phosphorylated whenever cyclic AMP or the adrenalin-related drug isoproterenol is added to the suspension of red blood cells. Furthermore, by simultaneously



CONTINUUM OF SYNAPTIC EVENTS in the nervous system stretches from the very brief to the very long-lived. At the short end of the spectrum are processes, such as voluntary muscle contraction, that are initiated in the span of a few milliseconds by the direct action of a neurotransmitter to open an ion channel in the membrane. Events lasting from hundreds of milliseconds to minutes appear to be mediated indirectly through the neurotransmitter-induced synthesis of cyclic AMP, which initiates the phosphorylation of membrane proteins to produce a relatively slow change in ion permeability. Events ranging from hours to years in duration, such as memory, may involve the synthesis of new proteins directed by the cyclic-AMP-stimulated phosphorylation of proteins that regulate gene expression.

measuring the state of phosphorylation of this protein and the movement of sodium ions into the cell in the presence of isoproterenol he has found that the time required for the membrane protein to reach half its maximum level of phosphorylation coincides closely with the time required for the flow of sodium ions across the membrane to reach half its maximum rate.

The results of these and other experiments conducted with a variety of systems demonstrate a close correlation between the movement of ions across cell membranes and the cyclic-AMP-dependent phosphorylation of specific membrane proteins. It should be pointed out, however, that correlation does not always mean causality. There is at present no direct proof that phosphorylation is a necessary event for the observed changes in permeability; both events could conceivably be secondary effects of some other process. The definitive proof of a causal connection between membrane-protein phosphorylation and permeability changes remains one of the major challenges for experimentation. We hope it will eventually be possible to dissect out the individual molecular components responsible for controlling ion permeability, put them into synthetic membranes and determine whether prior cyclic-AMP-dependent phosphorylation of one or more of these components will result in permeability changes similar to those observed in intact cells.

Because of the complex nature of cyclic-AMP-mediated processes, involving several biochemical steps, these processes are best suited to the regulation of synaptic events that are relatively long-lasting. This may be the reason the neuronal pathways that appear to make use of cyclic nucleotides often play a modulatory role in the nervous system, regulating activity rather than initiating it. For example, the cyclic AMP formed in the cerebellum by the activation of norepinephrine fibers in the locus coeruleus tones down the "spontaneous" firing rate of the Purkinje neurons. In ganglia of the sympathetic nervous system the dopamine synapses mediated by cyclic AMP adjust the level of activity at other synapses not mediated by cyclic AMP. Even in its role in movement disorders such as Parkinson's disease the cyclic AMP formed in the caudate nucleus by the activation of dopamine-sensitive adenylate cyclase appears to regulate movements only after they have been initiated by signals from other areas of the brain.

Observed changes in membrane excitability seen at cyclic-AMP-mediated synapses in the cerebellum and in the sympathetic ganglion may last for hundreds of milliseconds or longer—a long time on the scale of neural events. Recent investigations by Eric R. Kandel of Columbia University College of Physicians and Surgeons, working with a ganglion in a mollusk, and by Benjamin Libet of the University of California School of Medicine at San Francisco, working with a ganglion in the sympathetic nervous system of the rabbit, have demonstrated events of even greater duration that appear to be related to cyclic AMP, some of them lasting for hours. Observations such as these support the possibility that synaptic events mediated by cyclic nucleotides could be the basis for certain long-term changes in the central nervous system of man.

This notion has been supported by evidence indicating that the phosphorylation of protein mediated by cyclic AMP may influence events in the nucleus of the cell. A number of investigators, starting with Thomas A. Langan of the University of Colorado School of Medicine, have shown that histones, positively charged proteins that bind to the negatively charged phosphate backbone of the double helix of DNA, are susceptible to phosphorylation by a cyclic-AMP-dependent protein kinase. When histones are intimately associated with DNA, they appear to inhibit the expression of the DNA's information content by blocking access to the enzymes that effect the transcription of DNA to RNA. Phosphorylation by protein kinase makes the charge of the histones more negative, thereby reducing their ability to bind to DNA.

From a mechanistic point of view the linkage between adenylate cyclase and protein kinase supplies a ready-made system for the transformation of short-term synaptic events into longer-lived biochemical changes. The cyclic AMP generated as a result of synaptic activity could conceivably activate a protein kinase, which would phosphorylate histone or some other nuclear regulatory protein and cause its removal from the DNA of inactive genes. The exposed genes would then be available for transcription into messenger RNA and ultimately translation into protein. Indeed, experiments in the laboratory of Erimino Costa at the National Institute of Mental Health suggest that cyclic AMP can induce the synthesis of tyrosine hydroxylase, an enzyme involved in the manufacture of dopamine and norepinephrine.

Thus the cyclic AMP system provides a mechanism by which the continued activation of a particular synapse could lead to the synthesis of new enzyme molecules or new receptor molecules, permanently altering the electrical properties of the receiving neuron. Whether such changes could constitute a molecular basis for information storage in the nervous system is hard to say, but it is certainly an attractive hypothesis. One can even speculate that the cyclic-AMP-dependent phosphorylation of synaptic-membrane proteins represents a type of short-term memory and that the phosphorylation of nuclear proteins and the synthesis of new protein molecules could represent a more permanent change—a long-term memory.

In sum, there exists in the brain a continuum of functional events that stretches from the brief to the very long-lived. At the short end of the scale are synaptic events, such as the contraction of voluntary muscle, that last for only a few milliseconds; at the long end are memories that can last for 50 years or more. On the basis of our present knowledge it seems possible that all but the very briefest of these events may involve cyclic-AMP-associated mechanisms.

Most of this discussion has emphasized the possible role of cyclic AMP in synapses and its importance in mediating the effects of many neurotransmitters and the actions of drugs that affect behavior. These areas have been a major focus of research on the role of cyclic nucleotides in the nervous system largely because of the conceptual and mechanistic analogies between the action of hormones and the transmission of nerve impulses. Mostly because of a comparative lack of evidence other roles for cyclic nucleotides in the nervous system have been somewhat underplayed. Considerable research over the past two or three years, both in our laboratory and elsewhere, suggests, however, that processes mediated by cyclic nucleotides are involved not only in the mechanism of changes in ion permeability induced by neurotransmitters but also in the regulation of a variety of other phenomena, including the activation of enzymes, the synthesis and release of neurotransmitters, intracellular movements, carbohydrate metabolism in the cerebrum and possibly even processes of growth and development. We shall not elaborate on these recent advances, but we should like to emphasize their potential importance to an overall understanding of the functioning of the nervous system.

In particular, it may be useful to view the synaptic and nonsynaptic actions of cyclic AMP not so much in isolation but rather as part of an integrating system. It may be that enough synaptic stimulation not only raises cyclic AMP levels sufficiently to alter the permeability of membranes but also initiates a logical sequence of events mediated by cyclic nucleotides. This sequence might include a decrease or an increase in the synthesis of a neurotransmitter in response to synaptic stimulation, an initiation of intracellular movements in order to transport newly synthesized products, the activation of carbohydrate metabolism to supply the necessary cellular energy requirements, and direct effects on the genetic material in the cell nucleus that may lead to long-term alterations of behavior, such as memory.

MATHEMATICAL GAMES

A new kind of cipher that would take millions of years to break

by Martin Gardner

"Few persons can be made to believe that it is not quite an easy thing to invent a method of secret writing which shall baffle investigation. Yet it may be roundly asserted that human ingenuity cannot concoct a cipher which human ingenuity cannot resolve."

—EDGAR ALLAN POE

The upward creep of postal rates accompanied by the deterioration of postal service is a trend that may or may not continue, but as far as most private communication is concerned, in a few decades it probably will not matter. The reason is simple. The transfer of information will probably be much faster and much cheaper by "electronic mail" than by conventional postal systems. Before long it should be possible to go to any telephone, insert a message into an attachment and dial a number. The telephone at the other end will print out the message at once.

Government agencies and large businesses will presumably be the first to make extensive use of electronic mail, followed by small businesses and private individuals. When this starts to happen, it will become increasingly desirable to have fast, efficient ciphers to safeguard information from electronic eavesdroppers. A similar problem is involved in protecting private information stored in computer memory banks from snoopers who have access to the memory through data-processing networks.

It is hardly surprising that in recent years a number of mathematicians have asked themselves: Is it possible to devise a cipher that can be rapidly encoded and decoded by computer, can be used repeatedly without changing the key and is unbreakable by sophisticated cryptanalysis? The surprising answer is yes. The breakthrough is scarcely two years old, yet it bids fair to revolutionize the entire field of secret communication. Indeed, it is so revolutionary that all previous ciphers, together with the techniques for cracking them, may soon fade into oblivion.

An unbreakable code can be unbreakable in theory or unbreakable only in practice. Edgar Allan Poe, who fancied himself a skilled cryptanalyst, was convinced that no cipher could be invented that could not also be "unriddled." Poe was certainly wrong. Ciphers that are unbreakable even in theory have been in use for half a century. They are "one-time pads," ciphers that are used only once, for a single message. Here is a simple example based on a shift cipher, sometimes called a Caesar cipher because Julius Caesar used it.

First write the alphabet, followed by the digits 0 through 9. (For coding purposes 0 represents a space between words, and the other digits are assigned to punctuation marks.) Below this write the same sequence cyclically shifted to the right by an arbitrary number of units, as is shown in color in the illustration on this page. Our cipher consists in taking each symbol in the plaintext (the message), finding it in the top row, and replacing it with the symbol directly below it. The result is a simple substitution cipher, easily broken by any amateur.

In spite of its simplicity, a shift cipher can be the basis of a truly unbreakable code. The trick is simply to use a different shift cipher for each symbol in the plaintext, each time choosing the amount of shift at random. This is easily done with the spinner shown in the top illustration on the opposite page. Suppose the first word of plaintext is THE. We spin the arrow and it stops on K. This tells us to use for encoding T a Caesar cipher in which the lower alphabet is shifted 10 steps to the right, bringing A below K as is shown in the illustration. T, therefore, is encoded as J. The same procedure is followed for every symbol in the plaintext. Before each symbol is encoded, the arrow is spun and the lower sequence is shifted accordingly. The result is a ciphertext starting with J and a cipher "key" starting with K. Note that the cipher key will be the same length as the plaintext.

To use this one-time cipher for sending a message to someone—call him *Z*—we must first send *Z* the key. This can be done by a trusted courier. Later we send to *Z*, perhaps by radio, the ciphertext. *Z* decodes it with the key and then destroys the key. The key must not be used again because if two such ciphertexts were intercepted, a cryptanalyst might have sufficient structure for breaking them.

It is easy to see why the one-time cipher is uncrackable even in principle. Since each symbol can be represented by any other symbol, and each choice of representation is completely random, there is no internal pattern. To put it another way, any message whatever having the same length as the ciphertext is as legitimate a decoding as any other. Even if the plaintext of such a coded message is found, it is of no future help to the cryptanalyst because the next time the system is used the randomly chosen key will be entirely different.

One-time pads are in constant use today for special messages between high military commanders, and between governments and their high-ranking agents. The "pad" is no more than a long list of random numbers, perhaps printed on many pages. The sender and receiver must of course have duplicate copies. The sender uses page 1 for a cipher, then destroys the page. The receiver uses his page 1 for decoding, then destroys his page. When the Russian agent Rudolf Abel was captured in New York in 1957, he had a one-time pad in the form of a booklet about the size of a postage stamp. David Kahn, who tells the story in his marvelous history *The Codebreakers,* says that the one-time pad is the standard method of secret radio communication used by the U.S.S.R. The famous "hot line" between Washington and Moscow also makes use of a one-time pad, the keys being periodically delivered through the two embassies.

If the one-time pad provides absolute secrecy, why is it not used for all secret communication? The answer is that it is too impractical. Each time it is employed a key must be sent in advance, and the key must be at least as long as the anticipated message. "The problem of producing, registering, distributing and canceling the keys," writes Kahn, "may seem slight to an individual who has not had experience with military communications, but in wartime the volumes of traffic stagger even the signal staffs. Hundreds of thousands of words may be enciphered in a day; simply to generate the millions of key characters required would be enormously expensive and time-consuming. Since each

A Caesar cipher with a 10-shift

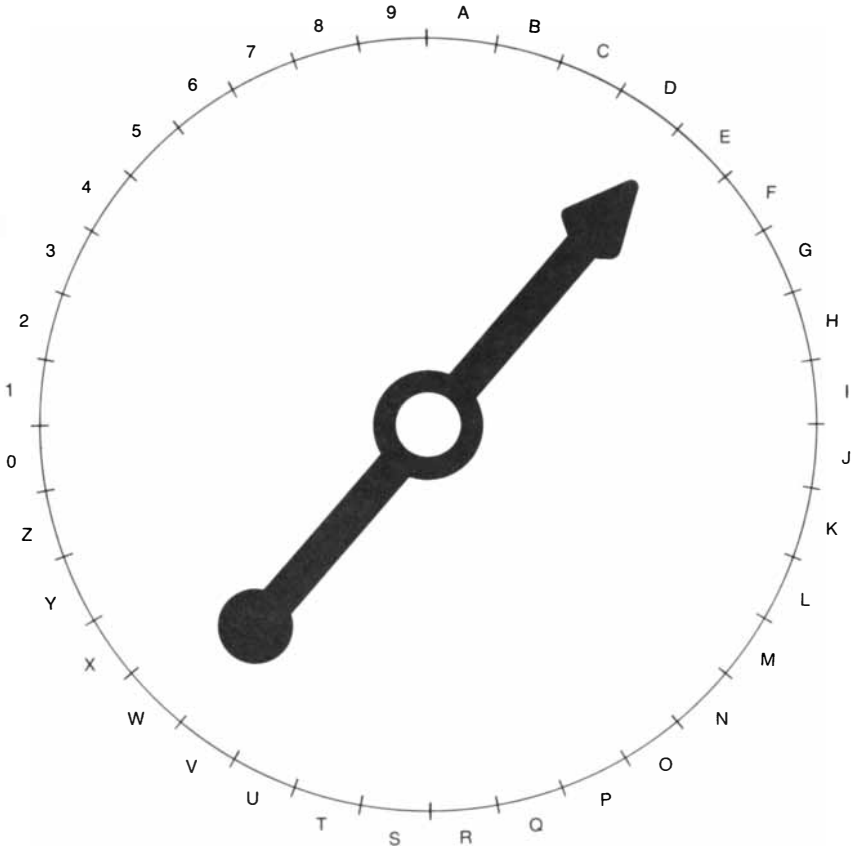
message must have its unique key, application of the ideal system would require shipping out on tape at the very least the equivalent of the total communications volume of a war."

Let us qualify Poe's dictum by applying it only to ciphers that are used repeatedly without any change in the key. Until recently all cipher systems of this kind were known to be theoretically breakable provided the code breaker has enough time and enough ciphertext. Then in 1975 a new kind of cipher was proposed that radically altered the situation by supplying a new definition of "unbreakable," a definition that comes from the branch of computer science known as complexity theory. These new ciphers are not absolutely unbreakable in the sense of the one-time pad, but in practice they are unbreakable in a much stronger sense than any cipher previously designed for widespread use. In principle these new ciphers can be broken, but only by computer programs that run for millions of years!

The two men responsible for this remarkable breakthrough are Whitfield Diffie and Martin E. Hellman, both electrical engineers at Stanford University. Their work was partly supported by the National Science Foundation in 1975 and was reported in their paper "New Directions in Cryptography" (*IEEE Transactions on Information Theory,* November, 1976). In it Diffie and Hellman show how to create unbreakable ciphers that do not require advance sending of a key or even concealment of the method of encoding. The ciphers can be efficiently encoded and decoded, they can be used over and over again and there is a bonus: the system also provides an "electronic signature" that, unlike a written signature, cannot be forged. If Z receives a "signed" message from A, the signature proves to Z that A actually sent the message. Moreover, A's signature cannot be forged by an eavesdropper or even by Z himself!

These seemingly impossible feats are made possible by what Diffie and Hellman call a trapdoor one-way function. Such a function has the following properties: (1) it will change any positive integer x to a unique positive integer y; (2) it has an inverse function that changes y back to x; (3) efficient algorithms exist for computing both the forward function and its inverse; (4) if only the function and its forward algorithm are known, it is computationally infeasible to discover the inverse algorithm.

The last property is the curious one that gives the function its name. It is like a trapdoor: easy to drop through but hard to get up through. Indeed, it is impossible to get up through the door unless one knows where the secret button is hidden. The button symbolizes the "trapdoor information." Without it one cannot open the door from below, but the button is so carefully concealed that the probability of finding it is practically zero.



Randomizer for encoding a "one-time pad"

Before giving a specific example, let us see how such functions make the new cryptographic systems possible. Suppose there is a group of businessmen who want to communicate secrets to one another. Each devises his own trapdoor function with its forward and backward algorithms. A handbook is published in which each company's encoding (forward) algorithm is given in full. The decoding (inverse) algorithms are kept secret. The handbook is public. Anyone can consult it and use it for sending a secret message to any listed company.

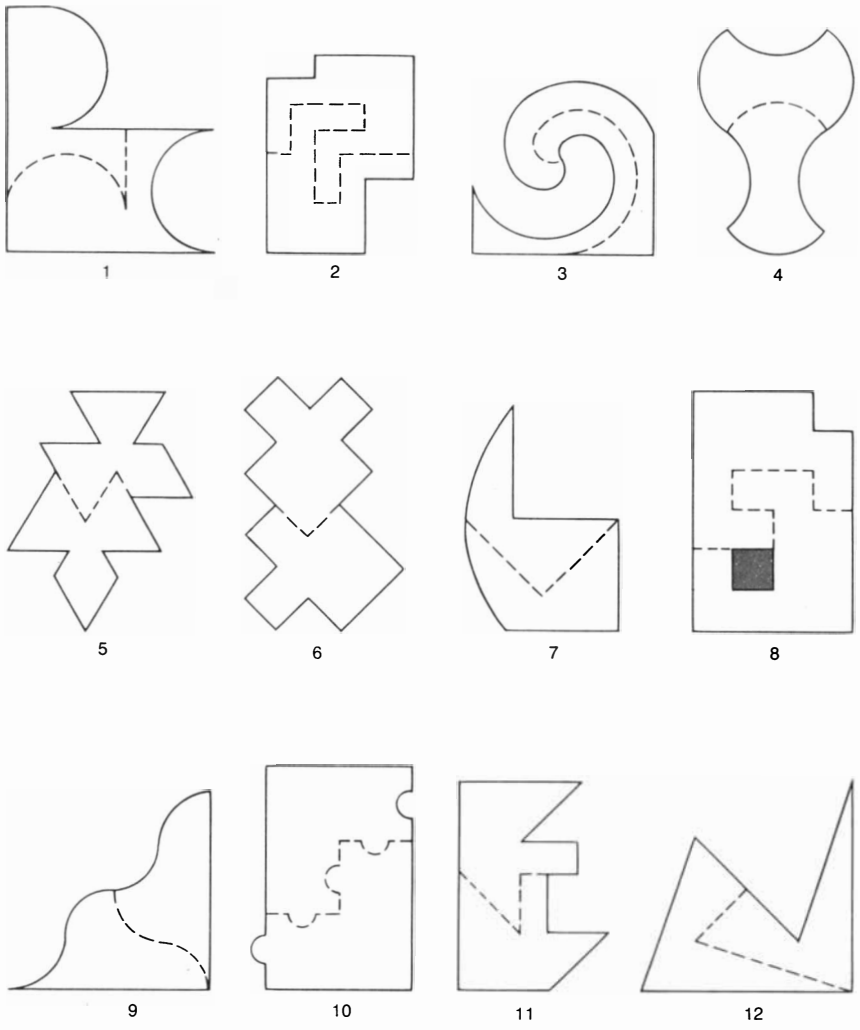
Suppose you are not a member of the group but you want to send a secret message to member Z. First you change your plaintext to a long number, using a standard procedure given in the handbook. Next you look up Z's forward algorithm and your computer uses it for rapid encoding of the ciphertext. This new number is sent to Z. It does not matter at all if the ciphertext is overheard or intercepted because only Z knows his secret decoding procedure. There is no way a curious cryptanalyst, studying Z's public encoding algorithm, can discover Z's decoding algorithm. In principle he might find it, but in practice that would require a supercomputer and a few million years of running time.

An outsider cannot "sign" a message to Z, but any member of the group can. Here is the devilishly clever way the signature works. Suppose A wants to sign a message to Z. He first encodes the plaintext number by using his own secret inverse algorithm. Then he encodes the ciphertext number a second time, using Z's public algorithm. After Z receives the ciphertext he first transforms it by applying his own secret decoding algorithm, then he applies A's public encoding algorithm. Out comes the message!

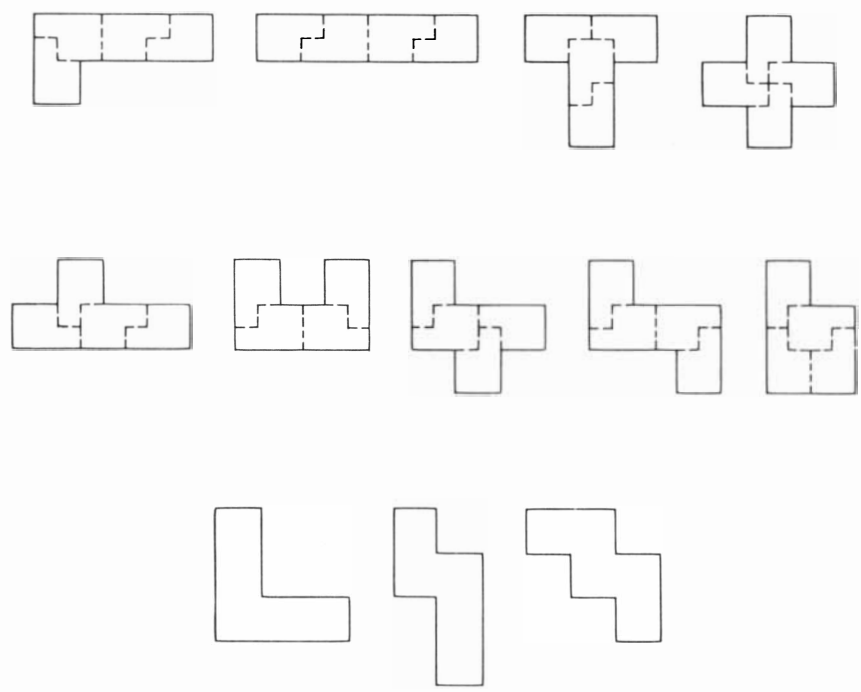
Z knows that only A could have sent this doubly encoded ciphertext because it made use of A's secret algorithm. A's "signature" is clearly unforgeable. Z cannot use it to send a message purporting to come from A because Z still does

9686	9613	7546	2206
1477	1409	2225	4355
8829	0575	9991	1245
7431	9874	6951	2093
0816	2982	2514	5708
3569	3147	6622	8839
8962	8013	3919	9055
1829	9451	5781	5154

A ciphertext challenge worth $100



The answers to last month's bisection problems



Dividing polyominoes into four congruent parts

not know *A*'s secret decoding algorithm. Not only that, but if it were to become necessary at some future time to prove to a third party, say a judge in a court of law, that *A* did in fact send the message, this can be done in a way that neither *A*, *Z* nor anyone else can dispute.

Diffie and Hellman suggested in their paper a variety of trapdoor functions that might be used for such systems. None is quite what is desired, but early this year there was a second breakthrough. Ronald L. Rivest, Adi Shamir and Leonard Adleman, computer scientists at the Massachusetts Institute of Technology, developed an elegant way to implement the Diffie-Hellman system by using prime numbers.

Rivest obtained his doctorate in computer science from Stanford University in 1973 and is now an associate professor at M.I.T. Once he had hit on the brilliant idea of using primes for a public cipher system, he and his two collaborators had little difficulty finding a simple way to do it. Their work, supported by grants from the NSF and the Office of Naval Research, appears in *On Digital Signatures and Public-Key Cryptosystems* (Technical Memo 82, April, 1977), issued by the Laboratory for Computer Science, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Mass. 02139. The memorandum is free to anyone who writes Rivest at the above address enclosing a self-addressed, 9-by-12-inch clasp envelope with 35 cents in postage.

To explain Rivest's system we need a bit of background in prime-number theory. The fastest-known computer programs for deciding whether a number is prime or composite (the product of primes) are based on a famous theory of Fermat's stating that if p is prime, and a is any positive number less than p, then $a^{p-1} = 1$ (modulo p). Suppose we want to test a large odd number n (all primes except 2 are of course odd) for primality. A number a is selected at random and raised to the power of $n - 1$, then divided by n. If the remainder is not 1, n cannot be prime. For example, $2^{21-1} = 4$ (modulo 21), therefore 21 is composite. What, however, is the connection between 2 (the randomly chosen a) and 3 and 7, the two prime factors of 21? There seems to be no connection whatever. For this reason Fermat's test is useless in finding prime factors. It does, however, provide a fast way of proving that a number is composite. Moreover, if an odd number passes the Fermat test with a certain number of random a's, it is almost certainly prime.

This is not the place to go into more details about computer algorithms for testing primality, which are extremely fast, or algorithms for factoring composites, all of which are infuriatingly slow. I content myself with the following facts, provided by Rivest. They dramatize the staggering gap in the re-

quired computer time between the two kinds of testing. For example, to test a 130-digit odd number for primality requires at the most (that is, when the number actually is prime) about seven minutes on a PDP-10 computer. The same algorithm takes only 45 seconds to find the first prime after 2^{200}. (It is a 61-digit number equal to $2^{200} + 235$.)

Contrast this with the difficulty of finding the two prime factors of a 125- or 126-digit number obtained by multiplying two 63-digit primes. If the best algorithm known and the fastest of today's computers were used, Rivest estimates that the running time required would be about 40 quadrillion years! (For a good discussion of computer methods of factoring into primes, see Donald E. Knuth's *Seminumerical Algorithms,* Section 4.5.4.) It is this practical impossibility, in any foreseeable future, of factoring the product of two large primes that makes the M.I.T. public-key cipher system possible.

To explain how the system works, the M.I.T. authors take as an example of plaintext a paraphrase of a remark in Shakespeare's *Julius Caesar* (Act 1, Scene 2): ITS ALL GREEK TO ME.

This is first changed to a single number, using the standard key: A = 01, B = 02, . . . , Z = 26, with 00 indicating a space between words. The number is 09201900011212000718050511002015001305.

The entire number is now encoded by raising it to a fixed power *s,* modulo a certain composite number *r.* The composite *r* is obtained by randomly selecting (using a procedure given in the M.I.T. memorandum) two primes, *p* and *q,* each of which is at least 40 digits long, and multiplying them together. The number *s* must be relatively prime to $p - 1$ and $q - 1$. Numbers *s* and *r* are made public, to be used in the encoding algorithm. The encoding operation can be done very efficiently even for enormous values of *r;* indeed, it requires less than a second of computer time.

The two prime factors of *r* are withheld, to play a role in the secret inverse algorithm. This inverse algorithm, used for decoding, consists in raising the ciphertext number to another power *t,* then reducing it to modulo *r.* As before, this takes less than a second of computer time. The number *t,* however, can be calculated only by someone who knows *p* and *q,* the two primes that are kept secret.

If the message is too long to be handled as a single number, it can be broken up into two or more blocks and each block can be treated as a separate number. I shall not go into more details. They are a bit technical but are clearly explained in the M.I.T. memo.

To encode ITS ALL GREEK TO ME, the M.I.T. group has chosen $s = 9{,}007$ and $r = 11438162575788886766923577997614661201021829672124236256256184293570693524573389783059712356395870505898907514759929002687954354 1.$

The number *r* is the product of a 64-digit prime *p* and a 65-digit prime *q,* each randomly selected. The encoding algorithm changes the plaintext number (09201 . . .) to the following ciphertext number: 19993513149780510045231712274026064742320400170583914631037037174062597160894892750430992096267258267501289355446135382376974802 6.

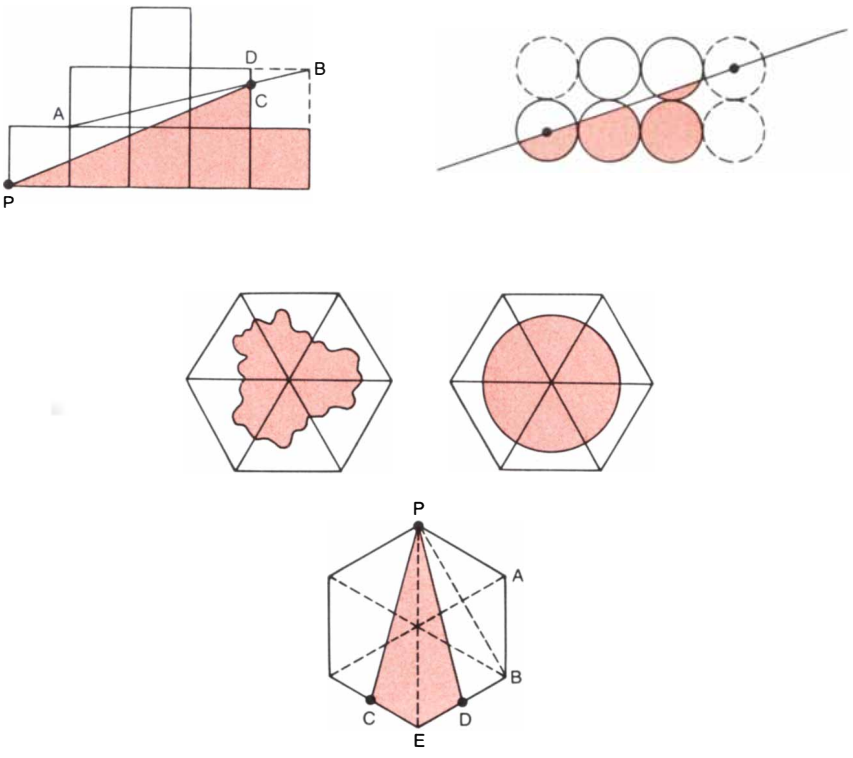
As a challenge to *Scientific American* readers the M.I.T. group has encoded another message, using the same public algorithm. The ciphertext is shown in the bottom illustration on page 121. Its plaintext is an English sentence. It was first changed to a number by the standard method explained above, then the entire number was raised to the 9,007th power (modulo *r*) by the shortcut method given in the memorandum. To the first person who decodes this message the M.I.T. group will give $100.

To prove that the offer actually comes from the M.I.T. group, the following signature has been added: 1671786115038084424601527138916839824543690103235831121783503844692906265544879223711449050957860865566249657797484000405702037 3.

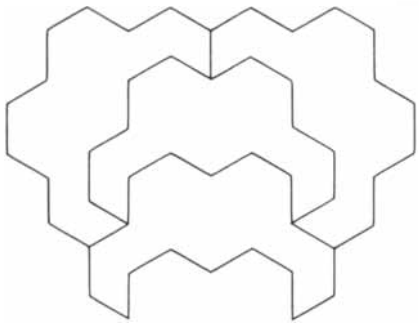
The signature was encoded by using the secret inverse of the encoding algorithm. Since the reader has no public encoding algorithm of his own, the second encoding operation has been omitted. Any reader who has access to a computer and the instructions in the M.I.T. memorandum can easily read the signature by applying the M.I.T. group's public encoding algorithm, that is, by raising the above number to the power of 9,007, then reducing it to modulo *r.* The result is 0609181920001915122205180023091419001514050008211404180504000415121201181 9. It translates (by the use of the standard key) to FIRST SOLVER WINS ONE HUNDRED DOLLARS. This signed ciphertext could only come from the M.I.T. group because only its members know the inverse algorithm by which it was produced.

Rivest and his associates have no proof that at some future time no one will discover a fast algorithm for factoring composites as large as the *r* they used or will break their cipher by some other scheme they have not thought of. They consider both possibilities extremely remote. Of course any cipher system that cannot be proved unbreakable in the absolute sense of one-time pads is open to sophisticated attacks by modern cryptanalysts who are trained mathematicians with powerful computers at their elbow. If the M.I.T. cipher withstands such attacks, as it seems almost certain it will, Poe's dictum will be hard to defend in any form.

Even in the unlikely event that the M.I.T. system is breakable there are probably all kinds of other trapdoor functions that can provide virtually unbreakable ciphers. Diffie and Hellman are applying for patents on cipher devices based on trapdoor functions they have not yet disclosed. Computers and complexity theory are pushing cryptog-



Solutions to four equal-division problems



A bilaterally symmetric tetrad with 18 sides

raphy into an exciting phase, and one that may be tinged with sadness. All over the world there are clever men and women, some of them geniuses, who have devoted their lives to the mastery of modern cryptanalysis. Since World War II even those government and military ciphers that are not one-time pads have become so difficult to break that the talents of these experts have gradually become less useful. Now these people are standing on trapdoors that are about to spring open and drop them completely from sight.

The top illustration on page 122 shows how the 12 shapes given last month can be divided into congruent halves. The bottom illustration on the same page shows how nine of the 12 order-5 polyominoes can be dissected into the same four congruent parts. The three blank polyominoes cannot be cut into four congruent parts of any shape.

The illustration on the preceding page answers the four problems at the end of last month's column. To bisect the nine squares draw the 10th square shown with broken lines. Rule *AB* to get point *C*, then join *P* to *C*. If the squares have sides of length 1, then *CD* equals 1/4, and it is easy to see that *PC* bisects the original figure. To bisect the five circles add three additional circles as shown by the broken lines. The line through the centers of two circles obviously halves the total area. (Both problems are from *A Problem a Day*, by R. M. Lucey, Penguin Books, 1937.)

The hexagon at the bottom is trisected by joining *P* to *C* and *D*, the midpoints of two sides. Assume that the equilateral triangles have areas of 1. The area of *PAB* is 1, therefore the area of *PBE* is 2 and the rest follows. I was unable to find any comparably simple way to trisect a regular pentagon with a line through a corner.

The middle two hexagons show how Leo Moser proved that the minimum-length curve bisecting an equilateral triangle is the arc of a circle. Whatever the shape of the bisecting curve, it will form a closed curve if the triangle is reflected around one vertex as is shown. Such a curve cuts the hexagon in half, and it has a fixed area. The figure of minimum perimeter that encloses a given area is the circle, therefore the minimum-length bisecting curves inside each triangle are arcs of a circle. (This exercise is from *Mathematical Quickies*, by Charles W. Trigg, McGraw-Hill, 1967.)

Comments on the mail response to April's short problems follow:

The generalization of the pool-ball problem to triangles of order *n*, bearing consecutive numbers starting with 1, has been solved. Herbert Taylor found an ingenious way to prove that no TAD (triangle of absolute differences) could be made with triangular arrays of order 9 or higher. Computer programs eliminated TAD's of orders 6, 7 and 8, therefore the unique solution for the 15 pool balls is the largest TAD of this type.

Solomon W. Golomb proposed three candidates for further investigation:

1. If all numbers in a TAD of order greater than 5 are distinct but not consecutive, how big is the largest number forced to be? (Example: An order-6 TAD is possible with the largest number as low as 22.)

2. Using all numbers from 1 to *k*, but allowing repeats, how big can *k* be in a TAD of order *n*? (Example: An order-6 TAD is possible with *k* as high as 20.)

3. For what orders is it possible to form a TAD modulo *m*, where *m* is the number of elements in the triangle and the numbers are consecutive from 1 to *m*? Each difference is expressed modulo *m*. Such triangles can be rotated so that every element below the top row is the sum (modulo *m*) of the two numbers above it. Here, in rotated form, are the four order-4 solutions:

1 6 9 4	2 7 8 3	6 1 4 9	7 2 3 8
7 5 3	9 5 1	7 5 3	9 5 1
2 8	4 6	2 8	4 6
0	0	0	0

A backtrack program by Golomb and Taylor found no solution for order 5. Col. George Sicherman, who invented the original pool-ball problem, reports a computer proof of impossibility for order 6. Higher orders remain open.

Robert Ammann, Greg Frederickson and Jean L. Loyer each found an 18-sided polygonal tetrad with bilateral symmetry [*see illustration on this page*], thus improving on the 22-sided solution I had published.

Dan Eilers, Allen I. Janis, Scott Kim, P. H. Lyons, Robert Mathews (with Martin G. Wallser), James Newton and Mike Tempest each found a second solution (there are no more) for the lost-king tour on the order-5 square.

When I ended the column with limericks of decreasing length, I referred to the one-line limerick as the "last of four." Draper L. Kauffman, John Little, John McKay, Thomas D. Nehrer and James C. Vibber were the first of many who told me I should have called it the last-but-one of five. The fifth, of course, has no lines, which is why other readers failed to notice it.

Tom Wright of Ganges, British Columbia, wrote: "I was interested in the limerick paradox, particularly in the decreasing two-line and one-line limericks. I wondered if you had, in fact, added the no-line limerick (about the man from Nepal), and I looked minutely to see if it wasn't there. On examination, my first impulse was to assume that it was indeed not there, since no space was provided, but further cogitation suggested that a no-line poem, requiring no space, might indeed be there. Unable to resolve this paradox by any logical proof, I am abjectly reduced to asking you whether or not a no-line limerick was not printed in the space not provided, or not."

A Cipher that Defeated Poe

"Ge Jeasgdxv,

Zij gl mw, laam, xzy zmlwhfzek ejlvdxw kwke tx lbr atgh lbmx aanu bai Vsmukkss pwn vlwk agh gnumk wdlnzweg jnbxvv oaeg enwb zwmgy mo mlw wnbx mw al pnfdcfpkh wzkex hssf xkiyahul. Mk num yexdm wbxy sbc hv wyx Phwkgnamcuk?"

In 1839, in a regular column Edgar Allan Poe contributed to a Philadelphia periodical, *Alexander's Weekly Messenger*, Poe challenged readers to send him cryptograms (monoalphabetic substitution ciphers), asserting that he would solve them all "forthwith." One G. W. Kulp submitted a ciphertext in longhand. It was printed as shown above in the issue of February 26, 1840. Poe "proved" in a subsequent column that the cipher was a hoax—"a jargon of random characters having no meaning whatsoever."

In 1975 Brian J. Winkel, a mathematician at Albion College, and Mark Lyster, a chemistry major in Winkel's cryptology class, cracked Kulp's cipher. It is not a simple substitution—Poe was right—but neither is it nonsense. Poe can hardly be blamed for his opinion. In addition to a major error by Kulp there are 15 minor errors, probably printer's mistakes in reading the longhand.

Winkel is an editor of a new quarterly, *Cryptologia*, available from Albion College, Albion, Mich. 49224, at $16 per year. The magazine stresses the mathematical and computational aspects of cryptology. The first issue (January, 1977) tells the story of Kulp's cipher and gives it as a challenge to readers. So far only three readers have broken it. I shall give the solution next month.

What's new in cosmology... computers...physics...chemistry... applied analytical math...geology... and the entire universe around us!

To keep in touch with everything that's happening in every scientific area that interests you, get set now for a great reading spree –

take any 3 books all for only $3.95

as a member of The

Library of Science

if you will join now for a trial period and accept only 3 more books–at member discount prices–within the next 12 months

42205. A DICTIONARY OF EARTH SCIENCES. *Edited by Stella Stiegler.* Covers geophysics, structural geology, geomorphology, oceanography, stratigraphy, paleontology, sedimentology, climatology, meteorology, cartography, mineralogy, petrology, etc. **$16.50**

00470. VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA. Fifth Edition. Hailed as the most comprehensive one-volume reference work of its kind; thoroughly revised and expanded. Illustrated with 2,450 photographs, diagrams, graphs, and over 550 tables. *Counts as 3 of your 3 books.* **$ 67.50**

66570. AN OUTLINE OF STRUCTURAL GEOLOGY. *Bruce E. Hobbs, Winthrop D. Means and Paul F. Williams.* Up-to-date account of every aspect of structural geology—from the nature of stress and strain to the latest information on plate tectonics. **$16.95**

55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. *Ian Ridpath.* Covers the full range of astronomical knowledge and space exploration, from ancient times to now. **$ 16.95**

70170. PRINCIPLES OF HOLOGRAPHY. Second Edition. *Howard M. Smith.* From fundamentals to rigorous theory, from chemical formulas to applications. **$ 16.50**

44900. THE ENCYCLOPEDIA OF COMPUTER SCIENCE. *Edited by Anthony Ralston and Chester L. Meek.* A monumental 1,550 page volume, offering condensed, accurate information on essential topics ranging from theory to applications, from program structures to the logical design of central processing units, from Boolean algebra to microprogramming. Bibliographies, cross references, complete indexing and 700 charts and tables. *Counts as 3 of your 3 books.* **$ 60.00**

70190. PRINCIPLES OF OPERATIONS RESEARCH. Second Edition. *Harvey M. Wagner.* This 1039-page book offers a complete overview of all the fundamental concepts. **$21.50**

49340. FUNDAMENTAL ASTRONOMY: SOLAR SYSTEM AND BEYOND. *Franklyn W. Cole.* Traces our assault on the mysteries of the universe from the earliest sensory-oriented glimmerings to modern astronomical thought. **$ 16.50**

61530. MATHEMATICS DICTIONARY. Fourth Edition. *Edited by Robert C. James and Edwin F. Beckenbach* Revised and expanded, offers clear, accurate definitions of more than 8,000 terms. **$17.95**

values to $67.50

77500. THE SEARCH FOR THE GOLD OF TUTANKHAMEN. *Arnold C. Brackman.* The first complete story of the search, discovery and aftermath of one of the great archaeological finds of all time. **$8.95**

82660. TECHNOLOGY MATHEMATICS HANDBOOK. *Jan J. Tuma.* Instant access to concise summaries of all major definitions, formulas, graphs, tables of elementary and intermediate mathematics — from algebra to determinants and vectors. **$15.95**

50380. GLOSSARY OF CHEMICAL TERMS. *Clifford A. Hampel and Gessner G. Hawley.* Over 2,000 concise definitions: all major chemical groups, important functional terms, basic phenomena and processes, all the chemical elements, and more. **$14.95**

49267. FRONTIERS OF ASTROPHYSICS. *Edited by Eugene Avrett.* A contemporary source on astrophysical topics from recent solar research, to neutron stars, supernovae and intergalactic matter. **$ 20.00**

34140. APPLIED ANALYTICAL MATHEMATICS FOR PHYSICAL SCIENTISTS. *James T. Cushing.* Deals with topics ranging from spectral analysis of linear operations to complex variables, to group theory. **$19.95**

79140. SOCIOBIOLOGY AND BEHAVIOR. *David T. Barash.* A brilliant introduction to the emerging science—sociobiology—a new, highly controversial way of looking at behavior. **$ 9.95**

36512. THE BOG PEOPLE/THE MOUND PEOPLE. *P. V. Glob.* Perfectly preserved 2,000-year-old bodies buried in peat bogs, all murdered...6-ft.-long oak coffins hidden within mounds of earth...two ancient mysteries solved by an eminent archaeologist. Richly illustrated. *The set counts as two of your 3 books.* **$25.00**

36785. THE BOWELS OF THE EARTH. *John Elder.* A fascinating, systematic text on the dynamics of the earth's interior, which joins together crustal geology and geophysics. **$ 14.00**

83760. THEORY OF COMPUTATION. *Walter S. Brainerd and Lawrence H. Landweber.* Fascinating, totally new presentation of the theory of computation. **$ 19.95**

61310. MATHEMATICAL IDEAS, MODELING AND APPLICATIONS. *Z. A. Melzak.* Probes underlying principles of problems in geometry, analysis, combinatorics, number theory, algebra, computing and much more. Companion to the immensely popular Concrete Mathematics, Vol. 1. *Counts as 2 of your 3 books.* **$ 29.95**

39745. COMPANION TO CONCRETE MATHEMATICS. *Z. A. Melzak.* "A delightful collection of problems organized into over sixty sections, together with historical notes, solutions, observations, and remarks." —*American Mathematical Monthly.* **$16.50**

34670. ASTRONOMY AND COSMOLOGY. *Sir Fred Hoyle.* One of the giants of modern astronomy clearly explains today's astronomy, and the physical processes and relationships underlying basic astronomical phenomena. Over 600 illustrations. **$16.50**

64145. A NEW DICTIONARY OF PHYSICS. *H. J. Gray and Alan Isaacs.* More than 9,000 entries provide a comprehensive focus on key concepts, formulas, men and machines that are crucial to understanding modern physics. Hundreds of illustrations, graphs, charts and tables. *Counts as 2 of your 3 books.* **$ 35.00**

42235. A DICTIONARY OF PHYSICAL SCIENCES. *Edited by John Daintith.* Physics, chemistry, astronomy, electronics, computer sciences and mathematics are covered by definitions, formulas and diagrams, 3,000 entries. **$ 16.50**

59800. MAN DISCOVERS THE GALAXIES. *Richard E. Berendzen, Richard C. Hart and Daniel Seeley.* The history-making discoveries by the band of ten great astronomers who led the way between 1915 and 1940, are recaptured through original letters, photographs and archival treasures. Profusely illustrated. **$ 15.95**

42210. A DICTIONARY OF LIFE SCIENCES. *Edited by Elizabeth Martin.* From *Abscisic acid* to *Zymogen,* this work covers taxonomy, zoology, botany, anatomy, physiology, cytology, genetics and ecology with 3,300 entries ranging from short 20-word entries to lengthy detailed articles. **$ 16.50**

If the reply card has been removed, please write to
Library of Science
Dept. 2-7AK, Riverside, N.J. 08075
to obtain membership information and an application.

THE AMATEUR SCIENTIST

Drops of water dance on a hot skillet and the experimenter walks on hot coals

by Jearl Walker

In his best-selling novel *Something of Value* Robert Ruark described a test of truth in an African village. Two men, one who was telling the truth and one who was lying, had "undergone the ordeal of the licking of the hot knife, so that he whose guilt was most apparent would scorch his tongue when the saliva dried up from the knowledge of fear." My grandmother does something similar when she tests a hot iron by touching it briefly with a wet finger. Touching hot metal was developed into a stunt in turn-of-the-century carnival sideshows where daredevils would briefly plunge their wet fingers into some molten metal, probably lead. Some candymakers similarly plunge wet fingers into hot, melted candy in order to test the temperature.

These examples of touching hot material with wet skin, which suffers no harm, are all related to the same physics. Jules Verne cleverly employed the physics in one of his novels, *Michael Strogoff*. Sent by the Czar into Siberia with a secret message for the Czar's brother, Strogoff was captured by the invading Tartars, identified as a courier for the Czar and sentenced to be blinded. The Tartar method of blinding was to pass a red-hot blade just in front of the victim's eyes; the intense heat radiated by the blade would damage the eyes beyond repair. Just as Strogoff was to be blinded, however, his mother fell before him, bringing tears to his eyes, since it would be his last glimpse of her. The watery layer of tears protected his eyes from the heat radiated by the blade, and he ultimately completed his journey.

The most remarkable exhibition of this protection afforded by a thin watery layer appears when people walk barefooted over hot coals or solidified hot lava. Probably you think there is some trick involved, the walker having surreptitiously applied a protective coating to his feet before the walk. As we shall see in the following experiments, however, no special coating is needed and no trick is involved, although perhaps a certain willingness to tolerate pain is needed.

A phenomenon of this kind is familiar to anyone who is skilled at making pancakes. Such a person will have the griddle very hot so that the batter solidifies quickly after being poured. Not having a thermostat on the griddle, one tests the griddle's temperature by sprinkling water on it. If the griddle is hot but not hot enough, the water drops spread out, wet the surface and evaporate within about two seconds. If the griddle is ready for the batter, the sprinkled drops dance, vibrate and skim over the griddle surface for from 30 to 100 seconds. This result seems all wrong. How can drops last longer on a hotter griddle?

The persistence of water drops on hot surfaces was first described by a German physician, Johann Gottlieb Leidenfrost, in "A Tract about Some Qualities of Common Water." Although the paper appeared in 1756, it was translated from the Latin only in 1965 and therefore has not been widely read. Leidenfrost placed small water drops in a red-hot iron spoon taken from the fireplace and noted how long they lasted by counting the swings of a pendulum. Among the several things he noticed about the suspended drops was that they appeared to suck light and fire from the spoon because under the drops the spoon immediately turned black. With the spoon initially red-hot, his first drop lasted for about 30 seconds. A second drop then placed on the spoon lasted for only 10 seconds. Further drops on the spoon lasted for only one or two seconds. Leidenfrost did not regard the longer-lasting drops as boiling, but modern science classifies them as a type of boiling different from the more common type (called nucleate boiling) that the quickly vanishing drops underwent.

Leidenfrost was unable to predict how long the drops would last for two main reasons. First, he was unable to measure the high temperatures of his spoon. (Indeed, he suggested using the length of time that drops persisted as a calibration of the higher temperatures.) Second, he could not calculate theoretically the heat supplied to a drop from the spoon because the concept of the latent heat of vaporization (the heat needed to transform liquid to vapor) did not exist at that time. (It was introduced 14 years later.) Leidenfrost apparently was also still captive to one of the Aristotelian errors, namely that fire and water may produce a solid. "These observations," he wrote, "[suggest] that water is changed into earth by a large fire, because always after the complete evaporation of the drop some terrestrial matter remains in the heated vessel." Finally, Leidenfrost thought the hotter spoon sustained drops longer because the heat increased the surface tension of the water, which is not true.

A Leidenfrost drop is supported by a layer of its own vapor approximately .09 millimeter thick if the metal plate on which it is suspended is sufficiently hot. On a metal plate at a relatively low temperature, although the plate is still hot, the drops fall on it, wet it and evaporate within several seconds. On a hotter plate the bottom of a falling drop vaporizes almost immediately as it nears the plate, leaving a layer of vapor to support the remaining portion of the drop. The drop dances and skims about in reasonable safety from the hot plate because of this protective vapor layer, which is constantly supplied by evaporation of the bottom of the drop. Eventually enough heat is conducted and radiated through the vapor layer to vaporize the remaining drop, and then the dancing is over. The primary mechanism of heat transfer to the drop appears to be conduction through the vapor layer, although at higher temperatures radiation is progressively more important. Water vapor is a relatively poor conductor of heat, typically an order of magnitude poorer than liquid water and several orders of magnitude poorer than a solid such as copper. The drops therefore have a surprisingly long duration.

A plot of the drop durations v. the temperature of the metal plate shows a fairly sharp transition between the drops that last a short time (not becoming suspended) and the drops that last a long time (with suspension). The temperature at which the drops last longest is the Leidenfrost point. The actual value of this temperature depends on, among other things, the saturation temperature of the fluid, that is, the temperature at which the fluid will boil in the normal way. If the saturation temperature is lowered, say by moving to a higher altitude, the Leidenfrost point will be lower. With plate temperatures somewhat higher than the Leidenfrost point, the drops last longer than the one or two seconds for the nonsuspension case but not as long as those right on the Leiden-

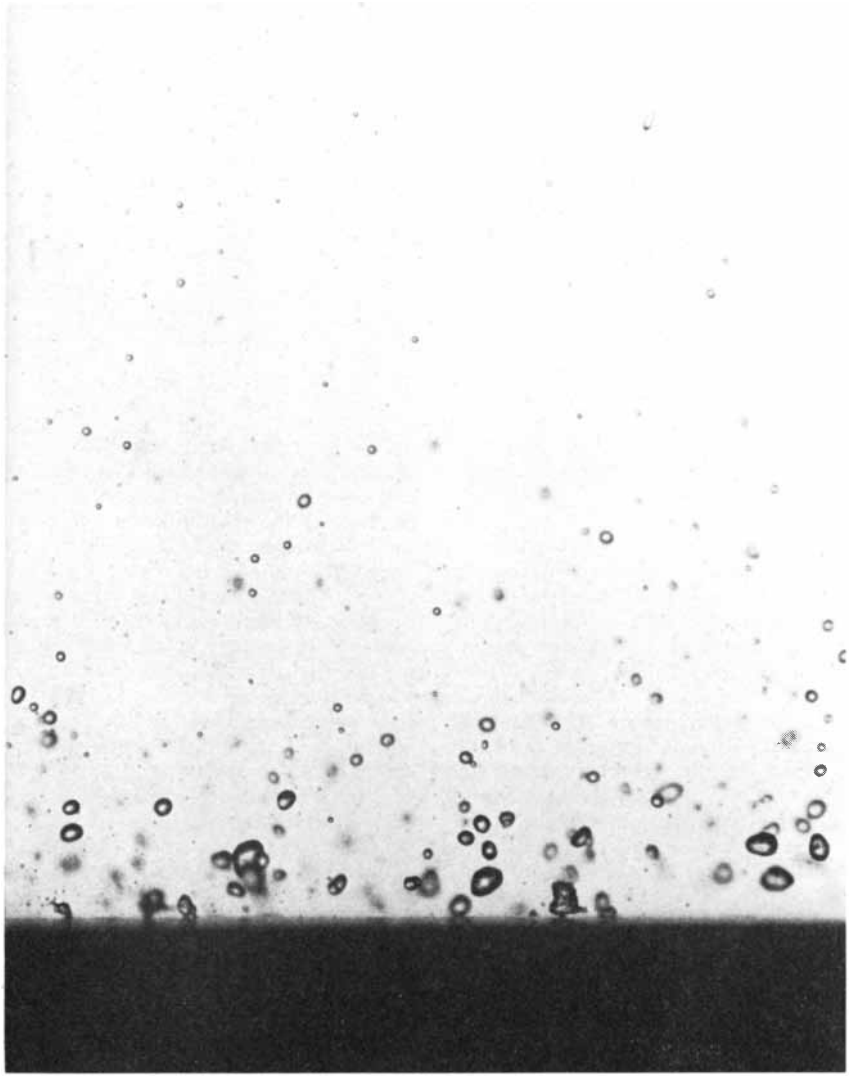
frost point. In the experiments described here one can determine the Leidenfrost point and the maximum duration for drops of water and of several other liquids. Not all the various aspects of the phenomenon have been explored, and I offer several of them that you might want to investigate on your own.

You can examine the duration of drops on a hot surface with some relatively simple arrangements in your kitchen. In obtaining my data I worked with a flat piece of 1/16-inch aluminum on a tripod straddling a Bunsen burner. You could use instead the burner on either a gas stove or an electric stove; for more control over the temperature of the surface you could plug a hot plate into a Variac so that by varying the voltage to the hot plate you can smoothly and easily vary the plate's temperature. I dented the metal plate with a ball-peen hammer so that the Leidenfrost drops would stay in one place.

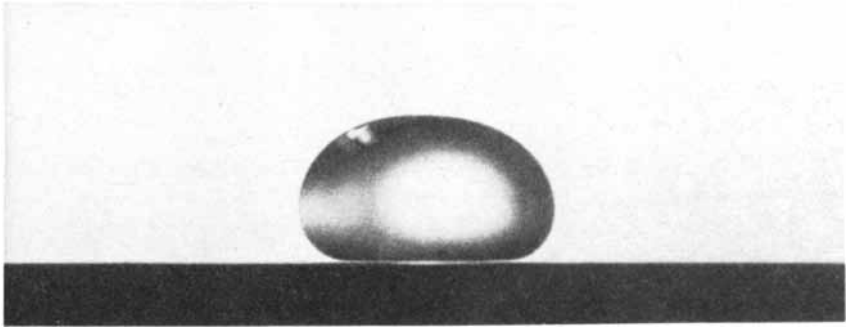
The drops are formed on the end of a hypodermic needle. A glass syringe is better, since the plastic type may react with some of the fluids. (If you have trouble getting a hypodermic needle, an eyedropper is a tolerable substitute.) Several needles of different diameter should be obtained so that drops of different initial radii can be formed. Hold the syringe just over the hot metal plate and slowly depress the plunger until a drop forms on the end of the needle and then falls of its own accord. This technique produces drops of just about the same size each time. To calculate the volume of each drop you can either count the number of drops needed to reduce the syringe contents by one or two cubic centimeters or, if the syringe does not have cubic centimeters marked off, count the number of drops needed to fill a known volume.

Each time I changed the plate temperature I allowed at least five minutes for the temperature to stabilize. The temperature can be roughly monitored up to about 430 degrees Fahrenheit by a candy thermometer. The readings are not accurate, however, because the end of the thermometer does not make good contact with the hot plate and the thermometer therefore reads low.

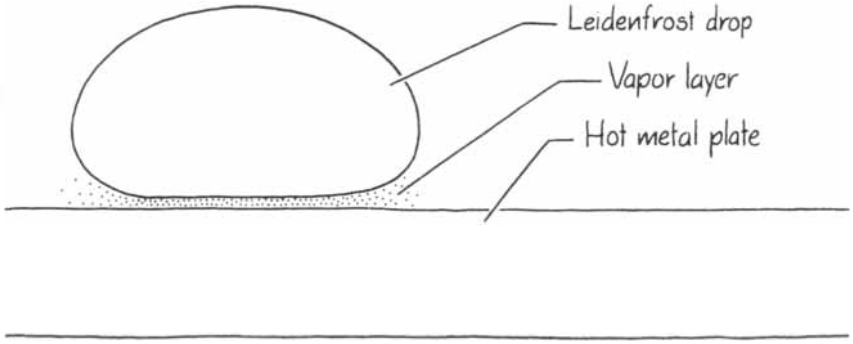
A better monitor and one that can be used at higher temperatures is a thermocouple [*see bottom illustration on next page*]. One end of the thermocouple is submerged in ice water (which, assuming it stays at a constant temperature, is at 0 degrees Celsius) and the other end is placed on the metal plate near the area in which the Leidenfrost drops are formed. Because the two wires in the thermocouple are of different metals an electric potential develops between the two ends according to their temperature difference. Having measured this voltage difference, you can convert it to a temperature difference by using a calibration table. Since one end is at 0 degrees C. the temperature difference is the temperature of the other end in degrees Celsius. I used a thermocouple with copper for one wire and the copper-nickel alloy constantan for the other and found the calibration table in the *Chemical Rubber Company Handbook*. The electric potential was measured on a potentiometer by comparing it with a reference voltage. Other types of metal can serve in the thermocouple; one is chromel-alumel wire, available from



Drops of water above a hot skillet



A single drop riding about .1 millimeter above the skillet



The vapor layer that protects the drop

the Edmund Scientific Company (555 Edscorp Building, Barrington, N.J. 08007) for about $8 and probably from distributors in your area. (Look under thermocouples in the yellow pages.) A complete package consisting of thermocouple wire and a meter can be bought from Edmund Scientific for about $50.

When I measured the temperature of molten lead, as I shall describe below, I used a resistance temperature detector from Engelhard Industries Division (700 Blair Road, Carteret, N.J. 01008). Encased in either glass or ceramic, the detector has a length of chemically pure platinum wire of temperature-dependent resistance. By measuring the resistance of the instrument on an ohmmeter (or a Multimeter) temperatures up to 750 degrees C. (corresponding to a resistance of 350 ohms) can be determined.

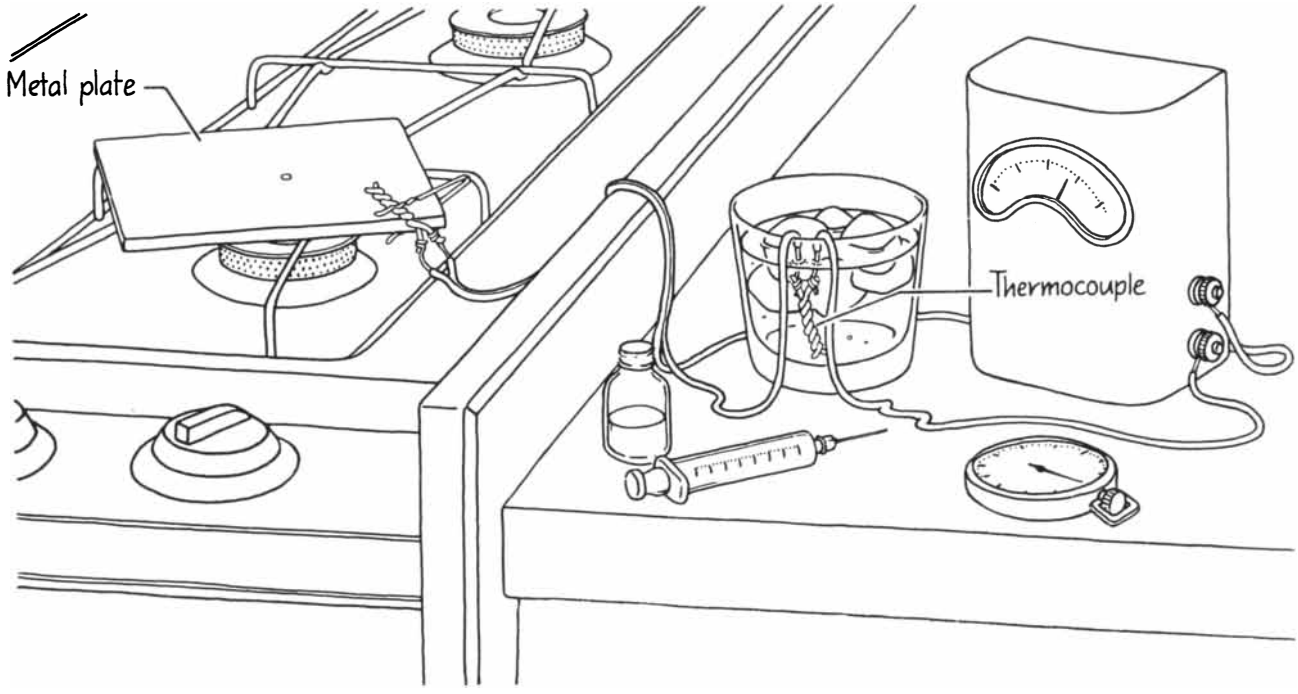
The persistence of the drops depends on the plate temperature. For distilled water there is a sharp transition from durations of 10 seconds or less to durations of 65 seconds or longer. The temperature of the maximum duration, the Leidenfrost point, was found to be between 210 and 240 degrees C., although other investigators find the maximum duration for water drops nearer 290 degrees C. under similar conditions. I do not know why my results were lower. At temperatures higher than the Leidenfrost point the duration slowly dropped off, but even for temperatures of from 400 to 500 degrees C. the drops lasted much longer than drops at temperatures below the Leidenfrost point. The Leidenfrost point that you find will depend on the temperature at which water boils where you live. At higher altitudes the boiling temperature is lower and so is the Leidenfrost point.

I found that the Leidenfrost point for smaller drops was slightly lower than the point for larger drops. More careful control over the experiment may produce the result found by others, namely that the Leidenfrost point is independent of the drop size.

Tap water produces no sharp transition at the Leidenfrost point. The difference might be that the particulate matter deposited by a fading drop of tap water bridges the vapor layer from the drop to the plate or at least prevents the necessary evaporation from the bottom of the drop. Either the bridge or the diminished vapor production would significantly shorten the duration of the drop and smooth out the Leidenfrost transition.

Taking data at or just below the Leidenfrost point is difficult because of the greater tendency for the drop to vibrate up and down at those temperatures. Invariably a vibrating drop lasts a short time because it periodically breaks through the vapor layer to touch the hot metal and is thereby warmed faster. The additional warming might decrease the lifetime of the drop by as much as 20 seconds. During the vibrations the drops jump as high as five millimeters above the hot surface.

In addition to jumping up and down the Leidenfrost drops may also oscillate radially in a variety of shapes, called normal modes, much as standing waves can be generated on a guitar string. Some of these radial oscillations are easily visible, but their visibility is increased by first dyeing the water black, dark blue or dark red and then viewing the drop under a stroboscope with flash frequencies between about four and 100 hertz. A well-behaved drop slowly shifts



Apparatus for measuring the lifetimes of drops

from one pattern (from one normal mode of radial oscillation) to another. If you have equipment for closeup photography, you can photograph the drop in one of these normal modes. The cause of the normal-mode oscillations is not understood but probably has to do with the uneven heating and uneven vaporization of the drop.

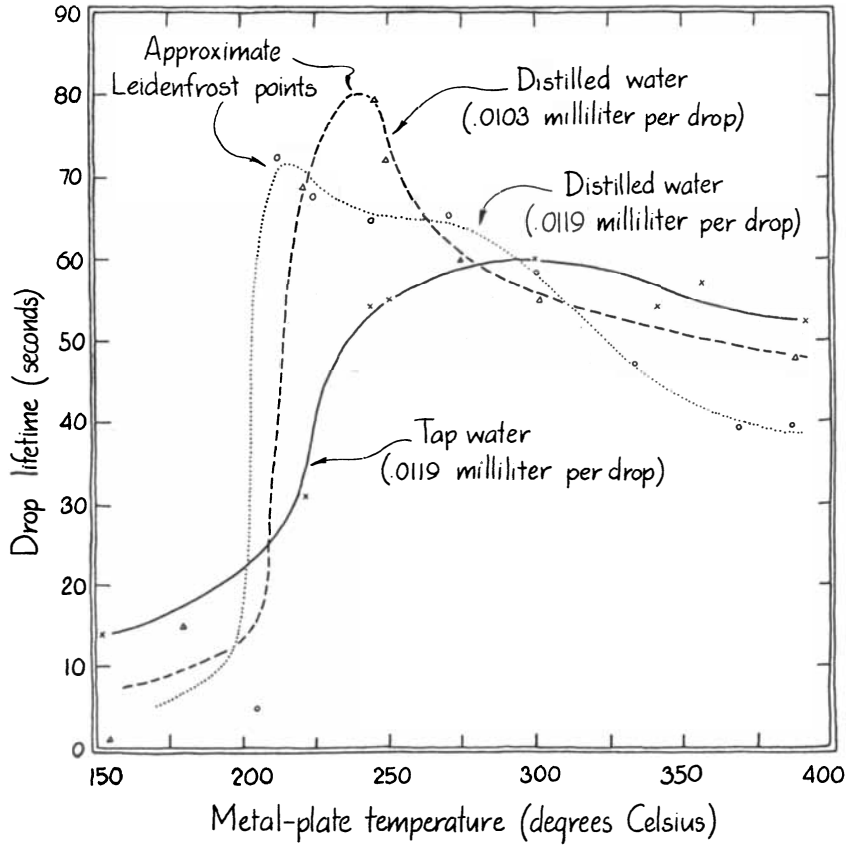
The vapor flow below a Leidenfrost drop can be demonstrated in two ways. A light powder, such as baby powder, initially sprinkled onto the metal surface will be blown by the vapor radially outward from under a drop. A similar flow can be seen on the surface of a molten-lead bath on which a Leidenfrost drop is floating, particularly if the drop is spinning.

Instead of depositing calibrated water drops on the metal plate, you can continuously feed water to the top of a Leidenfrost drop by inserting the needle into the top of the drop. Large drops can be built up, skimming and oscillating about the surface like some wild amoeba. If the drops become too large, however, they collapse under their own weight, suddenly vaporizing large portions of themselves with a sizzle.

To show that the Leidenfrost drop is really separated from the metal plate you can place a wire in the top of a large drop, attach another wire to the metal plate and connect the two leads through a battery and a small light bulb. The vapor layer below the drop is not electrically conductive, and the circuit remains open (light bulb unlighted) unless the drop happens to break through the vapor layer and momentarily touch the metal plate.

I did not investigate how the Leidenfrost point for distilled water varied with the type and nature of the metal. How the type of metal affects the Leidenfrost point is not clear in the writing on the subject. You might want to investigate that aspect yourself. The surface roughness of the metal plate is definitely a factor in the persistence of the drop; a rough surface generally has drops of shorter duration. Spraying the metal surface with a vegetable shortening such as Pam (a commercially available cooking spray used on pots and pans to decrease the sticking of food) has no apparent effect on either the Leidenfrost point or the persistence of the drops. You might try other surface coatings such as Crisco (another vegetable shortening) or a Teflon spray.

You may use many common household liquids in place of water in investigating the Leidenfrost phenomenon, but first you should eliminate any that are flammable or likely to explode near an open fire or on a hot surface. Karo corn syrup beads into Leidenfrost drops at temperatures near 300 degrees C. Instead of floating quietly until it vanishes, as a water drop does, the syrup drop puffs up into a large hollow sphere and then cooks into a mess on the metal. Save this liquid for last or you will have to clean the metal. White vinegar has a Leidenfrost point near 225 degrees C. My sample of a commercial white vinegar had been reduced with water to an acetic value of 4 percent. Nearly all the Leidenfrost vinegar drops popped out of existence at the end of their dance, perhaps because only the water was left then and they suddenly heated. I think it would be fun to try other fluids that are combinations of liquids with different boiling points. Does the Leidenfrost point become less pronounced with such combination liquids? Do the drops last longer than drops of the individual pure liquids?



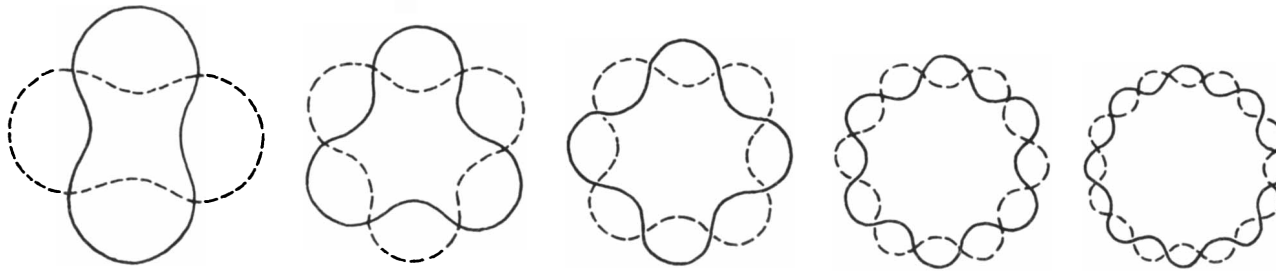
Results of measurements with water drops

Trying to float Leidenfrost drops on surfaces of a superheated fluid would also be fun because the vapor released by the superheated fluid would help to support the drop. Work has already been done on floating water drops on superheated water, the drops in such cases being called globules or boules [see "The Amateur Scientist," SCIENTIFIC AMERICAN, April, 1974]. In the Leidenfrost case you would have to superheat another fluid so that its temperature was then at or above the Leidenfrost point for the drops.

The most difficult fluid I worked with was vodka. The Leidenfrost point was fairly low, about 160 degrees C., and thus easy to reach. Although the duration of the drops was typically less than that of water drops because of the increased evaporation rate, the durations were still measurable. The trouble was that as I waited for the metal plate to stabilize in temperature between each set of drops I kept sampling the fluid.

Leidenfrost drops are not limited to fluids that must be heated above room temperature to bead up and skim about. Liquid nitrogen sprinkled on the floor behaves in the same way that drops of water do on a hot skillet. The drops of liquid nitrogen last from five to 15 seconds. Again, the surface on which the drop is suspended provides heat to vaporize the bottom of the drop, and that vapor layer then supports the remaining drop.

An inverse Leidenfrost effect occurs if you heat small bits of metal a millimeter or less in diameter to a temperature near 1,000 degrees C. in a flame and then sprinkle them on a water surface. The metal pieces float over the surface while they are supported on a vapor layer emitted by the water. You can use small metal chips from a machine shop, but the irregular shapes limit the time of the skimming because as the chips tumble the vapor layer is periodically spoiled. If



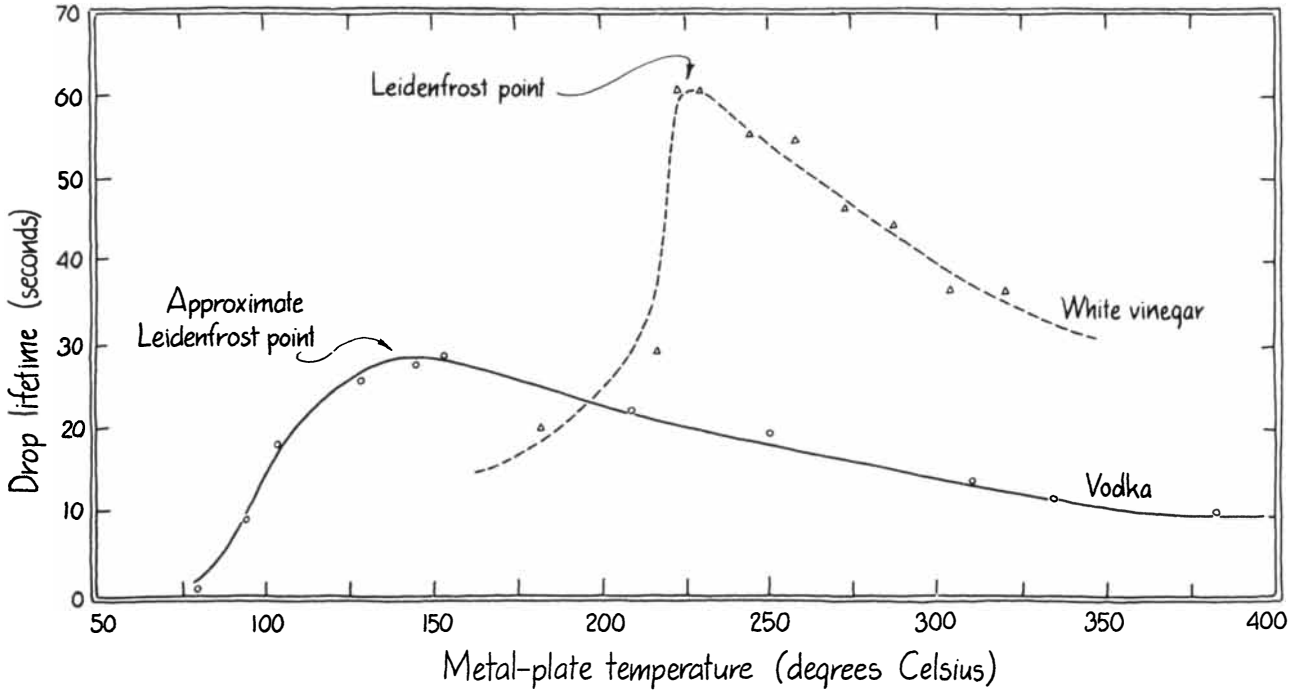
Radial oscillations of a drop as seen from above

you use small metallic spheres (I cannot find any), the pieces should last longer. Since only one brief paper has been published on the inverse Leidenfrost effect, you might want to investigate the duration of the skimmers as a function of the temperature of the main body of the water and the size of the skimmer.

Still another demonstration of the same principle, that a thin layer of water vaporizes when it is suddenly heated and then briefly provides protection against further heating, is the old sideshow stunt where the performer dips his wet fingers into a molten metal. I have done the demonstration with molten lead, dipping from one to five fingers to a depth of about two centimeters into lead at approximately 400 to 500 degrees C. (You can cook meat at 100 degrees C., and I stuck a piece of meat—my finger—into a fluid four or five times hotter.) Let me warn you, however, about this demonstration before I explain it. It can be dangerous. If the lead is near its melting point of 328 degrees C., it might suddenly solidify around the finger with disastrous results. If the lead is spilled or splashed, it can cause serious burns. The demonstration with hot lead is clearly not one to be performed by youngsters.

The protection to the finger in the molten lead comes from the water on the initially wet skin. On sudden contact with molten lead, the water on the skin vaporizes to create a brief protective sheath around the finger. One or two seconds are then needed to conduct and radiate sufficient heat through the vapor layer to the skin to bring the skin to a noticeable temperature. I find little sensation of heat in a very quick plunge. Of course I do not dillydally while my finger is in the lead. With an initially dry finger I cannot do more than quickly touch the molten lead.

I can simulate this demonstration by wrapping a thermocouple end with Teflon tape and then dipping the end into the molten lead while I monitor the temperature rise in the thermocouple. The analogy with a human finger is not too close, but the effect of water on the thermocouple can be appreciated. With the lead at 430 degrees C., the thermocouple takes about six seconds to come to the temperature of the lead if the tape is dry and eight or nine seconds if the tape is wet. This result means that in the first second there would be a difference of some 18 to 24 degrees C. in the temperature of the thermocouple in the wet and dry states. I attempted to simulate the demonstration more closely by burying a thermocouple end just under the skin of a fresh hot dog before dipping the hot dog into the molten lead. The thermocouple hardly rose in temperature, but within four or five seconds the skin of the hot dog was a black mess. Apparently the burning is confined to the first half millimeter on the surface of the hot dog (and would be for a finger if it were left in the lead that long), and the conduc-



Lifetimes of drops of white vinegar and vodka

tion of heat to the next several millimeters is relatively slow.

When a candymaker dips his wet finger into the melted candy to test the temperature, the same kind of vapor protection forms around his finger. If a finger is dipped into liquid nitrogen, the protecting vapor is created by the nitrogen bath rather than by any moisture on the skin, but the principle is the same.

Of all the demonstrations of the Leidenfrost effect, the most remarkable to me is when people walk on hot coals or lava. The *Guinness Book of World Records* describes a 25-foot walk over coals measured to be at 1,200 degrees F. Unverified stories of longer walks constantly emerge from the South Pacific and the Far East. In none of these cases is there any reason to invoke unusual powers. Although somehow shutting off pain information to the brain, having a deep-seated religious faith or just being dumb enough to try the stunt might help, the primary protection to the walker's feet comes from the natural moisture on them. Each step places parts of a foot in contact with the coals, and moisture at those places partially vaporizes to give momentary protection. Sweating between footfalls can replenish some of the moisture, but eventually most of it is depleted and the foot begins to warm up perceptibly. The walk usually ends then unless the walker has an unusual tolerance for pain. A thick layer of ash and heavily calloused feet might also lengthen the walk somewhat, but running does not help because the feet are then slammed down into the coals.

Having always been amazed by stories about people walking on hot coals, and having now become a firm believer in the Leidenfrost effect, I set up a five-foot bed of hot wood coals for such a walk. I suddenly found it is remarkably easy to believe in physics when it is on paper but remarkably hard to believe in it when the safety of one's own feet is at stake. As a matter of fact, walking on hot coals would be such a supreme test of one's true belief in what one had learned that I have suggested graduate schools might substitute it for the Ph.D. examination in physics. On one side of a pit of red-hot coals would be a line of fresh Ph.D. candidates. On the other side would be the department chairman with a handful of certificates. If a graduate student really believed in physics, he would stride across the coals without hesitation. If, however, he had any serious doubt, he would not be able to bring himself to do it.

I had to try it myself. Clutching my faded copy of Halliday and Resnick's *Physics* in one hand, I strode over the five feet of hot coals. Apparently I am a true believer in physics. I have to report, however, that my feet did get a bit hot. Oh well, I am *almost* a true believer.

BOOKS

Two marvelous views of the earth, Plato's universe and an Appalachian flood disaster

by Philip Morrison

HISTORY OF ANTARCTIC EXPLORATION AND SCIENTIFIC INVESTIGATION, ANTARCTIC MAP FOLIO SERIES, FOLIO 19. Vivian C. Bushnell, editor. American Geographical Society ($16). MISSION TO EARTH: LANDSAT VIEWS THE WORLD, by Nicholas M. Short, Paul D. Lowman, Jr., and Stanley C. Freden, NASA Goddard Space Flight Center, and William A. Finch, Jr., San Diego State University. NASA SP-360. U.S. Government Printing Office ($14). The last of a series issued with National Science Foundation backing over a decade, the portfolio of big folded plates with much text and clear maps, mainly in outline, celebrates a continent of the scientists. The biggest map, dotted with brown, locates and lists 75 Antarctic stations occupied during the years 1954 through 1975 by men and women of a dozen nations. Five sheets plot and list their scientific achievements year by year. The first nine sheets are prologue, from the earliest times. It all begins with a Roman geographer who invented the great southern continent out of sheer symmetry but placed it beyond an insurmountable barrier of heat at the Equator. It was Captain Cook who first made it clear that Antarctica was surely no rich and populous New World; he sailed all around it without one sighting. It was Yankee sealers, rapacious for fur, who probably first reached the mainland from the South Shetlands, whose seals they had extinguished in a few years of hunting. The whalers followed, once they had begun to drive the big mammals out of the seas around the other pole. By the 1890's the Norwegians had begun southern whaling in earnest, and by 1925 they came with seagoing factories and no need to have land bases and face regulation.

It remained whaler's waters; little was ashore except the succession of brave explorers. The whales are scarcer now but the bases flourish, amid aircraft, transponders, tracked weasels and deep drilling. The next prize may be krill, the whales' food, now in surplus in a world where even a shrimp paste that might glow in the dark fetches a good price. The little engravings and drawings that embellish the plates are good to see: here is Scott's doughty *Discovery,* and the *Admiralen,* first of the factory ships. Modern times are all business: we have maps and text only, no drawings or photographs. It is interesting to read of the world's longest antenna, installed in 1965 at the South Pole, but text cannot compare with the fine lithograph of Ross's two ships, the *Erebus* and the *Terror,* in the Ross Sea with Mount Erebus smoking in the background in 1842.

If the folio celebrates the old cartography and its makers, the bargain NASA volume displays the new age. Almost a foot wide, the book brings the reader about 400 plates in color, most seven inches across, with a couple of careful paragraphs explaining the main geographic features shown and their geologic context. Each photograph takes a vertical look at a piece of the earth about 115 miles on an edge (one centimeter for every 10 kilometers on the ground) seen generally in a striking coded false color, with blue printed for the green scan, green displayed for the natural red, and the infrared (haze-free) represented as red. The rain forest on the banks of the braided Congo (now the Zaïre) comes out as a clear red and the huge pit of the Bingham copper mine as a light blue; the built-up areas of Los Angeles, Tokyo, New York, London, Rio and Djakarta appear as a gray blue, their tree-green parks a bright infrared crimson. Every state of the U.S. is shown at least in part (a few states have earned mosaics displaying their entire area), constituting about 40 percent of the scenes shown.

The rest displays the world, sampled on every continent. Antarctic sea ice is shown, wonderfully fractured in a painter's image; the scenes are filled with atolls, volcanoes, lakes, scarps, dune fields, deltas and the intricacies of valley and mountain of every kind. A few images appear in a different color code or in black and white or at different scales. Roads and bridges, the dikes that hold the polders of the Zuider Zee, airport runways, snow-covered road nets, patterns of field crops and fenced grazing areas show up, in addition to the city surfaces, as the works of man. Resolution is better than 100 meters, but much of what we can see depends on contrast and other variables. We are told that segments of the Great Wall of China can be made out in the original by magnification; one can see plainly the square of old Peking.

The source of all this wonder is of course the satellite system *Landsat 1,* with one certifying image from its able successor, *Landsat 2,* launched early in 1975 and still in service. A fine index is offered, a brief account of the scanning system and a long introduction describing the uses and hopes of such a way of mapping. *Landsat 1* made more than 18,000 circuits of the globe and imaged all of the U.S. and 78 percent of the rest of the world's land areas under clear skies.

This atlas is less suited for the innocent than the earlier take from lower orbit, with long slant images and natural color. Those leaped to the eye, made ready by the experience of seeing a good many maps of large areas with their characteristic forms, usually seacoasts. Here the view is much more detailed, and it comes from a world of symbolic color. The hasty reader will feel less of the impact of the space view, but one who studies all this richness will, by the same argument, gain much more.

PLATO'S UNIVERSE, by Gregory Vlastos. University of Washington Press ($7.95). "In English *cosmos* is a linguistic orphan, a noun without a parent verb." At home in the Greek tongue the active verb *kosmeō* is much used. It describes what the commander does to range his forces for battle, what the official of the state does to maintain the law, what even a cook does to prepare an appealing dish: to set or to keep things in their proper and pleasing order. ("Cosmetic" gives a hint.) Pre-Socratic thinkers knew the word from childhood; in their flowering, say with Heraclitus a little after 500 B.C., they counted all that is made in the heavens or on the earth as one well-ordered system, *kosmos,* not even plural.

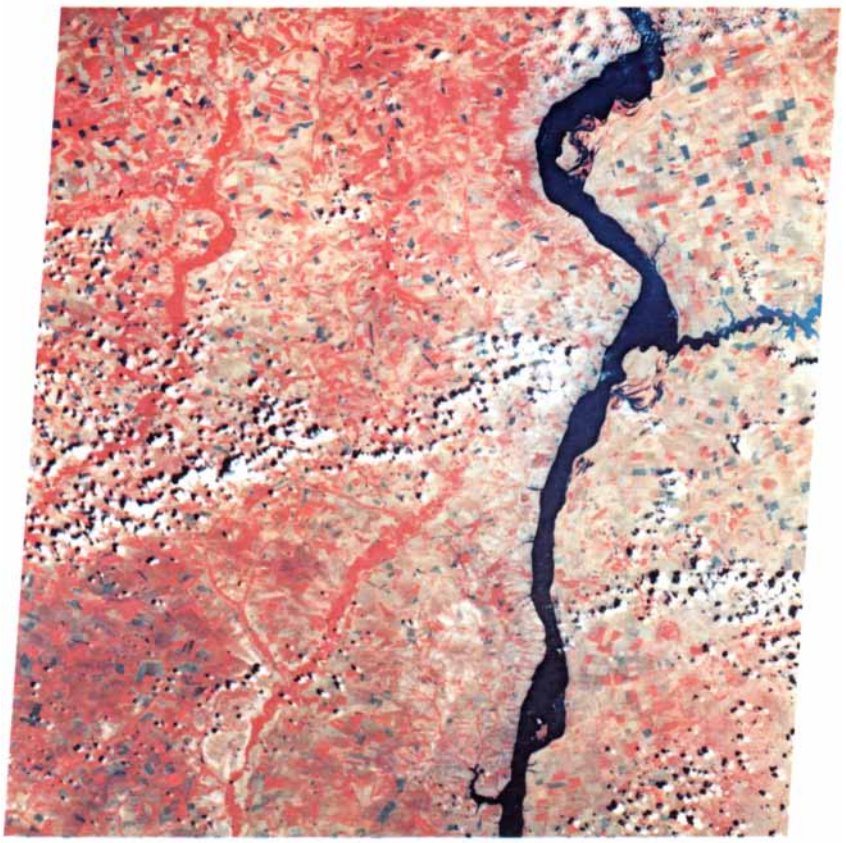
This small, elegant set of lectures by a distinguished classicist at Princeton brings the reader squarely into the domain of the scholars of texts. Meticulous translation of text, comparison of variants, allusions and implications, an effort to gain such empathy with an author in his context that we can see the options the translator faced when he chose his words—such are the delicate methods of this branch of learning, here directed at the few fragments we have from the Ionian thinkers after 600 B.C. and at the well-preserved full texts we have from their archenemy Plato, writing nearly three centuries later.

The earlier writers were the *physiologoi,* those who spoke of nature. They invented the *kosmos.* For them the nature of things, the *physis,* was not new. Greek concepts already included that

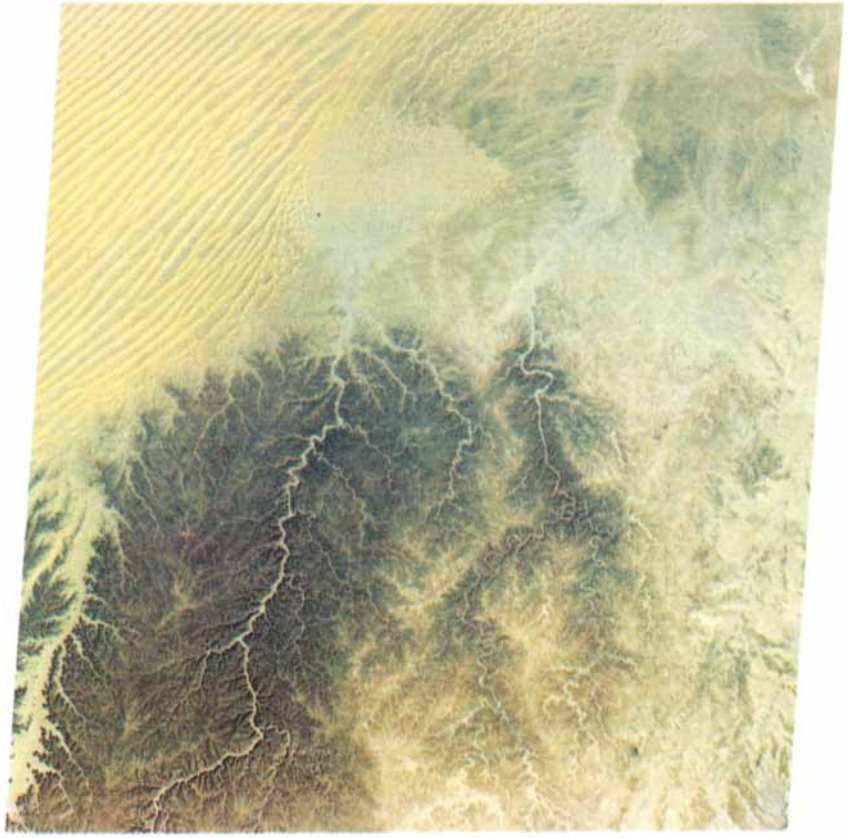
notion: the crocodile, say, had a tail according to his *physis,* his nature. Not all crocodiles have tails; they can lose them. That too is according to nature, held within certain bounds. The entire natural order, however, had its limits. The supernatural could intervene, and strange events could flow from the will of the gods.

It was the *physiologoi* who eliminated the exceptions. They achieved the "demolition of the supernatural" without a word of impiety, which might have brought them into conflict with the theocracy. They explained the heavens by a model of interactions and the events of the earth by a theory of invariants. They filled the universe with structure, so that whatever might happen—the solar eclipse no less than the bright noonday sun—would be well arranged, proper, *kosmos.* Some saw the moiety of water, earth and fire in all things, others arranged innumerable atoms in the void. For Heraclitus the sun was a bowl of fire, eclipses taking place when the bowl was turned away from earth. For Anaximander it was a great ring filled with fire, with a small orifice that enabled one to see the streaming flame from time to time. No matter. The models, grand or naïve, are many, but the framework is one. We dwell in a *kosmos* whose order we dimly perceive. The eclipses are not god-sent but follow from the well-ordered procession of natural things. "This cosmos, the same for all, no god or man has made, but it was, is, and will be for ever: ever-living fire, kindling according to measure and being extinguished according to measure," runs the fragment from Heraclitus.

Plato finds such thought worse than impious; it is dangerous. He would have his utopian state proscribe it: five years' solitary confinement and, failing recantation, even execution. "By nature and by chance, they say, fire and water and earth and air all exist—none of them exist by art . . . not by intelligence . . . nor by a god . . . but by . . . nature and by chance." The passage is awkward (Plato is writing in old age, says our guide) but perfectly clear. Yet Plato describes a cosmos too. Its structure is given in the *Timaeus,* the single dialogue that furnishes two-thirds of all the citations of Plato here. It has two parts, a theory of the heavens and a theory of matter. There was no question that the heavens were made by design. We learn of the Demiurge, their craftsman-creator. He makes the stars out of a superstuff, whose only property is motion. They are souls, part of the World Soul, a multitude of "everlasting gods." They move along the outermost of the soul circles in "the movement of the Same," the visible, perpetual, invariable circling of the fixed stars. The other bodies, the newly ordered planets, share that motion and another. They do not idly wander; rath-



Landsat 1 *satellite picture shows the lower Volga, dammed to form a long reservoir*



Landsat 1 *picture shows a part of the southern Arabian Peninsula, with dunes at upper left*

er they move by a superposition of circles, in the slanting "movement of the Different," as Vlastos translates it. The model is not simple: it embodies at least qualitatively the full material of contemporary astronomy, the fruit of observation in Greece and the Greeks' Middle Eastern sources of star lore. Plato had hit on the fundamental idea of analytic simplicity: Many simple elements (here the circles) may lie within what is perceptually not simple at all.

The supernatural is at the heart, it would appear. The world came by perfect design; the stars are gods and move by their own will. Yet they move in circles to fit the data. Why? Because their souls are superbly rational; they seek without envy only the beautiful and the good. Hence they can will only the circle. Plato has himself made a metaphysical circle. There is no more room for the whims of the gods in his mind-run cosmos than there is in the godless cosmos of his antique enemies. "A metaphysical fairy tale" led Plato to a precise and testable kinematics, a model that is opener to early scientific investigation than any vague analogies of rings and bowls and vortexes ever were.

Plato's theory of matter is treated at the same level of detail, with quite beautiful drawings (mainly from an edition of Plato prepared by P. Friedländer), of the regular solids that became Plato's atoms. These were no longer permanent and multifarious, like the atoms of Democritus, but freely changing, keeping the surface triangles and their areas perhaps as some sort of invariant. (The dodecahedron is plain left out, getting only vague reference at one point in the *Timaeus*. Four elements meant four solids; it was just too bad for the geometry.) There was no laboratory control of a theory of matter, to provide data like that for the theory of the heavens. Only traders and jewelers used the balance. The commonsense experiences of the bath or the oil lamp and its wick could be fitted easily to any kind of atomic theory in the hands of a clever theorist not constrained to make atoms directly observable.

But from Athens to Spinoza, Plato "gave rational men a pious faith to live by in two millennia all through which science was more prophecy than reality." When there is need for a spiritual origin of all the world, follow Plato. Rationalize the supernatural; in the beginning place the Word. Yet nature remains, save for the one great Act, cosmos forever.

This is a splendid work for the reflective reader inexperienced in the close study of Greek texts. The polar issues implied are hard to evade. One wonders, however, if we really know the Greek mind. Perhaps Plato actually saw and described a physical model of his World

Soul, its moving parts all made of copper. Did the Greeks indeed discover the system of wandering planets only a little while before Plato wrote? Or do we misjudge that ancient world by the few texts that remain? Underneath it all another rich and more specialized world of thought, closer to the craftsman's hand and the stargazer's ancient observations, might remain buried, to be revealed—or not—someday, even as the divers brought up the wonderful geared planet model (from Hellenistic times) out of the cove at Antikythera.

EVERYTHING IN ITS PATH: DESTRUCTION OF COMMUNITY IN THE BUFFALO CREEK FLOOD, by Kai T. Erikson. Simon and Schuster ($8.95). The Guyandotte River flows into the broad Ohio above the city of Huntington, W.Va. Like many of its fellow tributaries, it drains the dissected plateau of the western Appalachians through the branching pattern of smaller creeks, which make their way down from the uplands a little east of the river valley. Sixty winding miles above the Ohio you come to the settlement of Man, W.Va. At Man it is Buffalo Creek that offers its portion to the main stream, draining the highland through a slightly winding 17-mile hollow, which at Man is a couple of hundred yards wide but up at the head of the hollow, where near the mine three forks meet to form the creek itself, wide enough only to hold a road, the rail lines and a row or two of houses in the shadow of the slope.

Some 5,000 people lived along that hollow, their frame houses here and there on the banks in a loose string. Sixteen villages were distinct enough on the map, but in fact they formed a long unbroken settlement, small neighborhoods marked by a church, a store, a post office. Every day the C&O freight cars rattled up the creek pretty empty, and every day they came back, a 100-car trainload of coal. The men went daily to the top of the hollow, mostly to work underground at the coal face, as men have done in Buffalo Creek since the rail spur was built in 1912. Down from the mine waste pond one cold, rainy morning in February, 1972, came a wall of water black and heavy with slag, moving at the easy speed of a marathon runner, floating the houses before it like toys, battering and drowning as it came, taking everything in its path.

Professor Erikson entered the hollow straight from New Haven about a year after the flood, a sociologist engaged—in both senses of the word—in establishing the nature of the injury done the people of the hollow as part of their suit for damages against the Pittston Corporation, which owned the mine and had proved a poor steward of it. His book is an eloquent, perceptive, warm and authenticated tale of a complex people all too simple in disaster. Much of what he tells us is in the recorded speech of the community, talking reflectively and self-perceptively as they are questioned by a man of depth recognized as a friend. It comes clear that over their history the mountain people (in a long list of their names not one suggests an origin outside Britain and Ireland) have been, like all of us, both winners and losers. They lost the great forests to the lumbermen, who left a stump-filled green second growth everywhere to replace trees such as the "huge sycamore" George Washington had measured to be 45 feet around. "The grand prize, however, was coal." Pit and strip, the mine operators did about as they pleased with space, mountain and water.

Buffalo Creek was a successful coal camp; wages were high, the union writ ran pretty strong and although automation came to mean rather few jobs, those who remained were a "sort of blue-collar royalty." These mountain woodsmen brought their skills to the coal camp; "almost every man in the hollow" could wire a house, repair an engine, build a garage or find his way skillfully around the woods with a rifle and a fishing line. One old man reported crisply: "Before the flood was I'd get out and run the woods and deer hunt. And ginseng and bee hunt and squirrel hunt." The hardworking, prospering miners of Buffalo Creek had built a perilous community, balanced between the old individualism and the new need to see the hollow as a place of order and pride, a commitment to hard work in the company of neighbors who shared a common past, a common present and a common future.

The flood washed it all downstream. The day itself was terror. "I cannot explain that water as being water. It looked like a black ocean." "There was houses on top of it. A big tank blew up.... One mass of ugliness was all it was to me." The day ended silent, black, cold, lonely. "Everything was wet and black. I went into the bathroom and stepped on a body at the door." About 125 dead were flung through the hollow, hung on tipples and trees, lodged in houses, turned up by the blades of the bulldozers.

The first organized help at rehousing the hollow's people was provided by the Department of Housing and Urban Development: rent-free trailer camps for a year for about half of the original population. This quick, compassionate action had a hidden flaw. The survivors were put under a dry roof as fast as possible, first come, first served. The result was a scrambling of neighborhoods; the people, dazed and numbed by close contact with random death, bearing the guilt of survival, were joined in an understandable fear of protracted rain and lower-

ing skies. They now found themselves alone, without friends and neighbors, shuffled miles away. The community of the hills was broken; these people, who had had no fences, who were individualists and yet were realized only in neighborhoods, had no more support from their collectivity. "It has changed from the community of Paradise to Death Valley." "Seems like it's frozen." "What I miss most is the friendliness and closeness." "Now everybody is alone.... The people are gone, scattered... there's no next door. You can't laugh with friends." The water had indeed taken everything in its path.

There is now a small academic industry of disaster studies. Tornado, flood and hurricane are seen as stand-ins for wider disasters brought by war. The hill people have been the victims of "one long sustained disaster"; the old balance is gone from Buffalo Creek. When they report their loss of interest in an intact household, we hear the hint of the depressed and shiftless mountain hollow. When they tell us of illness and the need for compensation, we hear of the old dependency of the lonely and overlooked. There is a chronic source of neurotic trauma too, not only an acute one. Are we not, all of us, in a world where we feel a loss of power, a sense of generalized fear, a numbness to the suffering of those whom we see as television phantoms, in a kind of chronic phase of trauma? Historians may look back on our time to conclude that "the traumatic neuroses were its true clinical signature." We are all survivors, so far.

Energy is not the only measure of process, but it lies deep in the world. The potential energy released when that huge ponded slag mound turned soggy and gave way on Buffalo Creek lies in the range of 20 kilotons of TNT equivalent. Our country has some tens of thousands of nuclear warheads with that yield and more. We wait, ready to create 10,000 Buffalo Creeks across Europe and Asia some cold winter morning, each with its fireball instead of a wall of water, and an invisible radioactive stain instead of the black smear of wet muck and coal dust. The Pittston Corporation simply neglected the rising impondment at Buffalo Creek too long. Are our high statesmen more attentive?

ISOTOPE SEPARATION, by Stelio Villani. American Nuclear Society, 555 North Kensington Avenue, LaGrange Park, Ill. 60525 ($29.80). When the world was young, natural processes could once in a while lead to the nuclear chain reaction. It happened in Gabon during the Precambrian, as readers of the July 1976 issue of *Scientific American* know. Nowadays it takes more doing. Although there are important exceptions, most uses of the controlled nuclear chain reaction call for the enrichment of the natural isotope mixture of some chemical element in a single desirable nuclear species. The usual process is the enrichment of natural uranium in the fraction of the light isotope, mass 235, which undergoes fission with thermal neutrons; the ratio today in nature is one U-235 atom in about 140. Another important process removes from water nearly all its protons, which do the lion's share of neutron absorption in that abundant heat-transfer medium. There are some 6,500 protons for every deuteron in nature. One then uses pure heavy water to moderate the neutrons of the uranium chain reaction. Special isotopes of lithium and boron are also in trade in laboratory quantities.

This volume is an uncompromising survey of the principles and practices of isotope separation on the engineering scale. It includes the needed introduction to the general questions of isotopic physics, to isotope analysis (one remembers that the Gabon discovery came about because the French analysts could routinely notice discrepancies in isotope abundance to the fourth decimal place) and to the physical principles of separation. Fourteen distinct methods are surveyed. They are ingenious enough, depending on either the mass motion of the atoms or the effect of isotope mass on the zero-point energies and hence on chemical equilibrium. A final method, the photochemical, depends on taking advantage of the small differences in spectral-line energies to split one isotope from another. Nearly all these methods result in small enrichment ratios, particularly for the isotopes of uranium, whose difference in speed is less than 1 percent at equal kinetic energy.

The central burden of the book, then, is the discussion of the ways the modern physicochemical engineer has found to magnify any small separation by massive repetition. The trick resembles the electronic-circuit idea of feedback; the slightly enriched material is fed on to the next stage, to repeat the slight separation. The depleted fraction, however, is fed back to the beginning of the step to try again. A cascade results, a set of many stages with an ever enriched feed flowing forward and a steadily depleted stream coming back. The diagrams are like electric circuits, and the fundamental equations, expressed either as difference equations or in the limit as continuous differential equations, lead to a considerable body of complicated optimizations and other useful solutions. It is not hard to accept the main result, which is that the process is fundamentally logarithmic in nature. The details are long and abstract; Villani has a clear style but he does not lead his reader by easy stages. He is a Milan physicist, an international expert belonging to the

Brussels Joint Research Centre of the European Communities who is now on assignment to the laboratories at Chalk River in Canada.

He treats concretely the design and economics of the two main successes so far in uranium separation: the gaseous diffusion plants and their new rivals, the centrifuges. The big plants are at work in the U.S., France, Germany, the U.S.S.R. and China (although the last two countries are not discussed at all). In all these plants a hot vapor of uranium hexafluoride is pumped and diffused through many stages of sintered-metal barriers. The entire plant is under a heavy threat of corrosion; stainless steel alone works well. UF_6 is intolerant of humidity, of any organic contamination, of lubricants and gaskets; the development of the fluorine polymers we call Teflon arose from the properties of uranium hexafluoride.

Such a plant is a user of much electric power; the cost of its product is mostly the cost of its electricity. In burning the uranium there is a net gain of power, by a factor of about 10. Big, humming sealed pumps, tanks and pipes in a score of big factory bays—that is the diffusion plant. The centrifuge plant is quite different. It is a "short and wide" cascade, with fewer than 10 stages but many thousands of small, fast-spinning machines in parallel, where the diffusion plant is a "narrow, long" cascade, having 1,000 stages in series with big flow in each of the early ones.

The centrifuges are cheap, but they are many. The power costs go down; the real issue is the durability of the fast-spinning machines, their one-horsepower motors whirring them ceaselessly at 20,000 or 30,000 revolutions per minute. A British-Dutch-German demonstration plant is just under way. The possibilities of the subdivision of centrifuges among small, hidden parallel plants have not been lost on those who are worrying about nations evading international agreements on weapons production. The plutonium problem is not the only one in town.

The book is a useful, if austere, guide to this remarkable and not very transparent industry as of the year 1973. Changes are afoot, but they are not yet reported at book length.

SCIENTIFIC QUOTATIONS: THE HARVEST OF A QUIET EYE, selected by Alan L. Mackay, edited by Maurice Ebison and illustrated by John Taylor. Crane, Russak & Company, Inc. ($14). This attractive and useful anthology, a delightful "work of pure plagiarism," offers 167 pages of citations, drawn from the perception and wit of authors from the Buddha to Paul Ehrlich. The great are here; plenty is heard from Goethe and Einstein, Shakespeare and the Bible, the

items chosen for merit and not for mere currency. The obscure are here too. A good number of the quotations are offered for the savor of their wrongheadedness rather than for pat wisdom. Who could ignore that Francis, Lord Jeffery, once complained testily: "Damn the Solar System. Bad light; planets too distant; pestered with comets; feeble contrivance; could make a better myself."

Mackay is a crystallographer at Birkbeck College of the University of London, and his book celebrates the style of his great master J. D. Bernal. Like Bernal's, his is a mind of energy and catholic learning, with a good sense of the comical. He reserves his deep concern for the unity of all knowledge and its impact on society and human welfare. Most of the quotations will seem apt to anyone with a bent for learning, not only to a scientist. (A sprinkling of them remain directed at those who bring a feeling for the jargon and the experiences of the physics classroom or the biological laboratory.) The author-anthologist has offered us an unexpected number of gems out of the mine of the Marxist classics, a vein too neglected in the run of American thought. The volume is particularly well presented typographically, with a number of witty ink drawings, portraits of authors and unusual evocations of the text. (Ezra Pound's "Our science is from the watching of shadows" brings out a fine sketch of the great masonry gnomon at Jaipur.) The list has been twice selected; the editor (from the Institute of Physics in London), who abridged Mackay's big "exotic landscape of diverse interests," has pruned the book so that it is fresh, even perhaps a little wry. There is a useful index of "keywords and catch phrases"; one wishes only that the host of references could have been made bibliographically more detailed.

Both Rabelais and Montaigne warned us: "Science without conscience is but death of the soul." It was Thomas Mann who seems to have supplied Niels Bohr with an invaluable paradox: "A great truth is a truth whose opposite is also a great truth." V. I. Lenin understood what many bring forward now as a new viewpoint: "It is absurd to deny the role of fantasy in even the strictest science." The first of the scientific journals carried in its first issue a piece by the editor that said: "Nobody should find it strange to see here opinions different from his own concerning the sciences, because we aim to report the ideas of others without guaranteeing them." So wrote Denis de Sallo in the *Journal des Scavans* in 1665. (It is not clear whether he was already sending papers to a referee.) Leave the final unexpected words to our old foe George III: "I spend money on war because it is necessary, but to spend it on science, that is pleasant to me."

GET MOVING, AMERICA!

It feels good to feel fit
physically, mentally, emotionally
So learn a skill you can play for life

Sponsored by
American Alliance for Health, Physical Education and Recreation
1201-16th Street, N W , Washington, D C. 20036

AAHPER

BIBLIOGRAPHY

Readers interested in further exploration of the subjects covered by the articles in this issue may find the following lists of publications helpful.

THE SALT NEGOTIATIONS

SALT: THE MOSCOW AGREEMENTS AND BEYOND. Edited by Mason Willrich and John Rhinelander. The Free Press, 1974.

ANNUAL DEFENSE DEPARTMENT REPORT, FY 78. Donald Rumsfeld. Government Printing Office, January 17, 1977.

CRUISE MISSILES. Kosta Tsipis in *Scientific American,* Vol. 236, No. 2, pages 20–29; February, 1977.

PRESIDENT CARTER DISCUSSES STRATEGIC ARMS LIMITATION PROPOSALS in *The Department of State Bulletin,* Vol. LXXVI, No. 1974, pages 409–414; April 25, 1977.

BL LACERTAE OBJECTS

QUASI-STELLAR OBJECTS. Geoffrey Burbidge and Margaret Burbidge. W. H. Freeman and Company, 1967.

RADIO PROPERTIES OF BL LAC TYPE OBJECTS. D. R. Altschuler and J. F. C. Wardle in *Nature,* Vol. 255, No. 5506, pages 306–310; May 22, 1975.

THE BL LACERTAE OBJECTS. W. A. Stein, S. L. O'Dell and P. A. Strittmatter in *Annual Review of Astronomy and Astrophysics,* Vol. 14, pages 173–195; 1976.

LIGHT-WAVE COMMUNICATIONS

COMMUNICATION BY LASER. Stewart E. Miller in *Scientific American,* Vol. 214, No. 1, pages 19–27; January, 1966.

COMMUNICATION BY OPTICAL FIBER. J S. Cook in *Scientific American,* Vol. 229, No. 5, pages 28–35; November, 1973.

SPECIAL ISSUE: LIGHTWAVE COMMUNICATIONS in *Physics Today,* Vol. 29, No. 5, pages 23–60; May, 1976.

LIGHTWAVE COMMUNICATIONS PASSES ITS FIRST TEST. Ira Jacobs in *Bell Laboratories Record,* Vol. 54, No. 11, pages 290–297; December, 1976.

OPTICAL TRANSMISSION OF VOICE AND DATA. Ira Jacobs and Stewart E. Miller in *IEEE Spectrum,* Vol. 14, No. 2, pages 32–41; February, 1977.

THE FLOW OF HEAT FROM THE EARTH'S INTERIOR

GLOBAL HEAT FLOW: A NEW LOOK. David S. Chapman and Henry N. Pollack in *Earth and Planetary Science Letters,* Vol. 28, pages 23–32; November, 1975.

CONTINENTS ADRIFT AND CONTINENTS AGROUND: READINGS FROM *SCIENTIFIC AMERICAN.* Introductions by J. Tuzo Wilson. W. H. Freeman and Company, 1976.

AN ANALYSIS OF THE VARIATION OF OCEAN FLOOR BATHYMETRY AND HEAT FLOW WITH AGE. Barry Parsons and John G. Sclater in *Journal of Geophysical Research,* Vol. 82, No. 5, pages 803–827; 1977.

REGIONAL GEOTHERMS AND LITHOSPHERE THICKNESS. David S. Chapman and Henry N. Pollack in *Geology,* Vol. 5, No. 5, pages 265–268; May, 1977.

KANGAROOS

LIFE OF MARSUPIALS. H. Tyndale-Biscoe. Edward Arnold, 1973.

BIOGEOGRAPHICAL CONSIDERATIONS OF THE MARSUPIAL-PLACENTAL DICHOTOMY. Jason A. Lillegraven in *Annual Review Ecology and Systematics,* Vol. 5, pages 74–94; 1974.

THE BIOLOGY OF KANGAROOS (*MARSUPIALIA-MACROPODIDAE*). Eleanor M. Russell in *Mammal Review,* Vol. 4, No. 1 and No. 2, pages 1–59; March/June, 1974.

THE GAS VACUOLES OF BLUE-GREEN ALGAE

THE GAS-VACUOLES OF THE MYXOPHYCEAE (CYANOPHYCEAE). G. E. Fogg in *Biological Reviews of the Cambridge Philosophical Society,* Vol. 16, No. 3, pages 205–217; July, 1941.

BLUE-GREEN ALGAE: FINE STRUCTURE OF THE GAS VACUOLES. C. C. Bowen and T. E. Jensen in *Science,* Vol. 147, No. 3664, pages 1460–1462; March 19, 1965.

ISOLATION AND CHEMICAL CHARACTERIZATION OF GAS-VACUOLE MEMBRANES FROM *MICROCYSTIS AERUGINOSA* KUETZ. EMEND. ELENKIN. Daniel D. Jones and Michael Jost in *Archiv für Mikrobiologie,* Vol. 70, No. 1, pages 43–64; 1970.

STRUCTURE AND FUNCTION OF GAS VACUOLES. A. E. Walsby in *Bacteriological Reviews,* Vol. 36, No. 1, pages 1–32; March, 1972.

THE HISTORY OF THE AIRFLOW CAR

AUTOMOBILE QUARTERLY'S WORLD OF CARS. Edited by Richard M. Langworth. Clarke, Irwin & Company, 1971.

SEVENTY YEARS OF CHRYSLER. George H. Dammann. Crestline Publishing Co., 1974.

THE CAR CULTURE. James J. Flink. The MIT Press, 1975.

THE STREAMLINED DECADE. Donald J. Bush. George Braziller, 1975.

CHROME DREAMS: AUTOMOBILE STYLING SINCE 1893. Paul C. Wilson. Chilton Book Company, 1976.

"SECOND MESSENGERS" IN THE BRAIN

CYCLIC AMP. G. Alan Robison, Reginald W. Butcher and Earl W. Sutherland. Academic Press, 1971.

METABOLISM AND FUNCTIONS OF CYCLIC AMP IN NERVE. George I. Drummond and Yvonne Ma in *Progress in Neurobiology,* Vol. 2, Part 2, pages 119–176; 1973.

THE ROLE OF CYCLIC NUCLEOTIDES IN CENTRAL SYNAPTIC FUNCTION. Floyd E. Bloom in *Reviews of Physiology, Biochemistry and Pharmacology,* Vol. 74, pages 1–103; 1975.

POSSIBLE ROLE FOR CYCLIC NUCLEOTIDES AND PHOSPHORYLATED MEMBRANE PROTEINS IN POSTSYNAPTIC ACTIONS OF NEUROTRANSMITTERS. Paul Greengard in *Nature,* Vol. 260, No. 5547, pages 101–108; March 11, 1976.

CYCLIC NUCLEOTIDES IN THE NERVOUS SYSTEM. John W. Daly. Plenum Press, 1977.

CYCLIC NUCLEOTIDES AND NERVOUS SYSTEM FUNCTION. James A. Nathanson in *Physiological Reviews,* Vol. 57, pages 157–256; 1977.

MATHEMATICAL GAMES

THE CODEBREAKERS. David Kahn. The Macmillan Company, 1967.

A FAST MONTE-CARLO TEST FOR PRIMALITY. R. Solovay and V. Strassen in *SIAM Journal on Computation,* Vol. 6, No. 1, pages 84–85; March, 1977.

THE AMATEUR SCIENTIST

VIBRATIONS OF EVAPORATING LIQUID DROPS. Norman J. Holter and Wilford R. Glasscock in *The Journal of the Acoustical Society of America,* Vol. 24, No. 6, pages 682–686; November, 1952.

THE LEIDENFROST PHENOMENON: FILM BOILING OF LIQUID DROPLETS ON A FLAT PLATE. B. S. Gottfried, C. J. Lee and K. J. Bell in *International Journal of Heat and Mass Transfer,* Vol. 9, No. 11, pages 1167–1187; November, 1966.

ON THE FIXATION OF WATER IN DIVERSE FIRE. Johann G. Leidenfrost, translated by Carolyn Wares in *International Journal of Heat and Mass Transfer,* Vol. 9, No. 11, pages 1153–1166; November, 1966.

Amédée Gordini, known as "The Sorcerer" for his magic with racing cars, presents one of his greatest feats.

Throughout Europe the cars of Amédée Gordini represented France in Grand Prix and sports car racing for many years. The French bestowed the name "The Sorcerer" upon him because of his inimitable touch with racing engines and high performance cars. And he bestowed his name on this coupe-convertible.

At the heart of the Gordini is the engine designed by Gordini. This 1647-cc, four-stroke engine with a high compression ratio and electronic fuel injection will more than satisfy the requirements of any driver who admires pure performance. Hooked to the overhead valve engine is a smooth five-speed gearbox that gives you total command of acceleration.

When you add front-wheel drive, rack and pinion steering, Michelin steel-belted radials, and a sophisticated suspension system with front and rear stabilizer bars the end result is a highly responsive sports car. A car that can cruise effortlessly all day as fast as you want to go. A car with superior handling and the road-holding ability that the French are famous for. Now add another feature that the French are famous for — style and luxury.

In front, infinitely-adjustable soft reclining bucket seats. In back, gracefully sculptured contour seats.

The Gordini's fully carpeted interior is spacious enough to seat four people very comfortably. In the rear there's a roomy 10-cubic foot trunk.

The Gordini's three-way top is something you won't find in any other car at any price. The sturdy fiberglass hardtop lifts off to reveal a soft, supple convertible top that is electrically operated.

Brock Yates of Car and Driver concluded "with a price tag under $7,000 the Gordini provides power, speed, quality and handling generally unknown in automobiles costing less than $10,000." Now that's some trick. Call 800-631-1616 for your nearest Gordini dealer.

STANDARD FEATURES	Gordini	BMW 320i	Porsche 924
Rack & pinion steering	YES	YES	YES
5-speed transmission	YES	NO	NO
Front-wheel drive	YES	NO	NO
Electronic fuel injection	YES	YES	YES
Electric sun-roof	YES	NO	NO
PERFORMANCE			
*0-60	9.2	10.5	9.9
*Standing ¼ mile	17.1	17.3	17.3
**EPA fuel estimates	36 H'wy 21 City	29 H'wy 20 City	31 H'wy 17 City
†BASE PRICE	$6998	$7990	$9395

*Test results reported in July, August and December 1976 issues of Car and Driver magazine. Cars tested were a 1976 Gordini, 1976 Porsche and 1977 BMW.
**The actual mileage you get will vary depending on the type of driving you do, your driving habits, your car's condition and optional equipment. Not available in California. †East Coast P.O.E.

The Great Gordini

by Renault

Never buy a white Bordeaux by the bottle. Buy a crisp, white Bordeaux by the label.

B&G.
The B&G label.
250 years of wine making history from all the great wine regions of France: Bordeaux, The Loire Valley, Beaujolais, Côtes du Rhône and Burgundy.

Now, B&G brings you Pontet-Latour, the crisp white wine from Bordeaux.

Elegant. Affordable. Buy it by the label.

B&G.

Letters of recommendation from France