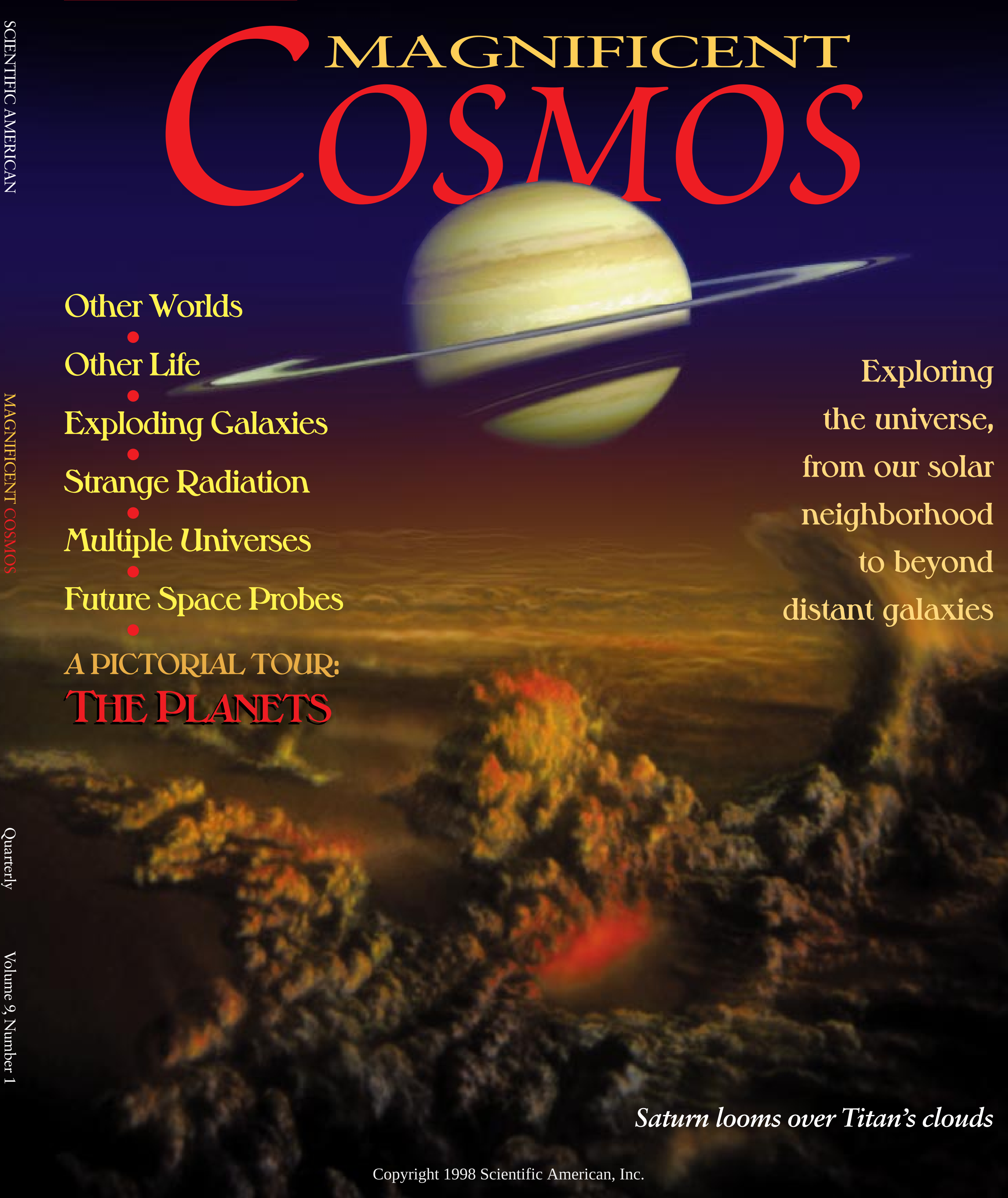
SCIENTIFIC AMERICAN
PRESENTS

Since the Big Bang • How Stars Live and Die • Dark Matter

QUARTERLY $4.95 DISPLAY UNTIL MAY 31, 1998

MAGNIFICENT
COSMOS



Other Worlds
•
Other Life
•
Exploding Galaxies
•
Strange Radiation
•
Multiple Universes
•
Future Space Probes
•
A PICTORIAL TOUR:
THE PLANETS

Exploring
the universe,
from our solar
neighborhood
to beyond
distant galaxies

Saturn looms over Titan's clouds

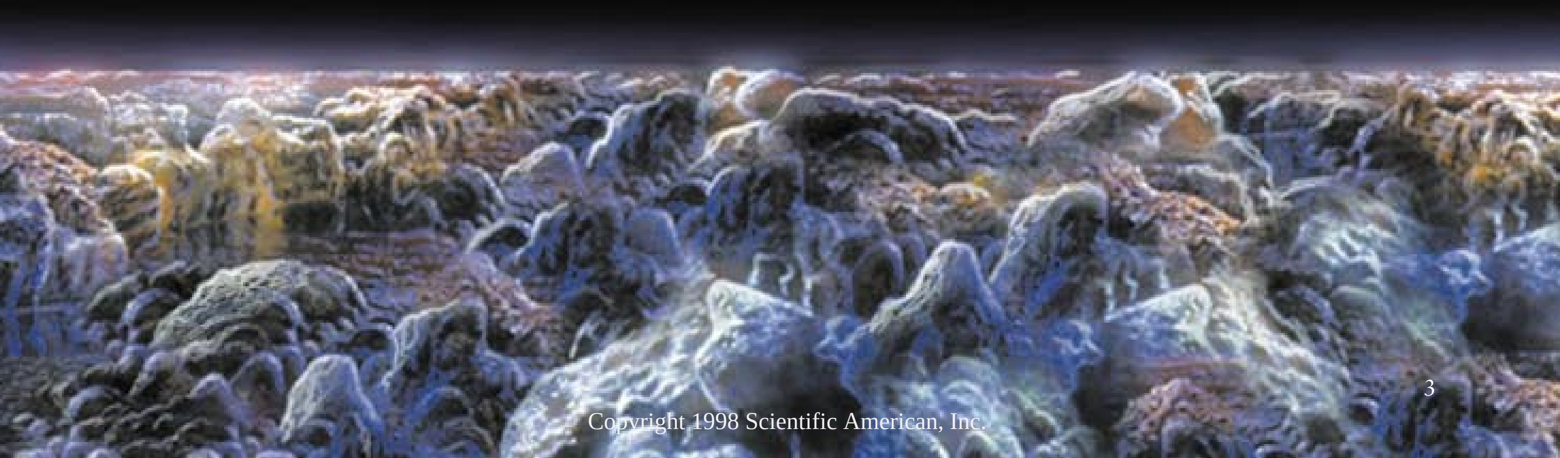
SCIENTIFIC AMERICAN MAGNIFICENT COSMOS Quarterly Volume 9, Number 1

SCIENTIFIC AMERICAN PRESENTS

Spring 1998
Volume 9 Number 1

MAGNIFICENT COSMOS

Scientific American quarterly (ISSN 1048-0943), Volume 9, Number 1, 1998, published quarterly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals Publication Rate Pending. Postage paid at New York, N.Y., and at additional mailing offices. Canadian BN No. 127387652RT; QT No. Q1015332537. Subscription rates: one year $19.80 (outside U.S. $23.80). To purchase additional quantities: 1 to 9 copies: U.S. $4.95 each plus $2.00 per copy for postage and handling (outside U.S. $5.00 P & H); 10 to 49 copies: $4.45 each, postpaid; 50 copies or more: $3.95 each, postpaid. Send payment to Scientific American, Dept. SAQ, 415 Madison Avenue, New York, N.Y. 10017-1111. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, IA 51537. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (800) 333-1199 or (515) 247-7631.**



FROM THE EDITORS

Treasures in the Stars

Exploration of space has sprinted forward over the past two decades, even though no human has ventured outside the lunar orbit. Thanks to strings of probes with names like Voyager, Pioneer, Galileo, Magellan and SOHO, planetary and solar science thrived. We have seen all the planets but Pluto from close by, visited Mars and Venus by proxy, and even witnessed the collision of Comet Shoemaker-Levy with Jupiter. The moons graduated from minor players to varied, exotic worlds in their own right and possibly to abodes for life. The sun revealed its complex internal anatomy. Whole new classes of frozen bodies beyond Neptune's orbit came into view.

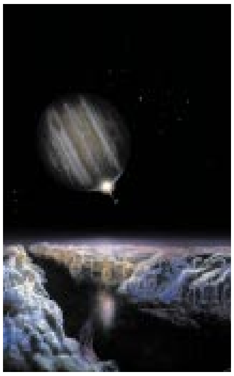
Meanwhile the magnificent Hubble Space Telescope, other orbiting instruments and their Earth-bound cousins peered clearly into deeper space. They showed us new types of galaxies and stars, spotted planets around other suns and took the temperature of the big bang. We better appreciated our own solar system after seeing how fiercely bright some corners of the universe burn.

With this issue, *Scientific American* summarizes the most extraordinary discoveries and still open mysteries of modern astronomy. It also debuts the new series of *Scientific American Presents* quarterlies, each of which will look in depth at a single topic in science or technology. (The regular monthly magazine will, of course, continue to scan the full range of disciplines.)

All the authors of this issue deserve thanks for their fully new articles or for the extensive updates they made to previous works. But I must with sadness extend special appreciation to the late cosmologist David N. Schramm, whose untimely death in December 1997 immediately followed our collaboration. We mourn him for both his many kindnesses and his scientific vision. I am grateful also to the Lockheed Martin Corporation for its generous offer to become the sole sponsor of this issue; such financial support, unfettered by editorial constraints, helps to ensure that we can bring to readers the information they crave at a price they can afford. My deepest gratitude, though, goes to editor Rick Lipkin and, as always, the rest of the staff of *Scientific American*, for their unfailing industry and love of good science.

John Rennie

JOHN RENNIE, *Editor in Chief*
editors@sciam.com



About the Cover and the Table of Contents

These paintings by Don Dixon imagine the views from two fascinating moons in our solar system. The scene at the left is set on the Jovian moon Europa, showing liquid water through a fissure in the icy surface. The cover image offers a perspective just above the methane clouds of the moon Titan as it orbits Saturn.

SCIENTIFIC AMERICAN® PRESENTS

Magnificent Cosmos is published by the staff of Scientific American, with project management by:

John Rennie, EDITOR IN CHIEF

Michelle Press, MANAGING EDITOR

Richard Lipkin, ISSUE EDITOR
Sasha Nemecek, ASSISTANT EDITOR

STAFF WRITERS: Timothy M. Beardsley;
Steve Mirsky; Madhusree Mukerjee; Glenn Zorpette

Art
Edward Bell, Jessie Nathans, ART DIRECTORS
Bridget Gerety, PHOTOGRAPHY EDITOR
Meghan Gerety, PRODUCTION EDITOR

Copy
Maria-Christina Keller, COPY CHIEF
Molly K. Frances; Daniel C. Schlenoff;
Katherine A. Wong; Stephanie J. Arthur; William Stahl

Administration
Rob Gaines, EDITORIAL ADMINISTRATOR

Production
Richard Sasso, ASSOCIATE PUBLISHER/
VICE PRESIDENT, PRODUCTION
William Sherman, DIRECTOR, PRODUCTION
Janet Cermak, MANUFACTURING MANAGER
Tanya DeSilva, PREPRESS MANAGER
Silvia Di Placido, QUALITY CONTROL MANAGER
Carol Hansen, COMPOSITION MANAGER
Madelyn Keyes, SYSTEMS MANAGER
Carl Cherebin, AD TRAFFIC; Norma Jones

Circulation
Lorraine Leib Terlecki, ASSOCIATE PUBLISHER/
CIRCULATION DIRECTOR
Katherine Robold, CIRCULATION MANAGER
Joanne Guralnick, CIRCULATION PROMOTION MANAGER
Rosa Davis, FULFILLMENT MANAGER

Advertising
Kate Dobson, ASSOCIATE PUBLISHER/ADVERTISING DIRECTOR
OFFICES: NEW YORK:
Thomas Potratz, EASTERN SALES DIRECTOR;
Kevin Gentzel; Stuart M. Keating; Timothy Whiting.
DETROIT, CHICAGO: 3000 Town Center, Suite 1435,
Southfield, MI 48075;
Edward A. Bartley, DETROIT MANAGER; Randy James.
WEST COAST: 1554 S. Sepulveda Blvd., Suite 212,
Los Angeles, CA 90025;
Lisa K. Carden, WEST COAST MANAGER; Debra Silver.
225 Bush St., Suite 1453,
San Francisco, CA 94104
CANADA: Fenn Company, Inc. DALLAS: Griffith Group

Business Administration
Joachim P. Rosler, PUBLISHER
Marie M. Beaumonte, GENERAL MANAGER
Alyson M. Lane, BUSINESS MANAGER
Constance Holmes, MANAGER, ADVERTISING ACCOUNTING
AND COORDINATION

Chairman and Chief Executive Officer
John J. Hanley

Corporate Officers
Joachim P. Rosler, PRESIDENT
Robert L. Biewen, Frances Newburg, VICE PRESIDENTS
Anthony C. Degutis, CHIEF FINANCIAL OFFICER

Program Development
Linnéa C. Elliott, DIRECTOR

Electronic Publishing
Martin O. K. Paul, DIRECTOR

Ancillary Products
Diane McGarvey, DIRECTOR

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue • New York, NY 10017-1111
(212) 754-0550
PRINTED IN U.S.A.

I

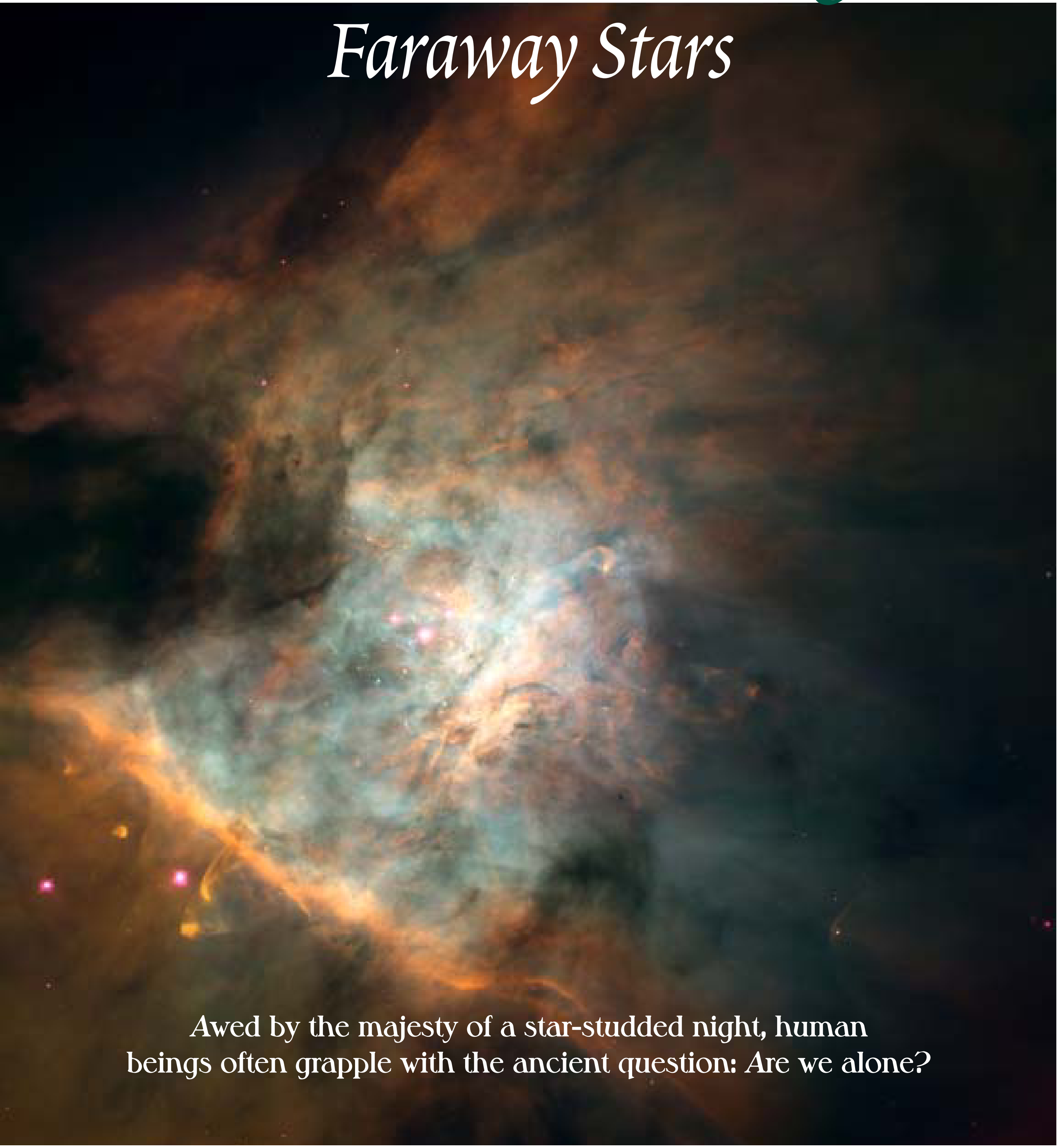
Discovering Worlds

- GIANT PLANETS ORBITING FARAWAY STARS
- SEARCHING FOR LIFE IN OUR SOLAR SYSTEM
- SEARCHING FOR LIFE IN OTHER SOLAR SYSTEMS
- PLANETARY TOUR

ILLUSTRATION BY DON DIXON

JUPITER AND IO RISING, as seen from Europa, a moon of Jupiter

Giant Planets Orbiting Faraway Stars



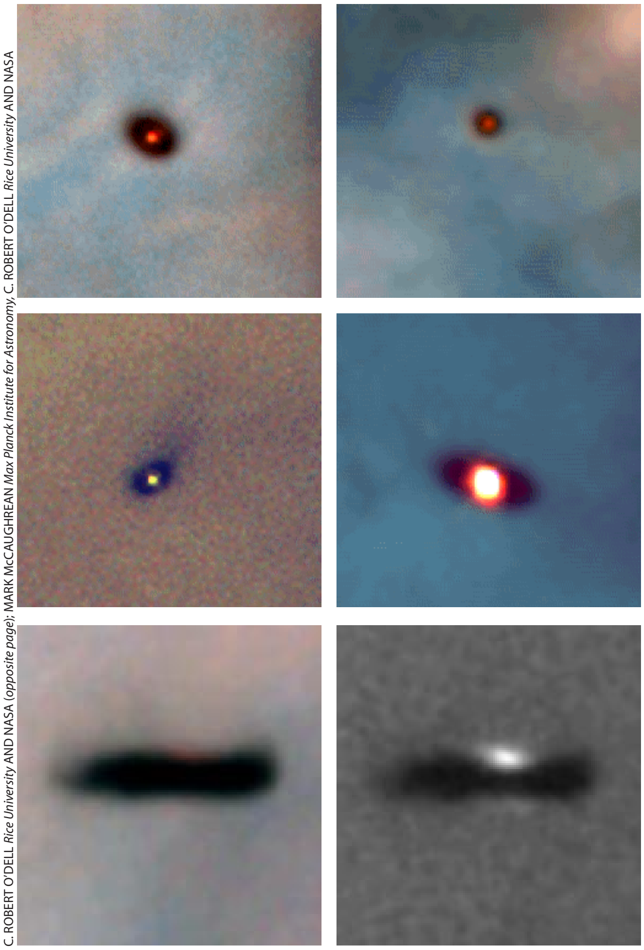
Awed by the majesty of a star-studded night, human beings often grapple with the ancient question: Are we alone?

by Geoffrey W. Marcy and R. Paul Butler

No doubt humans have struggled with the question of whether we are alone in the universe since the beginning of consciousness. Today, armed with evidence that planets do indeed orbit other stars, astronomers wonder more specifically: What are those planets like? Of the 100 billion stars in our Milky Way galaxy, how many harbor planets? Among those planets, how many constitute arid deserts or frigid hydrogen balls? Do some contain lush forests or oceans fertile with life?

For the first time in history, astronomers can now address these questions concretely. During the past two and a half years, researchers have detected eight planets orbiting sunlike stars. In October 1995 Michel Mayor and Didier Queloz of Geneva Observatory in Switzerland reported finding the first planet. Observing the star 51 Pegasi in the constellation Pegasus, they noticed a telltale wobble, a cyclical shifting of its light toward the blue and red ends of the spectrum. The timing of this Doppler shift suggests that the star wobbles because of a closely orbiting planet, which revolves around the star fully every 4.2 days—at a whopping speed of 482,000 kilometers (299,000 miles) an hour, more than four times faster than Earth orbits the sun.

ORION NEBULA (*left*), a turbulent maelstrom of luminous gas and brilliant stars, shows stellar formation under way. Located 1,500 light-years from Earth in the Milky Way's spiral arm, the nebula formed from collapsing interstellar gas clouds, yielding many hot, young stars. Among those are at least 153 protoplanetary disks believed to be embryonic solar systems. Below are six views of disks: four disks seen from above, plus a fifth viewed edge-on in two different wavelengths. Together they reveal gas and dust, circling million-year-old stars, that should eventually form planets. The disks' diameters range from two to 17 times that of our solar system.



C. ROBERT O'DELL *Rice University* AND NASA *(opposite page)*; MARK McCAUGHREAN *Max Planck Institute for Astronomy*, C. ROBERT O'DELL *Rice University* AND NASA

Another survey of 107 sunlike stars, performed by our team at San Francisco State University and the University of California at Berkeley, has turned up six more planets. Of those, one planet circling the star 16 Cygni B was independently discovered by astronomers William D. Cochran and Artie P. Hatzes of the University of Texas McDonald Observatory on Mount Locke in western Texas.

Detection of an eighth planet was reported in April 1997, when a nine-member team led by Robert W. Noyes of Harvard University detected a planet orbiting the star Rho Coronae Borealis. A ninth large object, which orbits the star known by its catalogue number HD114762, has also been observed—an object first detected in 1989 by astronomer David W. Latham of the Harvard-Smithsonian Center for Astrophysics and his collaborators. But this bulky companion has a mass more than 10 times that of Jupiter—large, though not unlike another large object discovered around the star 70 Virginis, a similar object with a mass 6.8 times that of Jupiter. The objects orbiting both HD114762 and 70 Virginis are so large that most astronomers are not sure whether to consider them big planets or small brown dwarfs, entities whose masses lie between those of a planet and a star.

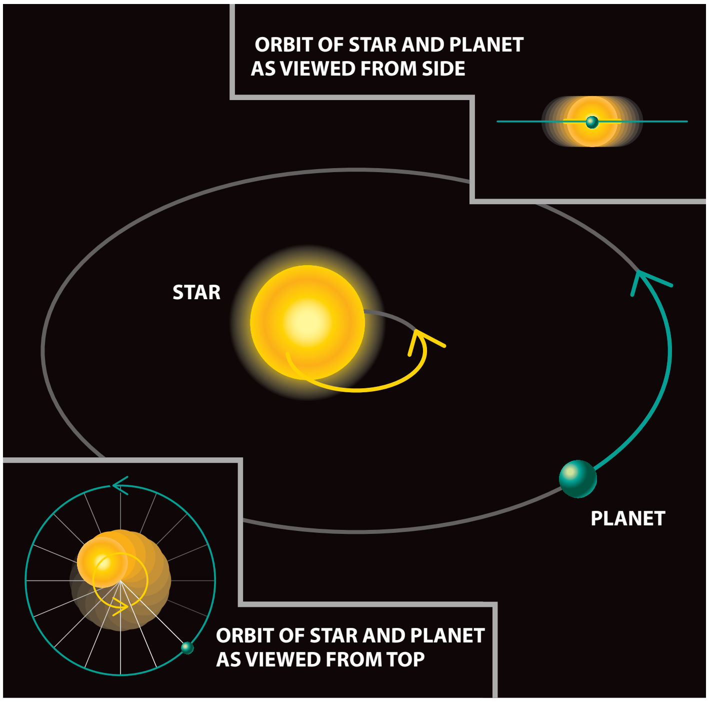
Detecting Extrasolar Planets

Finding extrasolar planets has taken a long time because detecting them from Earth, even using current technology, is extremely difficult. Unlike stars, which are fueled by nuclear reactions, planets faintly reflect light and emit thermal infrared radiation. In our solar system, for example, the sun outshines its planets about one billion times in visible light and one million times in the infrared. Because of the distant planets' faintness, astronomers have had to devise special methods to locate them. The current leading approach is the Doppler planet-detection technique, which involves analyzing wobbles in a star's motion.

Here's how it works. An orbiting planet exerts a gravitational force on its host star, a force that yanks the star around in a circular or oval path—which mirrors in miniature the planet's orbit. Like two twirling dancers tugging each other in circles, the star's wobble reveals the presence of orbiting planets, even though we cannot see them directly.

The trouble is that this stellar motion appears very small from a great distance. Someone gazing at our sun from 30 light-years away would see it wobbling in a circle whose radius measures only one seventh of one millionth of one degree. In other words, the sun's tiny, circular wobble appears only as big as a quarter viewed from 10,000 kilometers away.

Yet the wobble of the star is also revealed by the Doppler



PLANET ORBITING ITS HOST STAR causes the star to wobble. Although Earth-based astronomers have not yet been able to see an orbiting planet, they can deduce its size, mass and distance from its host by analyzing the to-and-fro oscillation of that star's light.

JARED SCHNEIDMAN DESIGN

effect of the starlight. As a star sways to and fro relative to Earth, its light waves become cyclically stretched, then compressed—shifting alternately toward the red and blue ends of the spectrum. From that cyclical Doppler shifting, astronomers can retrace the path of the star's wobble and, from Newton's law of motion, compute their masses, orbits and distances from their host stars. The cyclical Doppler shift itself remains extremely tiny: stellar light waves shrink and expand by only about one part in 10 million because of the pull of a large, Jupiter-like planet. The sun, for example, wobbles with a speed of only about 12.5 meters per second, pivoting around a point just outside its surface. To detect planets around other stars, measurements must be highly accurate, with errors in stellar velocities below 10 meters per second.

Using the Doppler technique, our group can now measure stellar motions with an accuracy of plus or minus three meters per second—a leisurely bicycling speed. To do this, we use an iodine absorption cell—a bottle of iodine vapor—placed near a telescope's focus. Starlight passing through the iodine is stripped of specific wavelengths, revealing tiny shifts in its remaining wavelengths. So sensitive is this technique that we can measure wavelength changes as small as one part in 100 million.

As recorded by spectrometers and analyzed by computers, a star's light reveals the telltale wobble produced by its orbiting companions. For example, Jupiter, the largest planet in our solar system, is one thousandth the mass of the sun. Therefore, every 11.8 years (the span of Jupiter's orbital period) the sun oscillates in a circle that is one thousandth the size of Jupiter's orbit. The other eight planets also cause the sun to wobble, albeit by smaller amounts. Take Earth, having a mass $^{1}/_{318}$ that of Jupiter and an orbit five times closer: it causes the sun to move a mere nine centimeters a second.

Yet some uncertainty about each extrasolar planet's mass remains. Orbital planes that astronomers view edge-on will give the true mass of the planet. But tilted orbital planes reduce the Doppler shift because of a smaller to-and-fro motion, as witnessed from Earth. This effect can make the mass appear smaller than it is. Without knowing a planet's orbital inclination, astronomers can compute only the least possible mass for the planet; the actual mass could be larger.

Thus, using the Doppler technique to analyze light from about 300 stars similar to the sun—all within 50 light-years of Earth—astronomers have turned up eight planets similar in size and mass to Jupiter and Saturn. Specifically, their masses range from about a half to seven times that of Jupiter, their orbital periods span 3.3 days to three years, and their distances from their host stars extend from less than one twentieth of Earth's distance to the sun to more than twice that distance [*see illustration on opposite page*].

To our surprise, the eight newly found planets exhibit two unexpected characteristics. First, unlike planets in our solar system, which display circular orbits, two of the new planets move in eccentric, oval orbits around their hosts. Second, five of the new planets orbit very near their stars—closer, in fact, than Mercury orbits the sun. Exactly why these huge planets orbit so closely—some skim just over their star's blazing coronal gases—remains unclear. These findings are mysterious, given that the radius of Jupiter's orbit is five times larger than that of Earth. These observations, in turn, provoke questions about our own solar system's origin, prompting some astronomers to revise the standard explanation of planet formation.

Reconsidering How Planets Form

What we have learned about the nine planets in our own solar system has constituted the basis for the conventional theory of planet formation. The theory holds that planets form from a flat, spinning disk of gas and dust that bulges out of a star's equatorial plane, much as pizza dough flattens when it is tossed and spun. This model shows the disk's material orbiting circularly in the same direction and plane as our nine planets do today. Based on this theory, planets cannot form too close to the star, because there is too little disk material, which is also too hot to coalesce. Nor do planets clump extremely far from the star, because the material is too cold and sparse.

Considering what we now know, such expectations about planets in the rest of the universe seem narrow-minded. The planet orbiting the star 47 Ursae Majoris in the Big Dipper constellation stands as the only one resembling what we expected, with a minimum bulk of 2.4 Jupiter-masses and a circular orbit with a radius of 2.1 astronomical units (AU)—1 AU representing the 150-million-kilometer distance from Earth to the sun. Only a bit more massive than Jupiter, this planet orbits in a circle farther from its star than Mars does from the sun. If placed in our solar system, this new planet might appear as Jupiter's big brother.

But the remaining planetary companions around other stars baffle us. The two planets with oval orbits have eccentricities of 0.68 and 0.40. (An eccentricity of zero is a perfect circle, whereas an eccentricity of 1.0 is a long, slender oval.) In contrast, in our solar system the greatest eccentricities appear in the orbits of Mercury and Pluto, both about 0.2; all other planets show nearly circular orbits (eccentricities less than 0.1).

These eccentric orbits have prodded astronomers to scratch their heads and revise their theories. Within two months of the first planet sighting, theorists hatched new ideas and adjusted the standard planet formation theory.

For instance, astronomers Pawel Artymowicz of the University of Stockholm and Patrick M. Cassen of the National Aeronautics and Space Administration Ames Research Center recalculated the gravitational forces at work when planets emerge from disks of gas and dust seen swirling around young, sunlike stars. Their calculations show that gravitational forces exerted by protoplanets—planets in the process of forming—on the gaseous, dusty disks create alternating spiral "density waves." Resembling the "arms" of spiral galaxies, these waves exert forces back on the forming planets, driving them from circular motion. Over millions of years, planets can easily wander from circular orbits into eccentric, oval ones.

A second theory also accounts for large orbital eccentricities. Suppose, for instance, that Saturn had grown much larger than it actually is. Conceivably, all four giant planets in our solar system—Jupiter, Saturn, Uranus and Neptune—could have swelled into bigger balls if our original protoplanetary disk had contained more mass or had existed longer. In this case, the solar system would contain four superplanets, exerting gravitational forces on one another, perturbing one another's orbits and causing them to intersect.

Eventually, some of the superplanets might be gravitationally thrust inward, others outward, an unlucky few even ejected from the planetary system. Like balls ricocheting on a billiards table, the scattered giant planets might adopt extremely eccentric orbits, as we now observe for three of the new planets. Interestingly, this billiards model for eccentric planets shows that we should be able to detect the massive planets causing eccentric orbits—planets perhaps orbiting farther out than the planets we have detected thus far. A variation on this theme suggests that a companion star, rather than other planets, might gravitationally scatter planet orbits.

The most bizarre of the new planets are the four so-called 51 Peg planets, which show orbital periods shorter than 15 days. The four members of this class are 51 Peg itself, Tau Bootis, 55 Cancri and Upsilon Andromedae, which have orbital periods of just 4.2, 3.3, 14.7 and 4.6 days, respectively.

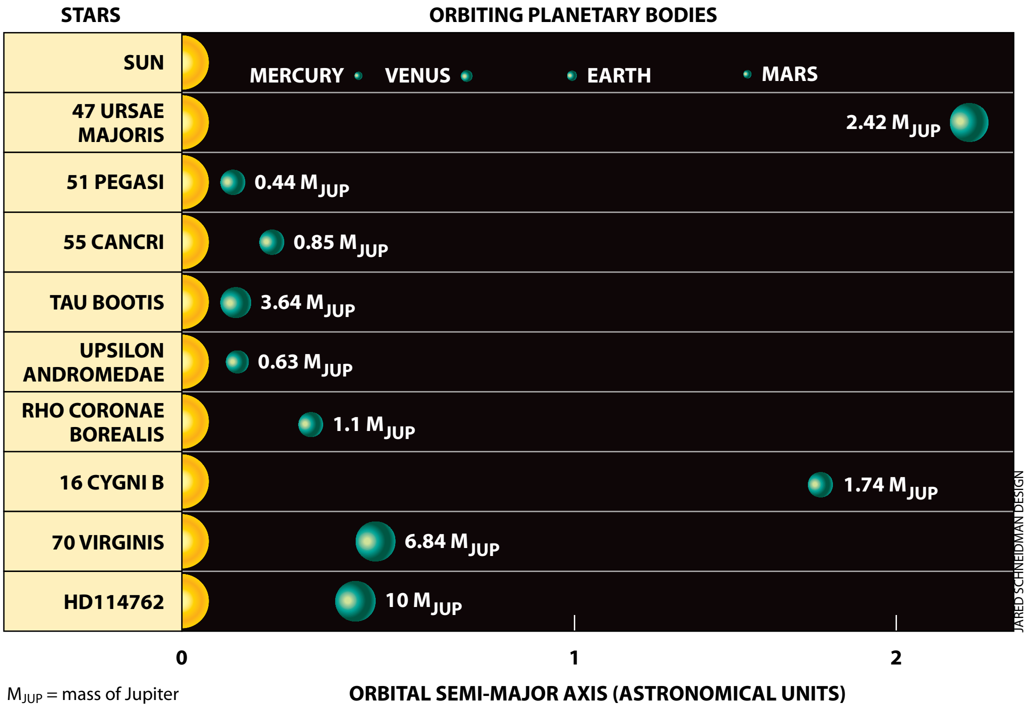
These orbits are all small, with radii less than one tenth the distance between Earth and the sun—indeed, less than one third of Mercury's distance from the sun. Yet these planets are as big as, or bigger than, the largest planet in our solar system. They range in mass from 0.44 of Jupiter's mass for 51 Peg to 3.64 of Jupiter's mass for Tau Bootis. Their Doppler shifts suggest that these planets orbit in circles.

Mysterious 51 Pegasi–Type Planets

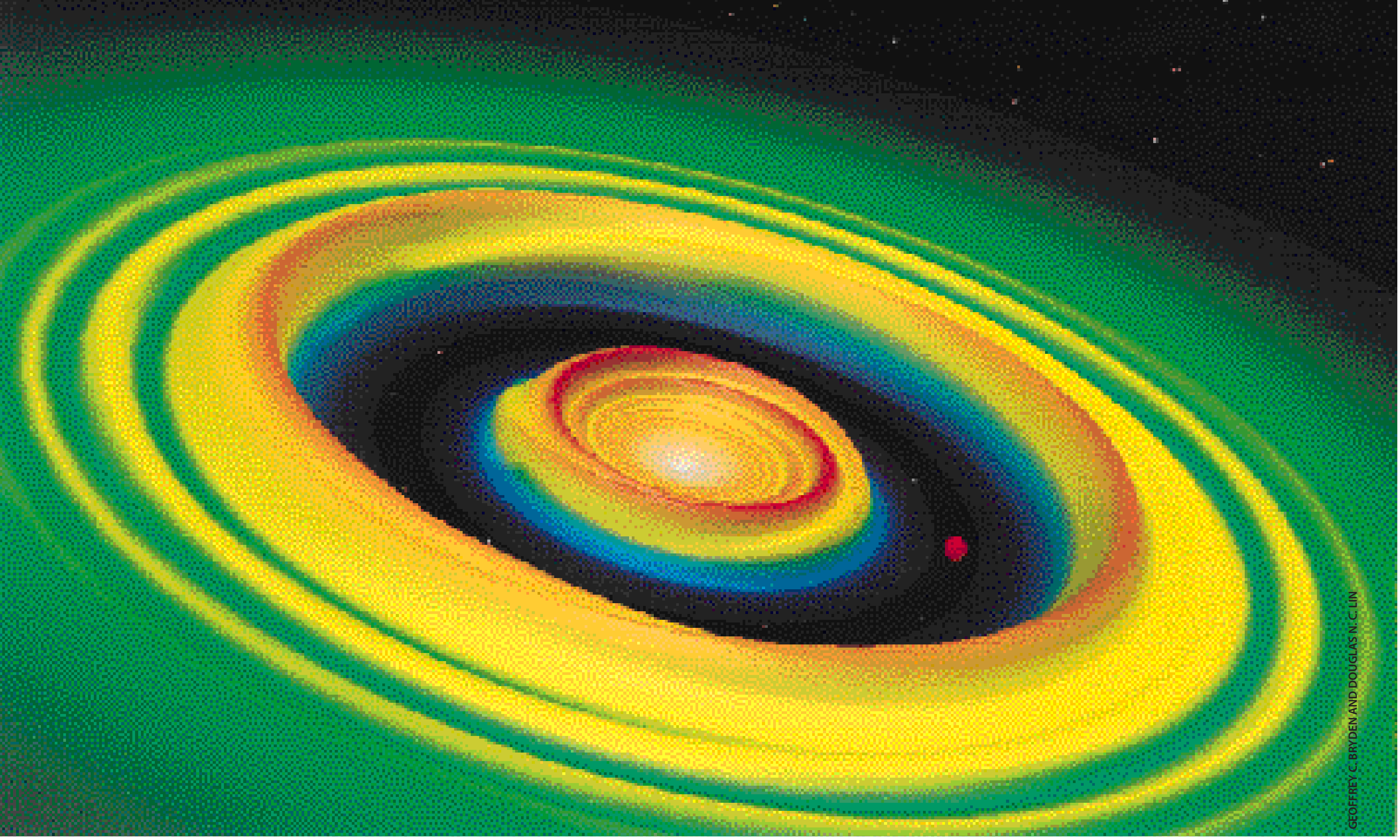
The 51 Peg planets defy conventional planet formation theory, which predicts that giant planets such as Jupiter, Saturn, Uranus or Neptune would form in the cooler outskirts of a protoplanetary disk, at least five times the distance from Earth to the sun.

To account for these planetary oddities, a revised planet formation theory is making the rounds in theorists' circles. Astronomers Douglas N. C. Lin and Peter Bodenheimer, both of the University of California at Santa Cruz, and Derek C. Richardson of the University of Washington extend the standard model by arguing that a young protoplanet precipitating out of a massive protoplanetary disk will carve a groove in the disk, separating it into inner and outer sections. According to their theory, the inner disk dissipates energy because of dynamical friction, causing the disk material and the protoplanet to spiral inward and eventually plunge into the host star.

A planet's salvation stems from the young star's rapid rotation, spinning every five to 10 days. Approaching its star,



PLANETARY OBJECTS ORBITING DISTANT STARS include eight planets, plus HD114762, which—with its large mass—may be a planet or a brown dwarf. These planets show a wide range of orbital distances and eccentricities, which has prompted theorists to revise standard planet-formation theories.



GEOFFREY C. BRYDEN AND DOUGLAS N. C. LIN

JUPITER-MASS PROTOPLANET excites "density waves" in the gas and dust of a planetary disk, as shown in this model by astronomers Douglas N. C. Lin and Geoffrey Bryden of the University of California at Santa Cruz. Those waves, seen as spiral patterns, create regions of high (*red*), medium (*green*) and low (*blue*) density in the disk. The protoplanet accretes gas and dust until its gravity can no longer attract surrounding material. The resulting planetary body ultimately settles into a stable orbit.

a planet would cause tides on the star to rise, just as the moon raises tides on Earth. With the young star rotating faster than the protoplanet orbiting the star, the star would tend to sprout a bulge whose gravity would tug the planet forward. This effect would tend to whip the protoplanet into a larger orbit, halting its deathly inward spiral.

In this model, the protoplanet hangs poised in a stable orbit, delicately balanced between the disk's drag and the rotating star's forward tug. Even before the discovery of the 51 Peg planets, Lin predicted that Jupiter should have spiraled into the sun during its formation. If this were so, then why did Jupiter survive? Perhaps our solar system contained previous "Jupiters" that did indeed spiral into the sun, leaving our Jupiter as the sole survivor.

Why, we wonder, does no large 51 Peg–like planet orbit close to our sun? Perhaps Jupiter formed near the end of our protoplanetary disk's lifetime. Or the protoplanetary disk may have lacked enough gas and dust to exert sufficient tidal drag. Perhaps protoplanetary disks come in a wide range of masses, from a few Jupiter-masses to hundreds of Jupiter-masses. In that case, the diversity of new planets may correspond to different disk masses or disk lifetimes, perhaps even to different environments, including the presence or absence of nearby radiation-emitting stars.

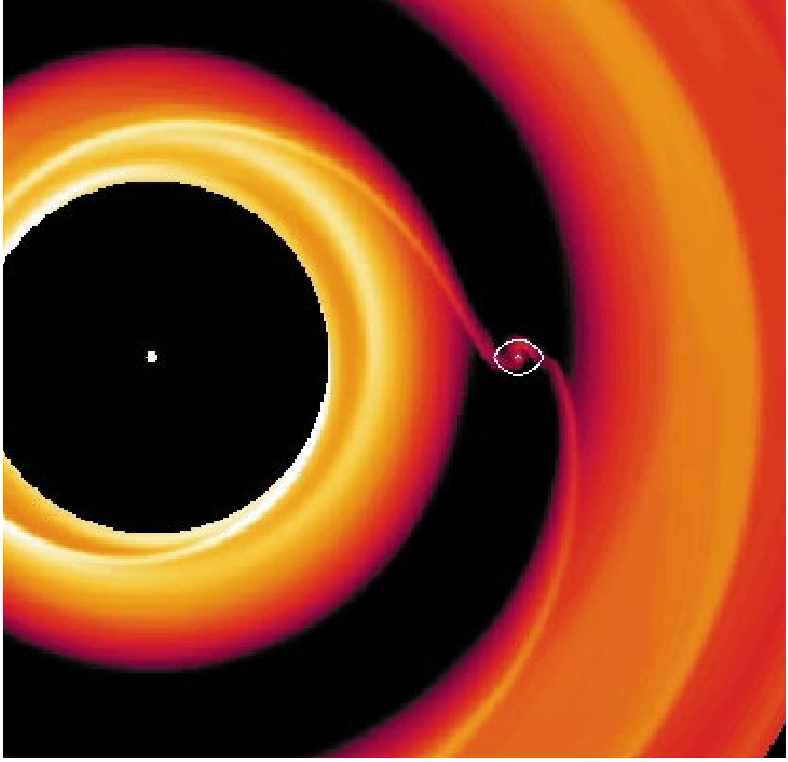
On the other hand, astronomer David F. Gray of the University of Western Ontario in Canada has challenged the existence of the 51 Peg planets altogether. Gray argues that the alleged planet-bearing stars are themselves oscillating—almost like wobbling water balloons. In his view, the cyclical Doppler shifts in these stars stem from inherent stellar wobbles, not planets tugging at stars.

Armed with new data, astronomers now largely dismiss the existence of the oscillations. The strongest argument against the oscillations stems from the single period and frequency seen in the Doppler variations from the star. Most oscillating systems, such as tuning forks, display a set of harmonics, or several different oscillations occurring at different frequencies, rather than just one frequency. But the 51 Peg stars show only one period each, quite unlike harmonic oscillations.

Moreover, ordinary physical models predict that the strongest wobbles would occur at higher frequencies than those of the observed oscillations of these stars. In addition, the 51 Peg stars show no variations in brightness, suggesting that their sizes and shapes are not changing.

Planetary Comparisons

Although we are tempted to compare the eight new planets with our own nine, the comparison is, unfortunately, quite challenging. No one can draw firm conclusions from only eight new planets. So far our ability to spot other types of planets remains limited. At present, our instruments cannot even detect Earth-size companions. Although the extrasolar planets found to date have orbital periods no longer than three years, this finding does not necessarily represent planetary systems in general. Rather it arises from the fact that astronomers have searched for other planets with better techniques for only about a



PAWEL ARTYMOWICZ

PROTOPLANET FORMS in the disk material circling a star, opening up a gap in the gas and dust from which it coalesces. In this model by Pawel Artymowicz of the University of Stockholm and his colleagues, the protoplanet is surrounded by a gravitational field, or Roche lobe, in which raw disk material accumulates, clumping together into a body that is recognizable as a massive planet.

decade. With more time and improved Doppler precision, more planets with longer orbital periods may be found.

Curiously, finding these new planets proves that our own history could easily have played out quite differently. Suppose that gravitational scattering of planets occurs commonly in planetary systems. We see in our own solar system evidence that during its first billion years, planetesimals—fragmentary bodies of rock and ice—hurtled through space. Our cratered moon and Uranus's highly tilted axis—nearly perpendicular to the axes of all its neighbors—show that collisions were common, some involving planet-size objects. The neatly carved orbits of our now stable solar system emerged from the collision-happy orbits of its youth.

We should consider ourselves lucky that Jupiter ended up in a nearly circular orbit. If it had careened into an oval orbit, Jupiter might have scattered Earth, thwacking it out of the solar system. Without stable orbits for Earth and Jupiter, life might never have emerged.

The Future of Planet Hunting

In July 1996 we began a second Doppler survey of 400 stars, using the 10-meter Keck telescope at Mauna Kea Observatory in Hawaii. Mayor and Queloz of Geneva Observatory recently tripled the size of their Northern Hemisphere Doppler survey to about 400 stars, and soon they will begin a Southern Hemisphere survey of 500 more stars. Within the next year, Doppler surveys of several hundred additional stars will begin at the nine-meter Hobby-Eberly Telescope located at McDonald Observatory.

By the year 2000 two Keck telescopes on Mauna Kea and a binocular telescope at the University of Arizona will become optical interferometers, precise enough to image extrasolar planets. NASA plans to launch at least three spaceborne telescopes to detect planets in infrared light.

One proposed NASA space-based interferometer, a second-generation telescope known as the Terrestrial Planet Finder, should obtain pictures of candidate habitable planets orbiting distant stars. Arguably the greatest telescope ever conceived, Planet Finder could spot other Earths, starting in about 2010. Using a spectrometer, it could analyze light from far-off planets to determine the chemical makeup of their atmospheres—data to determine if biological activity is proceeding. This monumental, spaceborne telescope would span a football field and sport four huge mirrors.

Drawing from the data on planets found so far, we believe other planets orbit similar stars, many the size of Jupiter, some the size of Earth. It may be that as many as 10 percent of all stars in our galaxy host planetary companions. Based on this estimate, 10 billion planets would exist in our Milky Way galaxy alone.

Seeking the ideal Earth-like planet on which life could flourish, astronomers will search for planets that are neither too cold nor too hot, temperate enough to sustain liquid water to serve as the mixer and solvent for organic chemistry and biochemistry. Planets with the perfect blend of molecular constituents orbiting at just the right distance from the sun enjoy what astronomers call a "Goldilocks" orbit.

Seeing such a planet would spawn an endless stream of questions: Does its atmosphere contain oxygen, nitrogen, and carbon dioxide, like Earth's, or sulfuric acid and CO_2, the deadly combination on Venus? Is there a protective ozone layer, or is the surface scorched by harmful ultraviolet rays? Even if a planet has oceans, does the water have a pH neutral enough to permit cells to grow?

There may even exist some other biology that thrives on sulfuric acid—even starves without it. Indeed, if primitive life does arise on another Earth, does it always evolve toward intelligence, or is our human technology some fluke of Darwinian luck? Are we humans a rare quirk of nature, destined to appear on Earth-like planets only once in a universe that otherwise teems with primitive life?

Amazing as it seems, answers to some of these questions may arise during our lifetimes, using tools such as telescopes already in existence or on the drawing board. We can only barely imagine what the next generation will see in our reconnaissance of the galactic neighborhood. Human destiny lies in exploring the galaxy and finding our roots, biologically and chemically, out among the stars.

The Authors

GEOFFREY W. MARCY and R. PAUL BUTLER together have found six of the eight planets around sunlike stars reported to date. Marcy is a Distinguished University Professor at San Francisco State University and an adjunct professor at the University of California, Berkeley. Butler is a staff astronomer at the Anglo-Australian Observatory. For more information on extrasolar planets, visit the authors' site (http://cannon.sfsu.edu/~gmarcy/planetsearch/planetsearch.html) on the World Wide Web.

Searching for Life in Our Solar System

If life evolved independently on our neighboring planets or moons, then where are the most likely places to look for evidence of extraterrestrial organisms?

by Bruce M. Jakosky

Since antiquity, human beings have imagined life spread far and wide in the universe. Only recently has science caught up, as we have come to understand the nature of life on Earth and the possibility that life exists elsewhere. Recent discoveries of planets orbiting other stars and of possible fossil evidence in Martian meteorites have gained considerable public acclaim. And the scientific case for life elsewhere has grown stronger during the past decade. There is now a sense that we are verging on the discovery of life on other planets.

To search for life in our solar system, we need to start at home. Because Earth is our only example of a planet endowed with life, we can use it to understand the conditions needed to spawn life elsewhere. As we define these conditions, though, we need to consider whether they are specific to life on Earth or general enough to apply anywhere.

Our geologic record tells us that life on Earth started shortly after life's existence became possible—only after protoplanets (small, planetlike objects) stopped bombarding our planet near the end of its formation. The last "Earth-sterilizing" giant impact probably occurred between 4.4 and 4.0 billion years ago. Fossil microscopic cells and carbon isotopic evidence suggest that life had grown widespread some 3.5 billion years ago and may have existed before 3.85 billion years ago.

Once it became safe for life to exist, no more than half a billion years—and perhaps as little as 100 million to 200 mil-

DENDRITIC VALLEYS ON MARS resemble river drainage systems on Earth, spanning roughly one kilometer across and several hundred meters deep. Occurring primarily on ancient, cratered terrain, the valleys may have formed from atmospheric precipitation or from underground water that flowed onto the surface. Compared with Earth's drainage systems, the Martian valleys show a lower channel density (number of channels per square kilometer), suggesting that on early Mars water was less abundant than it is on Earth.

COURTESY OF BRUCE M. JAKOSKY

lion years—passed before life rooted itself firmly on Earth. This short time span indicates that life's origin followed a relatively straightforward process, the natural consequence of chemical reactions in a geologically active environment. Equally important, this observation tells us that life may originate along similar lines in any place with chemical and environmental conditions akin to those of Earth.

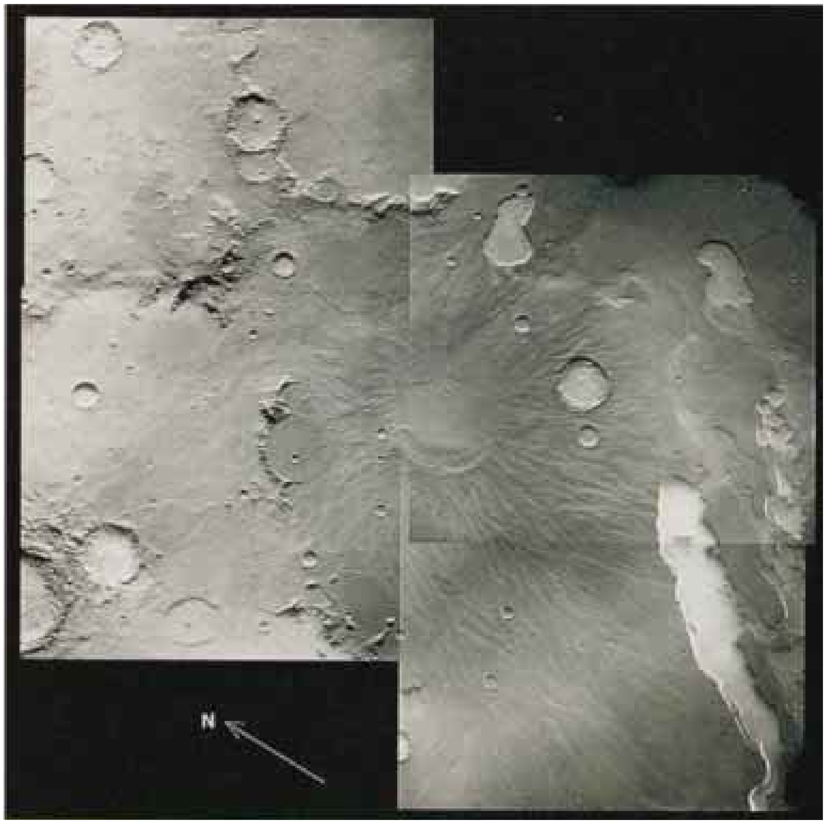
The standard wisdom of the past 40 years holds that prebiological organic molecules formed in a so-called reducing atmosphere, with energy sources such as lightning triggering chemical reactions to combine gaseous molecules. A more recent theory offers a tantalizing alternative. As water circulates through ocean-floor volcanic systems, it heats to temperatures above 400 degrees Celsius (720 degrees Fahrenheit). When that superhot water returns to the ocean, it can chemically reduce agents, facilitating the formation of organic molecules. This reducing environment also provides an energy source to help organic molecules combine into larger structures and to foster primitive metabolic reactions.

Where Did Life Originate?

The significance of hydrothermal systems in life's history appears in the "tree of life," constructed recently from genetic sequences in RNA molecules, which carry forward genetic information. This tree arises from differences in RNA sequences common to all of Earth's living organisms. Organisms evolving little since their separation from their last common ancestor have similar RNA base sequences. Those organisms closest to the "root"—or last common ancestor of all living organisms—are hyperthermophiles, which live in hot water, possibly as high as 115 degrees C. This relationship indicates either that terrestrial life "passed through" hydrothermal systems at some early time or that life's origin took place within such systems. Either way, the earliest history of life reveals an intimate connection to hydrothermal systems.

As we consider possible occurrences of life elsewhere in the solar system, we can generalize environmental conditions required for life to emerge and flourish. We assume that liquid water is necessary—a medium through which primitive organisms can gain nutrients and disperse waste. Although other liquids, such as methane or ammonia, could serve the same function, water is likely to have been much more abundant, as well as chemically better for precipitating reactions necessary to spark biological activity.

To create the building blocks from which life can assemble itself, one needs access to biogenic elements. On Earth, these elements include carbon, hydrogen, oxygen, nitrogen, sulfur and phosphorus, among the two dozen or so others playing a pivotal role in life. Although life elsewhere might not use exactly the same elements, we would expect it to use many of them. Life on Earth utilizes carbon (over silicon, for example) because of its versatility in forming chemical bonds, rather than strictly its abundance. Carbon also exists readily as



CATASTROPHIC OUTFLOW CHANNEL on Mars—Dao Vallis—is on the flanks of the volcano Hadriaca Patera. Scientists believe the volcano's heat may have caused groundwater to well up, or erupt, onto Mars's surface at this location. The possible combination of volcanic energy and water makes this an intriguing place to search for life.

COURTESY OF BRUCE M. JAKOSKY

carbon dioxide, available as a gas or dissolved in water. Silicon dioxide, on the other hand, exists plentifully in neither form and would be much less accessible. Given the ubiquity of carbon-containing organic molecules throughout the universe, we would expect carbon to play a role in life anywhere.

Of course, an energy source must drive chemical disequilibrium, which fosters the reactions necessary to spawn living systems. On Earth today, nearly all of life's energy comes from the sun, through photosynthesis. Yet chemical energy sources suffice—and would be more readily available for early life. These sources would include geochemical energy from hydrothermal systems near volcanoes or chemical energy from the weathering of minerals at or near a planet's surface.

Possibilities for Life on Mars

Looking beyond Earth, two planets show strong evidence for having had environmental conditions suitable to originate life at some time in their history—Mars and Europa. (For this purpose, we will consider Europa, a moon of Jupiter, to be a planetary body.)

Mars today is not very hospitable. Daily average temperatures rarely rise much above 220 kelvins, some 53 kelvins below water's freezing point. Despite this drawback, abundant evidence suggests that liquid water has existed on Mars's surface in the past and probably is present within its crust today.

Networks of dendritic valleys on the oldest Martian surfaces look like those on Earth formed by running water. The water may have come from atmospheric precipitation or "sapping," released from a crustal aquifer. Regardless of where it came from, liquid water undoubtedly played a role. The valleys' dendritic structure indicates that they formed gradually, meaning that water once may have flowed on Mars's surface, although we do not observe such signs today.

In addition, ancient impact craters larger than about 15 kilometers (nine miles) in diameter have degraded heavily, showing no signs of ejecta blankets, the raised rims or central peaks typically present on fresh craters. Some partly eroded craters display gullies on their walls, which look water-carved. Craters smaller than about 15 kilometers have eroded away entirely. The simplest explanation holds that surface water eroded the craters.

Although the history of Mars's atmosphere is obscure, the atmosphere may have been denser during the earliest epochs, 3.5 to 4.0 billion years ago. Correspondingly, a denser atmosphere could have yielded a strong greenhouse effect, which would have warmed the planet enough to permit liquid water to remain stable. Subsequent to 3.5 billion years ago, evidence tells us that the planet's crust did contain much water. Evidently, catastrophic floods, bursting from below the planet's surface, carved out great flood channels. These floods occurred periodically over geologic time. Based on this evidence, liquid water should exist several kilometers underground, where geothermal heating would raise temperatures to the melting point of ice.

Mars also has had rich energy sources throughout time. Volcanism has supplied heat from the earliest epochs to the recent past, as have impact events. Additional energy to sustain life can come from the weathering of volcanic rocks. Oxidation of iron within basalt, for example, releases energy that organisms can use.

The plentiful availability of biogenic elements on Mars's surface completes life's requirements. Given the presence of water and energy, Mars may well have independently originated life. Moreover, even if life did not originate on Mars, life still could be present there. Just as high-velocity impacts have jettisoned Martian surface rocks into space—only to fall on Earth as Martian meteorites—rocks from Earth could similarly have landed on the red planet. Should they contain organisms that survive the journey and should they land in suitable Martian habitats, the bacteria could survive. Or, for all we know, life could have originated on Mars and been transplanted subsequently to Earth.

An inventory of energy available on Mars suggests that enough is present to support life. Whether photosynthesis evolved, and thereby allowed life to move into other ecological niches, remains uncertain. Certainly, data returned from the Viking spacecraft during the 1970s presented no evidence that life is widespread on Mars. Yet it is possible that some Martian life currently exists, cloistered in isolated, energy-rich and water-laden niches—perhaps in volcanically heated, subsurface hydrothermal systems or merely underground, drawing energy from chemical interactions of liquid water and rock.

Recent analysis of Martian meteorites found on Earth has led many scientists to conclude that life may have once thrived on Mars—based on fossil remnants seen within the rock [*see box below*]. Yet this evidence does not definitively indicate biological activity; indeed, it may result from natural geochemical processes. Even if scientists determine that these rocks contain no evidence of Martian life, life on the red planet might still be possible—but in locations not yet searched. To draw a definitive conclusion, we must study those places where life (or evidence of past life) will most likely appear.

Europa

Europa, on the other hand, presents a different possible scenario for life's origin. At first glance, Europa seems an unlikely place for life. The largest of Jupiter's satellites, Europa is a little bit smaller than our moon, and its surface is covered with nearly pure ice. Yet Europa's interior may be less frigid, warmed by a combination of radioactive decay and tidal heating, which could raise the temperature above the melting point of ice at relatively shallow depths. Because the layer of surface ice stands 150 to 300 kilometers thick, a global, ice-covered ocean of liquid water may exist underneath.

Recent images of Europa's surface from the Galileo spacecraft reveal the possible presence of at least transient pockets of liquid water. Globally, the surface appears covered with long grooves or cracks. On a smaller scale, these quasilinear features show detailed structures indicating local ice-related tectonic activity and infilling from below. On the smallest scale, blocks of ice are present. By tracing the crisscrossing grooves, the blocks clearly have moved with respect to the larger mass. They appear similar to sea ice on Earth—as if large ice blocks had broken off the main mass, floated a small distance away and then frozen in place. Unfortunately, we cannot yet determine if the ice blocks floated through liquid water or slid on relatively warm, soft ice. The dearth of impact craters on the ice indicates that fresh ice continually resurfaces Europa. It is also likely that liquid water is present at least on an intermittent basis.

If Europa has liquid water at all, then that water probably exists at the interface between the ice and underlying rocky interior. Europa's rocky center probably has had volcanic activity—perhaps at a level similar to that of Earth's moon, which rumbled with volcanism until about 3.0 billion years ago. The volcanism within its core would create an energy source for possible life, as would the weathering of minerals reacting with water. Thus, Europa has all the ingredients from which to spark life. Of course, less chemical energy is likely to exist on Europa than Mars, so we should not expect to see an abundance of life, if any. Although the Galileo space probe has detected organic molecules and frozen water on Callisto and Ganymede, two of Jupiter's four Galilean satellites, these moons lack the energy sources that life would require to take hold. Only Io, also a Galilean satellite, has volcanic heat—yet

---

Microbial Remnants from Mars?

In 1984, surveying the Far Western Icefield of the Allan Hills Region of Antarctica, geologist Roberta Score plucked from a plain of wind-blasted, bluish, 10,000-year-old ice an unusual greenish-gray rock. Back at the National Aeronautics and Space Administration Johnson Space Center and at Stanford University, researchers confirmed that the 1.9-kilogram (four-pound), potato-size rock—designated ALH84001—was a meteorite from Mars, one with a remarkable history.

Crystallizing 4.5 billion years ago, shortly after Mars's formation, the rock was ejected from the red planet by a powerful impact, which sent it hurtling through space for 16 million years until it landed in Antarctica 13,000 years ago. Geochemists concluded that the rock's distribution of oxygen isotopes, minerals and structural features was consistent with those of five other meteorites identified as coming from Mars.

Lining the walls of fractures within the meteorite are carbonate globules, each a flattened sphere measuring 20 to 250 microns (millionths of meters). The globules appear to have formed in a carbon-dioxide-saturated fluid, possibly water, between 1.3 and 3.6 billion years ago. Within those globules, provocative features vaguely resemble fossilized remnants of ancient Martian microbes.

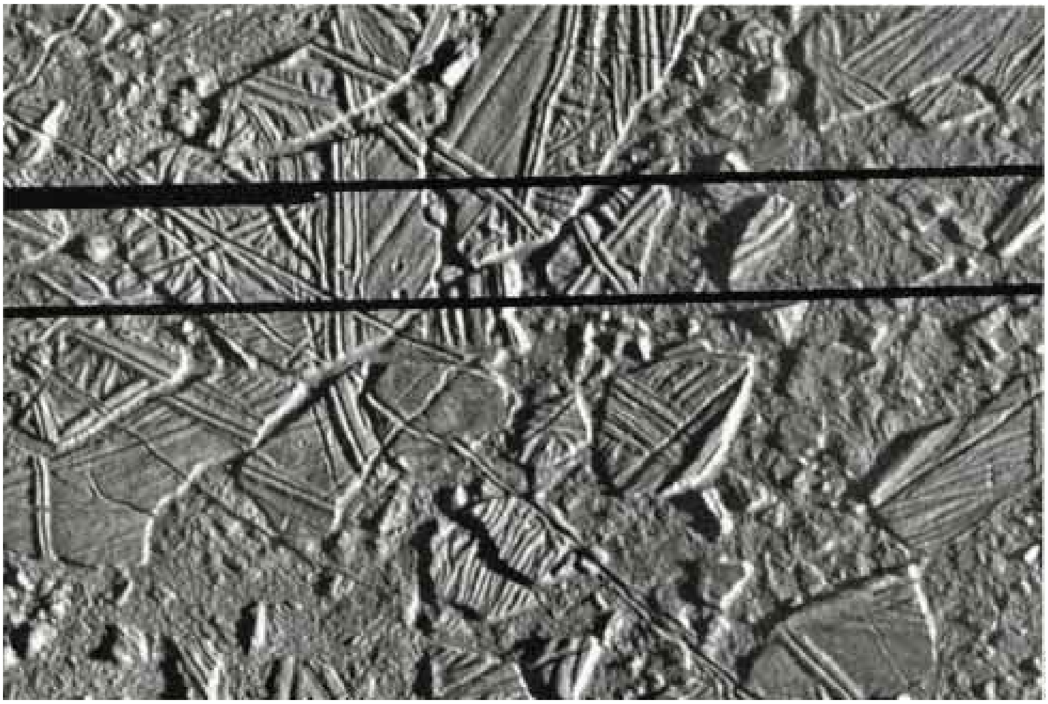
Tiny iron oxide and iron sulfide grains, resembling ones produced by bacteria on Earth, appear in the globules, as do particular polycyclic aromatic hydrocarbons, often found alongside decaying microbes. Other ovoid and tubular structures resemble fossilized terrestrial bacteria themselves. Although the structures range from 30 to 700 nanometers (billionths of meters) in length, some of the most intriguing tubes measure roughly 380 nanometers long—a size nearing the low end of that for terrestrial bacteria, which are typically one to 10 microns long. The tubes' size and shape indicate they may be fossilized pieces of bacteria, or tinier "nanobacteria," which on Earth measure 20 to 400 nanometers long.

These findings collectively led NASA scientists Everett K. Gibson, David S. McKay and their colleagues to announce in August 1996 that microbes might once have flourished on the red planet. Recent chemical analyses reveal, however, that ALH84001 is heavily contaminated with amino acids from Antarctic ice, a result that weakens the case for microfossils from Mars. —*Richard Lipkin*

NASA JOHNSON SPACE CENTER

JOHN W. VALLEY *University of Wisconsin–Madison*

CARBONATE GLOBULE (*right*), about 200 microns long, seemingly formed in the Martian meteorite ALH84001. In the globule, a segmented object (*left*), some 380 nanometers long, vaguely resembles fossilized bacteria from Earth.



EUROPA'S SURFACE is lined with features that suggest "ice tectonics." Blocks of ice appear to have broken up and shifted, perhaps sliding on slush or possibly even floating on liquid water. Either way, spectral analysis of reflected light indicates nearly pure water ice on Europa's surface. The horizontal black bars through the image designate data lost during interplanetary transmission.

JPL/NASA

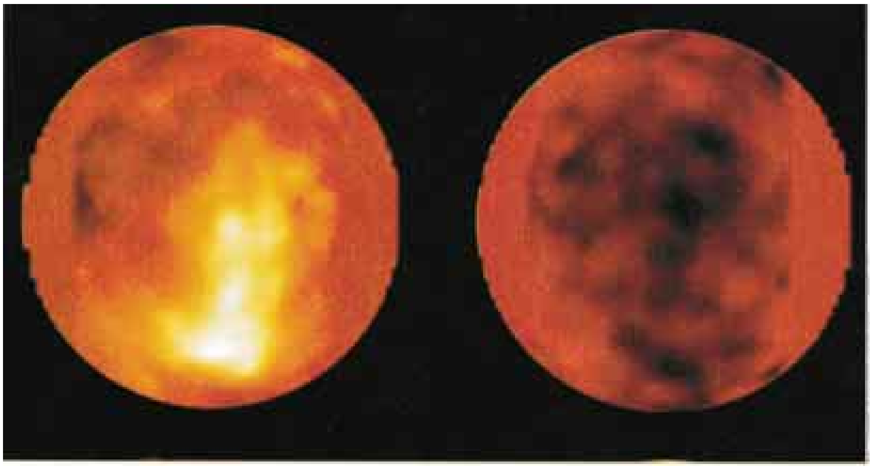
it has no liquid water, necessary to sustain life as we know it.

Mars and Europa stand today as the only places in our solar system that we can identify as having (or having had) all ingredients necessary to spawn life. Yet they are not the only planetary bodies in our solar system relevant to exobiology. In particular, we can look at Venus and at Titan, Saturn's largest moon. Venus currently remains too hot to sustain life, with scorching surface temperatures around 750 kelvins, sustained by greenhouse warming from carbon dioxide and sulfur dioxide gases. Any liquid water has long since disappeared into space.

Venus and Titan

Why are Venus and Earth so different? If Earth orbited the sun at the same distance that Venus does, then Earth, too, would blister with heat—causing more water vapor to fill the atmosphere and augmenting the greenhouse effect. Positive feedback would spur this cycle, with more water, greater greenhouse warming and so on saturating the atmosphere and sending temperatures soaring. Because temperature plays such a strong role in determining the atmosphere's water content, both Earth and Venus have a temperature threshold, above which the positive feedback of an increasing greenhouse effect takes off. This feedback loop would load Venus's atmosphere with water, which in turn would catapult its temperatures to very high values. Below this threshold, its climate would have been more like that of Earth.

Venus, though, may not always have been so inhospitable. Four billion years ago the sun emitted about 30 percent less energy than it does today. With less sunlight, the boundary between clement and runaway climates may have been inside Venus's orbit, and Venus may have had surface temperatures only 100 degrees C above Earth's current temperature. Life could survive quite readily at those temperatures—as we observe with certain bacteria and bioorganisms living near hot springs and undersea vents. As the sun became hotter, Venus would have warmed gradually until it would have undergone a catastrophic transition to a thick, hot atmosphere. It is possible that Venus originated life several billion years ago but that high temperatures and geologic activity have since obliterated all evidence of a biosphere. As the sun continues to heat up, Earth may undergo a similar catastrophic transition only a couple of billion years from now.



TITAN'S BLOTCHED SURFACE suggests that it is not uniformly coated with an ocean of methane and ethane, as scientists once thought. Instead a patchwork of lakes and solid regions may cover its surface. Enveloping the moon are thick clouds, rich in organic aerosols caused by atmospheric reactions. Scientists often compare Titan to the early Earth, before life began.

SPACE TELESCOPE SCIENCE INSTITUTE

Titan intrigues us because of abundant evidence of organic chemical activity in its atmosphere, similar to what might have occurred on the early Earth if its atmosphere had potent abilities to reduce chemical agents. Titan is about as big as Mercury, with an atmosphere thicker than Earth's, consisting predominantly of nitrogen, methane and ethane. Methane must be continually resupplied from the surface or subsurface, because photochemical reactions in the atmosphere drive off hydrogen (which is lost to space) and convert the methane to longer chains of organic molecules. These longer-chain hydrocarbons are thought to provide the dense haze that obscures Titan's surface at visible wavelengths.


EMORY KRISTOF *National Geographic Society*

MINERAL CHIMNEY
near an undersea hydrothermal vent is located off Mexico's west coast at the East Pacific Rise of the Galápagos Rift. More than two kilometers below the sea surface along this midocean ridge, mineral-rich water, up to 757 degrees Celsius, spews from volcanically heated seafloor vents, which sprout mineral chimneys six to nine meters tall. Unusual life-forms, including tiny, white alvinellid worms and heat-tolerant bacteria, thrive in this seemingly hostile environment. Some scientists believe such hydrothermal vents fostered the origin of life on Earth.

Surface temperatures on Titan stand around 94 kelvins, too cold to sustain either liquid water or nonphotochemical reactions that could produce biological activity—although Titan apparently had some liquid water during its early history. Impacts during its formation would have deposited enough heat (from the kinetic energy of the object) to melt frozen water locally. Deposits of liquid water might have persisted for thousands of years before freezing. Every part of Titan's surface probably has melted at least once. The degree to which biochemical reactions may have proceeded during such a short time interval is uncertain, however.

Exploratory Missions

Clearly, the key ingredients needed for life have been present in our solar system for a long time and may be present today outside of Earth. At one time or another, four planetary bodies may have contained the necessary conditions to generate life.

We can determine life's actual existence elsewhere only empirically, and the search for life has taken center stage in the National Aeronautics and Space Administration's ongoing science missions. The Mars Surveyor series of missions, scheduled to take place during the coming decade, aims to determine if Mars ever had life. This series will culminate in a mission currently scheduled for launch in 2005, to collect Martian rocks from regions of possible biological relevance and return them to Earth for detailed analysis. The Cassini spacecraft currently is en route to Saturn. There the Huygens probe will enter Titan's atmosphere, its goal to decipher Titan's composition and chemistry. A radar instrument, too, will map Titan's surface, looking both for geologic clues to its history and evidence of exposed lakes or oceans of methane and ethane.

Moreover, the Galileo orbiter of Jupiter is focusing its extended mission on studying the surface and interior of Europa. Plans are under way to launch a spacecraft mission dedicated to Europa, to discern its geologic and geochemical history and to determine if a global ocean lies underneath its icy shell.

Of course, it is possible that, as we plumb the depths of our own solar system, no evidence of life will turn up. If life assembles itself from basic building blocks as easily as we believe it does, then life should turn up elsewhere. Indeed, life's absence would lead us to question our understanding of life's origin here on Earth. Whether or not we find life, we will gain a tremendous insight into our own history and whether life is rare or widespread in our galaxy.


SPACE TELESCOPE SCIENCE INSTITUTE

The Author

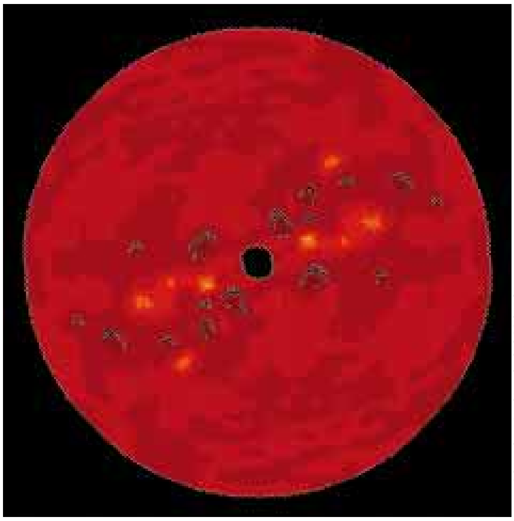
BRUCE M. JAKOSKY is professor of geology and a member of the Laboratory for Atmospheric and Space Physics at the University of Colorado at Boulder. He is also an investigator on the Mars Global Surveyor mission currently in orbit around Mars. His book *The Search for Life on Other Planets* will be published in the summer of 1998 by Cambridge University Press.

Searching for Life in Other Solar Systems

Life remains a phenomenon we know only on Earth. But an innovative telescope in space could change that by detecting signs of life on planets orbiting other stars

by Roger Angel and Neville J. Woolf

ALFRED T. KAMAJIAN



UNIVERSITY OF ARIZONA OASES PROJECT

IMAGE OF DISTANT PLANETS, created from simulated interferometer signals, indicates what astronomers might reasonably expect to see with a space-based telescope. This study displays a system about 30 light-years away, with four planets roughly equivalent in luminosity to Earth. (Each planet appears twice, mirrored across the star.) With this sensitivity, the authors speculate that the instrument could easily examine the planet found in 1996 orbiting 47 Ursae Majoris.

The search for extraterrestrial life can now be extended to planets outside our solar system. After years of looking, astronomers have turned up evidence of giant planets orbiting several distant stars similar to our sun. Smaller planets around these and other stars may have evolved living organisms. Finding extraterrestrial life may seem a Herculean task, but a space telescope mission called the Terrestrial Planet Finder, which the National Aeronautics and Space Administration plans to start in 2005, aims to locate such planets and search for evidence of life-forms, such as the primitive ones on Earth.

The largest and most powerful telescope now in space, the Hubble Space Telescope, can just make out mountains on Mars at 30 kilometers (19 miles). Pictures sharp enough to display geologic features of planets around other stars would require an array of space telescopes the size of the U.S. But pictures of Earth do not reveal the presence of life unless they are taken at very high resolution. Such images could be obtained with unmanned spacecraft sent to other solar systems, but the huge distance between Earth and any other planet makes this approach impractical.

Taking photographs, however, is not the best way to study distant planets. Spectroscopy, the technique astronomers use to obtain information about stars, can also reveal much about planets. In spectroscopy, light originating from an object in space is analyzed for unique markers that help researchers piece together characteristics such as the celestial body's temperature, atmospheric pressure and chemical composition. Simple life-forms on our planet have profoundly altered conditions on Earth in ways that a distant observer could perceive by spectroscopy of the planet atmosphere.

Fossil records indicate that within a billion years of Earth's formation, as soon as heavy bombardment by asteroids ceased, primitive organisms such as bacteria and algae evolved and spread around the globe. These organisms represented the totality of life here for the next two billion years; consequently, if life exists on other planets, it might well be in this highly uncommunicative form.

Earth's humble blue-green algae do not operate radio transmitters. Yet they are chemical engineers, honed by evolution, operating on a huge scale. As algae became more widespread, they began adding large quantities of oxygen to the atmosphere. The production of oxygen, fueled by energy derived from sunlight, is fundamental to carbon-based life: the simplest organisms take in water, nitrogen and carbon dioxide as nutrients and then release oxygen into the atmosphere as waste. Oxygen is a chemically reactive gas; without continued replenishment by algae and, later in Earth's evolution, by plants, its concentration would fall. Thus, the presence of large amounts of oxygen in a planet's atmosphere is a good indicator that some form of carbon-based life may exist there.

SPACE-BASED TELESCOPE SYSTEM that can search for life-bearing planets has been proposed by the authors. The instrument, a type of interferometer, could be assembled at the proposed international space station (*lower left*). Subsequently, electric propulsion would send the 50- to 75-meter-long device into an orbit around the sun roughly the same as Jupiter's. Such a mission is at the focus of the National Aeronautics and Space Administration's plans to study neighboring planetary systems.

In 1993 the Galileo space probe detected oxygen's distinctive spectrum in the red region of visible light from Earth. Indeed, this observation tells us that for a billion years—since plant and animal life has flourished on Earth—a signal of life's presence has radiated into space. The clincher that reveals life processes are occurring on Earth is the simultaneous presence in the planet's spectrum of methane, which is unstable around oxygen but which life continuously replenishes.

What constitutes detection of distant life? Some scientists hold that because life elsewhere is improbable, proof of its detection requires strong evidence. It seems likely, though, that life on other planets would have a carbon-based chemistry similar to our own. Carbon is particularly suitable as a building block of life: it is abundant in the universe, and no other known element can form the myriad of complex but stable molecules necessary for life as we know it. We believe that if a planet looks like Earth and has liquid water and oxygen (evident as ozone), then this would present strong evidence for its having life. If such a planet were found, subsequent investigations could strengthen the case by searching for the more elusive spectral observation of methane.

Of course, there could be some nonbiological oxygen source on a lifeless planet, a possibility that must be considered. Conversely, life could arise from some other type of chemistry that does not generate oxygen. Yet we still should be able to detect any stirrings from chemical residues.

Searching for Another Earth

Planets similar to Earth in size and distance from their sun—ones likely to have oceans of water—represent the most plausible homes for carbon-based life in other solar systems. Water provides a solvent for life's biochemical reactions and serves as a source of needed hydrogen. If each star has planets spanning a range of orbital distances, as occurs in our solar system, then one of those planets is likely to orbit at the right distance to sustain liquid water—even if the star shines more or less brightly than the sun.

Temperature, though, means little if a planet's gravitational pull cannot hold on to oceans and an atmosphere. If distance from a star were the only factor to consider, Earth's moon would have liquid water. But gravity depends on the size and density of the body. Because the moon is smaller and less dense than Earth, its gravitational pull is much weaker. Any water or layers of atmosphere that might develop on or around such a body would quickly be lost to space.

Clearly, we need a technique to reveal characteristics as specific as what chemicals can be found on a planet. Previously we mentioned that the visible radiation coming from a planet can confirm the presence of certain molecules, in particular oxygen, that are known to support life. But distinguishing faint oxygen signals in light reflected by a small planet orbiting even a nearby star is extraordinarily difficult.

A larger version of the Hubble Space Telescope, specially equipped for extremely accurate optical correction, possibly could spot Earth-like planets if they are orbiting the three nearest sunlike stars and search them spectroscopically for oxygen. A more robust method for sampling dozens of stars is needed.

Faced with this quandary, in 1986 we proposed, along with Andrew Y. S. Cheng, now at the University of Hong Kong, that midinfrared wavelengths would serve as the best spectral region in which to find planets and to search for extraterrestrial life. This type of radiation—really the planet's radiated heat—has a wavelength 10 to 20 times longer than that of visible light. At these wavelengths, a planet emits about 40 times as many photons—particles of light—as it does at shorter wavelengths. The nearby star would outshine the planet "only" 10 million times, a ratio 1,000 times more favorable than that which red light offers.

Moreover, three key compounds that we would expect to find on inhabited planets—ozone (a form of oxygen usually located high in the atmosphere), carbon dioxide and water—leave strong imprints in a planet's infrared spectrum. Once again, our solar system provides promising support for this technique: a survey of the infrared emissions of local planets reveals that only Earth displays the infrared signature of life. Although Earth, Mars and Venus all have atmospheres with carbon dioxide, only Earth shows the signature of plentiful water and ozone. Sensitively indicating oxygen, ozone would have appeared on Earth a billion years before oxygen's infrared spectral feature grew detectable.

What kind of telescope do we need to locate Earth-like planets and pick up their infrared emissions? Some of today's ground-based telescopes can detect strong infrared radiation emanating from stars. But the telescope's own heat plus atmospheric absorptions would swamp any sign of a planet. Obviously, we reasoned, we must move the telescope into space.

Even then, to distinguish a planet's radiation from that of its star, a traditional telescope must be much larger than any ground-based or orbiting telescope built to date. Because light cannot be focused to a spot smaller than its wavelength, even a perfect telescope cannot form ideal images. At best, light will focus to a fuzzy core surrounded by a faint halo. If the halo surrounding the star extends beyond the planet's orbit, then we cannot discern the much dimmer body of the planet inside it. By making a telescope mirror and the resulting image very large, we can, in principle, make the image of a star as sharp as desired.

Because we can predict a telescope's performance, we know in advance what kind of image quality to expect. For example, to monitor the infrared spectrum of an Earth-like planet circling, say, a star 30 light-years away, we need a supergiant space telescope, close to 60 meters in diameter. We have made recent steps toward the technology for such telescopes, but 60 meters remains far beyond reach.

Rethinking the Telescope

We knew that to develop a more compact telescope to locate small, perhaps habitable, planets would require some tricks. Twenty-three years ago Ronald N. Bracewell of Stanford University suggested a good strategy when he showed how two small telescopes could together search for large, cool planets similar to Jupiter. Bracewell's proposed instrument consisted of two one-meter telescopes separated by 20 meters. Each telescope alone yields blurred pictures, yet together the two could discern distant worlds.

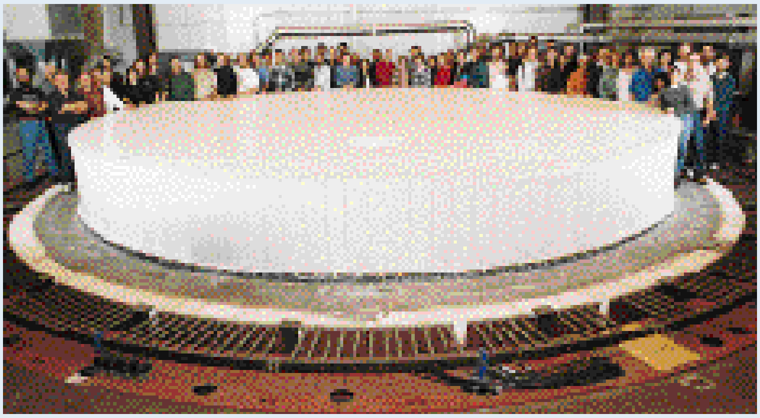
With both telescopes focused on the same star, Bracewell saw that he could invert light waves from one telescope (flipping peaks into troughs), then merge that inverted light with light from the second telescope. With precisely overlapping im-

Building an Earth-Based Interferometer

A consortium of American, Italian and German astronomers is now building a ground-based interferometer on Mount Graham in Arizona. At the Mirror Lab on the University of Arizona campus, where one of us (Angel) works, technicians have cast the first of two 8.4-meter-diameter mirrors (*right*), the largest ever made. Mounted side by side in the Large Binocular Telescope, two such mirrors will serve as a Bracewell interferometer, measuring heat emitted around nearby stars potentially hosting Earth-like planets.

Deformable secondary mirrors will correct for atmospheric blurring. This system is sensitive enough to detect giant planets and dust clouds around stars but not enough to spot another Earth-like planet. Designing a superior space-based interferometer depends on critical dust measurements. If dust clouds around other stars prove much denser than the cloud around the sun, then placing a Terrestrial Planet Finder instrument far from the sun (to avoid local heat from interplanetary dust) will offer no advantage. Instead an interferometer with larger mirrors that is closer to Earth will be needed.

—R.A. and N.J.W.

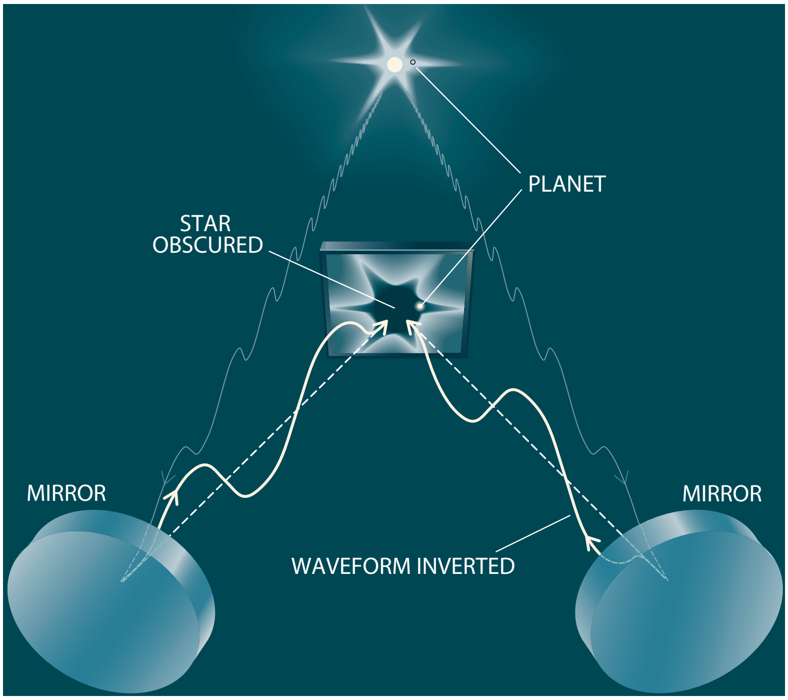

JOHN FLORENCE

GIANT MIRROR at the University of Arizona is to be mounted in the Large Binocular Telescope.

ages, the star's light—from its core and surrounding halo—would cancel out. Yet the planet's signal, which emanates from a slightly different direction, would remain intact. Scientists refer to this type of instrument as an interferometer because it reveals details about a light source by employing interference of light waves.

Bracewell's envisioned telescope would have enough sensitivity to spot Jupiter-size planets, although Earth-size planets would still be too faint to detect. To see Earth-size planets, an interferometer must cancel starlight more completely. In 1990, however, one of us (Angel) showed that such precision becomes possible if more than two telescopes are involved.

Another problem—even after canceling starlight completely—stems from background heat radiated from our solar system's cloud of dust particles, referred to as the zodiacal glow. As Bracewell realized, this glow would nearly overwhelm the signal of a giant planet, let alone that of an Earth-size one. Alain Léger and his collaborators at the University of Paris proposed the practical solution of placing the device in orbit around the sun, at roughly Jupiter's distance, where the dust is so cold that its background thermal radiation is negligible. He showed that an orbiting interferometer at that distance with telescopes as small as one meter in diameter would be sensitive enough to detect an Earth-size planet. Only if the star under study has its own thick dust cloud would detection be obscured, a difficulty that can be assessed with ground-based observations [*see box on opposite page*].



CANCELING STARLIGHT enables astronomers to see dim planets typically obscured by stellar radiance. Two telescopes focused on the same star (*top*) can cancel out much of its light: one telescope inverts the light—making peaks into troughs and vice versa (*right*). When the inverted light is combined with the noninverted starlight from the second telescope (*left*), the light waves interfere with one another, and the image of the star then vanishes (*center*).

Space-Based Interferometer

In 1995 NASA selected three teams to investigate various methods for discovering planets around other stars. We assembled an international team that included Bracewell, Léger and his colleague Jean-Marie Mariotti of the Paris Observatory, as well as some 20 other scientists and engineers. The two of us at the University of Arizona studied the potential of a new approach, an interferometer with two pairs of mirrors all arranged in a straight line.

Because this interferometer cancels starlight very effectively, it could span about 75 meters, a size offering important advantages. It permits astronomers to reconstruct actual images of planets orbiting a star, as well as to observe stars over a wide range of distances without expanding or contracting the device. As we envision the orbiting interferometer, it could point to a different star every day while returning to interesting systems for more observations.

If pointed at our own solar system from a nearby star, the interferometer could pick out Venus, Earth, Mars, Jupiter and Saturn. Its data could be analyzed to find the chemical composition of each planet's atmosphere. The device could easily study the newly discovered planet around 47 Ursae Majoris. More important, this interferometer could identify Earth-like planets that otherwise elude us, checking such planets for the presence of carbon dioxide, water and ozone—perhaps even methane.

Thanks to new ultralightweight mirrors developed for NASA's Next Generation Space Telescope, a space-based interferometer combining telescopes as large as six meters in diameter looks feasible. Such an interferometer would suffer less from background heat and would function effectively in a near-Earth orbit. Also, it could better handle emissions from dust clouds around nearby candidate stars, if these clouds prove denser than those around the sun.

Building the interferometer would be a substantial undertaking, perhaps an international project, and many of the details have yet to be worked out. NASA has challenged designers of the Terrestrial Planet Finder to keep construction and launch costs below $500 million. A first industrial analysis indicates the price tag is not unrealistic.

The discovery of life on another planet may arguably be the crowning achievement of the exploration of space. Finding life elsewhere, NASA administrator Daniel S. Goldin has said, "would change everything—no human endeavor or thought would be unchanged by that discovery." SA

The Authors

ROGER ANGEL and NEVILLE J. WOOLF have collaborated for 15 years on methods for making better telescopes. They are based at Steward Observatory at the University of Arizona. A fellow of the Royal Society, Angel directs the Steward Observatory Mirror Laboratory. Woolf has pioneered techniques to minimize the distortion of images caused by the atmosphere. Angel and Woolf consider the quest for distant planets to be the ultimate test for telescope builders; they are meeting this challenge by pushing the limits of outer-space observation technology, such as adaptive optics and space telescopes. This article updates a version that appeared in *Scientific American* in April 1996.

Planetary Tour

Some four and a half billion years ago, and for reasons that scientists have yet to agree upon, a flat, round cloud of gas and dust began to contract in the interstellar space of our Milky Way galaxy, itself already at least five billion years old. As this cloud collapsed toward its center, its relatively small initial rate of spin increased. This spinning, in turn, hurled agglomerations of dust outward, enabling them to resist the gravitational pull of a massive nebula at the center of the cloud.

As this giant central nebula—the precursor of our sun—collapsed in on itself, the temperature at its center soared. Eventually, the heat and pressure were enough to ignite the thermonuclear furnace that would make life possible and that will probably burn for another five billion years.

Over tens of millions of years, the agglomerations of dust surrounding the protosun became the nine planets, 63 moons, and myriad asteroids and comets of our solar system. One of the many unsolved puzzles about the formation of the solar system concerns the arrangement of these planets—specifically, why the first four are small and rocky, and the next four are giant and gaseous. A leading theory—that early, powerful solar flares blew the lighter elements out of the inner solar system—has been challenged by the discovery of gas giant–type planets orbiting very close to sunlike stars in the Milky Way.

In the pages that follow, Scientific American conducts a guided tour of the solar system. Its purpose, in this issue devoted to the grandeur and complexity of the cosmos, is to reassert the wonders that exist in our own infinitesimal corner of it.

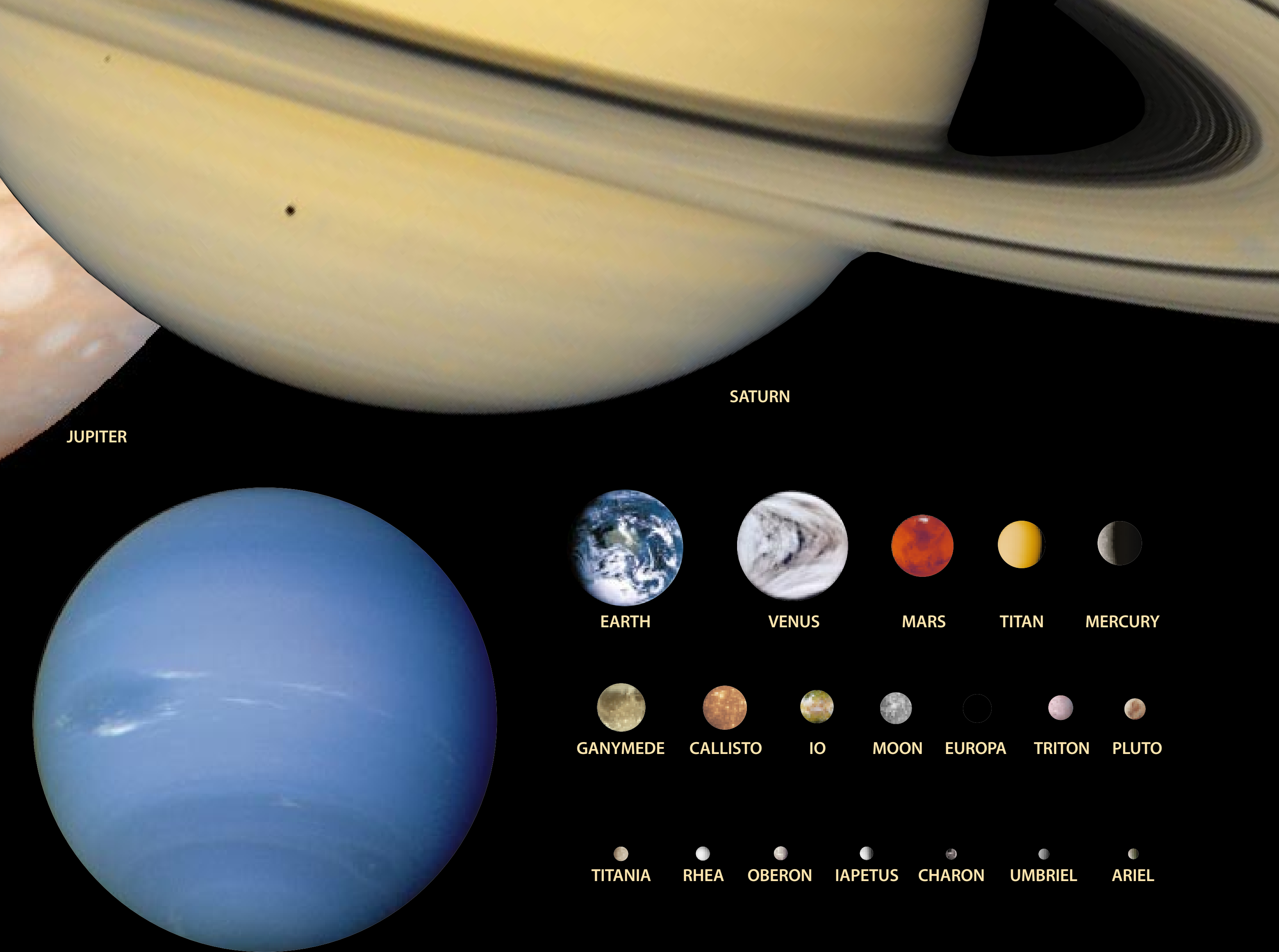
—*The Editors*

URANUS

The planets at a glance

	MERCURY	VENUS	EARTH	MARS
AVERAGE DISTANCE FROM SUN (kilometers)	57.9 million	108.2 million	149.6 million	227.94 million
EQUATORIAL DIAMETER (kilometers)	4,878	12,100	12,756.34	6,786
MASS (kilograms)	3.3×10^{23}	4.9×10^{24}	6.0×10^{24}	6.4×10^{23}
DENSITY (grams per cubic centimeter)	5.41	5.25	5.52	3.9
LENGTH OF DAY (relative to Earth)	58.6 days	243.0 days	23.93 hours	24.62 hours
LENGTH OF YEAR (relative to Earth)	87.97 days	224.7 days	365.26 days	686.98 days
NUMBER OF KNOWN MOONS	0	0	1	2
ATMOSPHERIC COMPOSITION	Negligible traces of sodium, helium, hydrogen and oxygen	96% carbon dioxide, 3.5% nitrogen	78% nitrogen, 21% oxygen, 0.9% argon	95% carbon dioxide, 3% nitrogen, 1.6% argon

JPL/CALTECH/NASA (*all images*); LAURIE GRACE (*table*)



The relative sizes of the largest bodies in the solar system

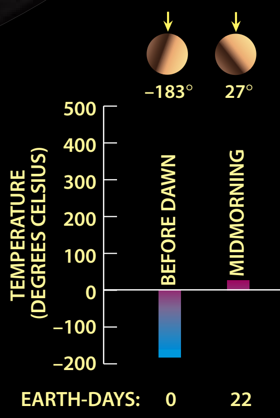
JUPITER	SATURN	URANUS	NEPTUNE	PLUTO
778.4 million	1,423.6 million	2,867.0 million	4,488.4 million	5,909.6 million
142,984	120,536	51,108	49,538	2,350
1.9×10^{27}	5.7×10^{26}	8.7×10^{25}	1.0×10^{26}	1.3×10^{22}
1.3	0.7	1.3	1.7	1.99
9.8 hours	10.2 hours	17.9 hours	19.1 hours	6.39 days
11.86 years	29.46 years	84 years	164.8 years	247.7 years
16	At least 19	17	8	1
90% hydrogen, 10% helium, traces of methane	97% hydrogen, 3% helium, traces of methane	83% hydrogen, 15% helium, 2% methane	74% hydrogen, 25% helium, 2% methane	Probably methane, possibly nitrogen and carbon monoxide

Mercury

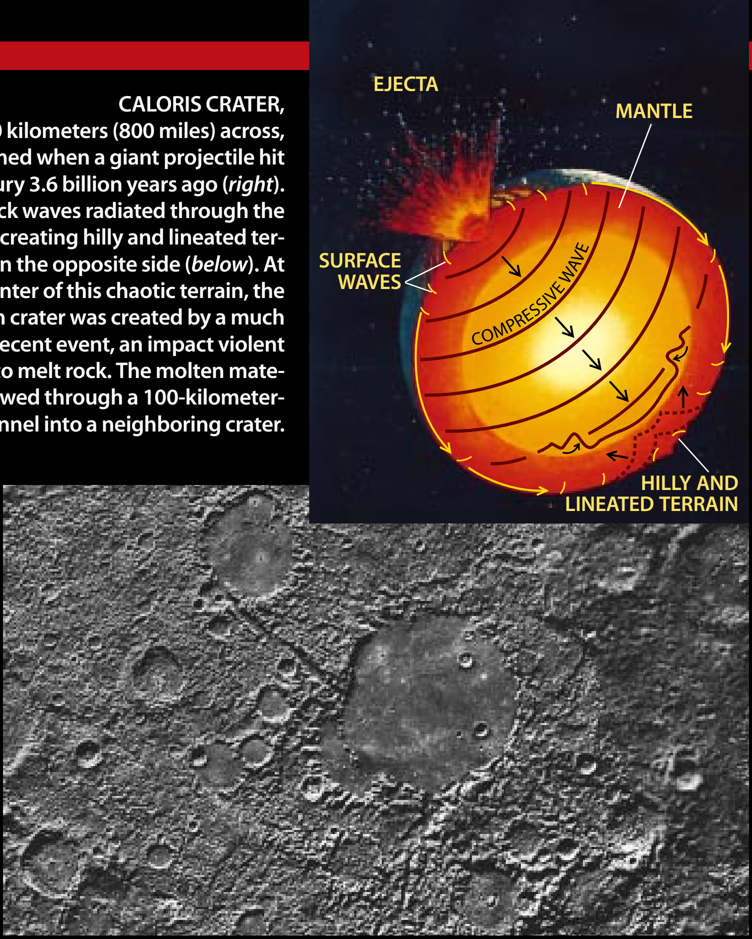
ASTROGEOLOGY TEAM, U.S. GEOLOGICAL SURVEY, FLAGSTAFF, ARIZ. *(middle)*; NASA *(bottom left)*; BRYAN CHRISTIE *(illustration)*

SIZE COMPARED WITH EARTH

MERCURIAN DAYTIME TEMPERATURE ranges above 400 degrees Celsius (750 degrees Fahrenheit)—and, at night, plummets to almost –200 degrees C. The high temperatures preclude the existence of a significant atmosphere, because gas molecules move faster than the planet's escape velocity.



CALORIS CRATER, 1,300 kilometers (800 miles) across, was formed when a giant projectile hit Mercury 3.6 billion years ago (*right*). Shock waves radiated through the planet, creating hilly and lineated terrain on the opposite side (*below*). At the center of this chaotic terrain, the Petrarch crater was created by a much more recent event, an impact violent enough to melt rock. The molten material flowed through a 100-kilometer-long channel into a neighboring crater.



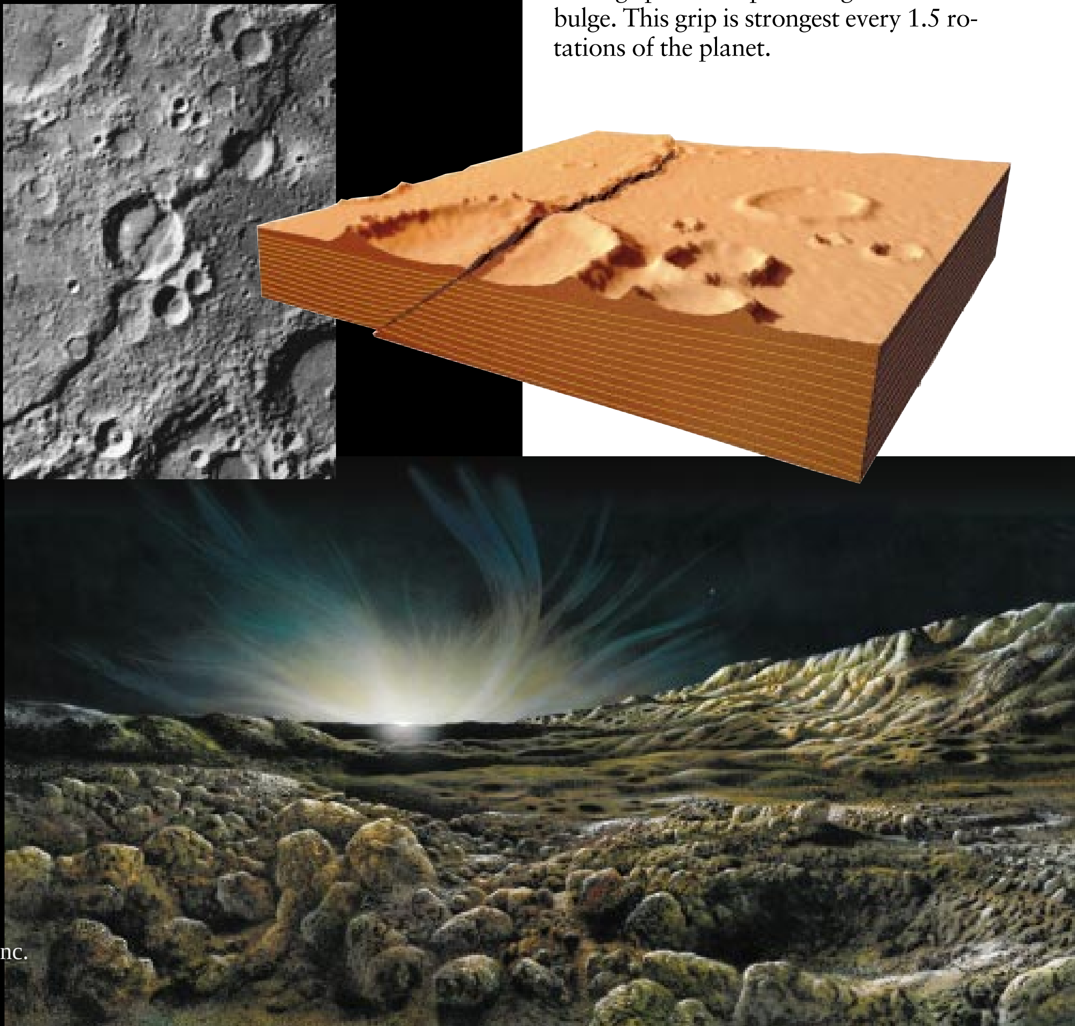
ALFRED T. KAMAJIAN, COURTESY OF P. H. SCHULTZ AND D. E. GAULT (*top*); NASA (*upper middle and lower middle left*); NASA AND SLIM FILMS (*lower middle right*); SLIM FILMS (*bottom left*); DON DIXON (*bottom right*)

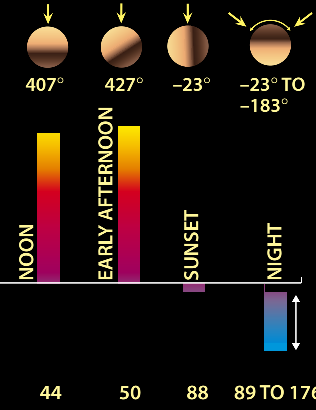
The innermost planet in the solar system, Mercury has the most extreme characteristics of the terrestrial bodies. Daytime temperatures on the planet reach 427 degrees Celsius (801 degrees Fahrenheit)—hot enough to melt zinc. At night, however, the lack of an atmosphere lets the temperature plunge to –183 degrees C, which is cold enough to freeze krypton.

Mercury is also unusually dense. To account for its density of 5.44 grams per cubic centimeter (0.20 pound per cubic inch), astronomers believe the planet must have a relatively huge core that is unusually iron-rich. The core probably takes up 42 percent of Mercury's volume; in comparison, Earth's core is only about 16 percent, and Mars's, about 9 percent.

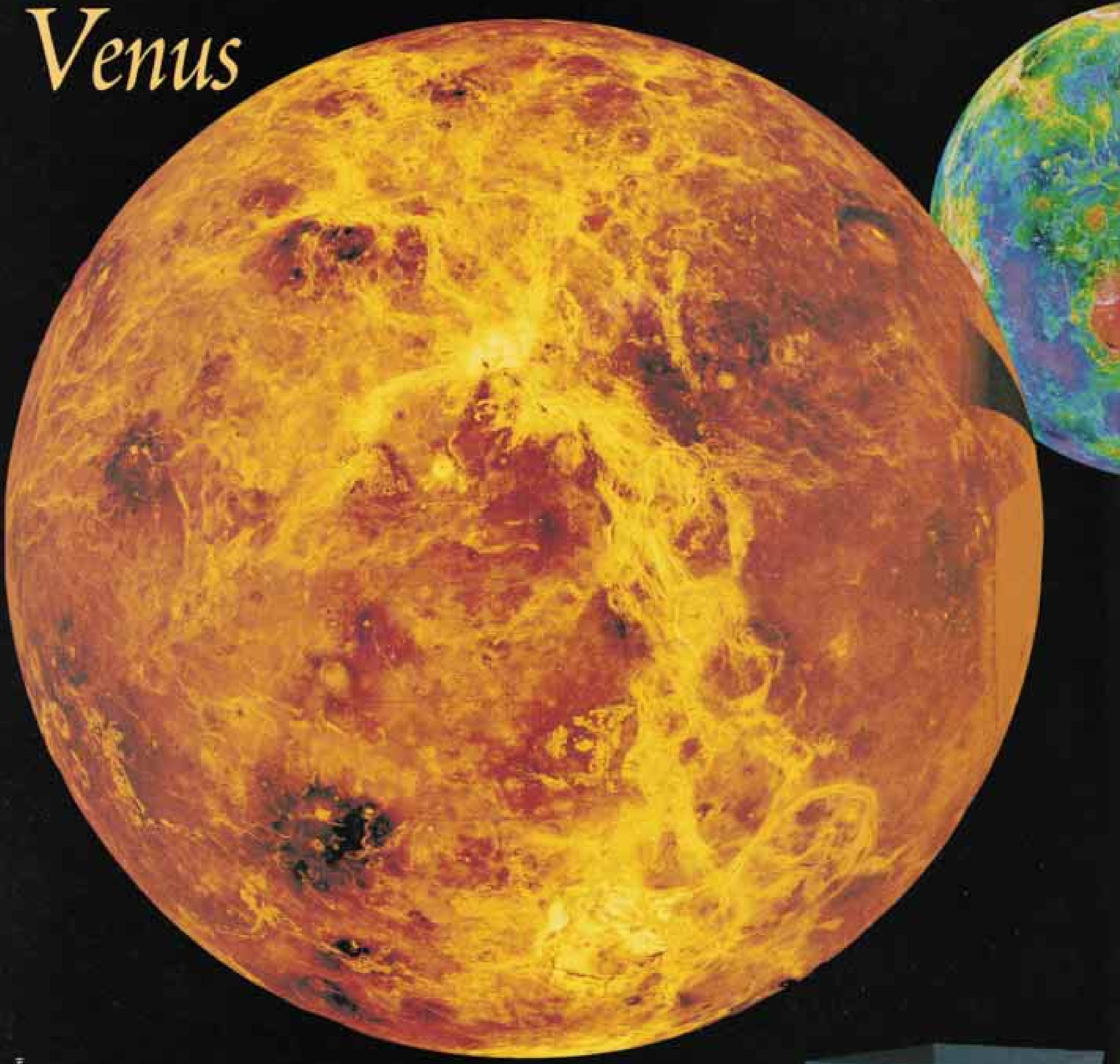
The planet also has an intriguing relation between the amount of time it takes to rotate—59 Earth-days—and the period required for it to complete a circuit of the sun—88 Earth-days. Mercury appears locked into this 2:3 ratio of rotational to revolutionary periods by the sun's grip on the planet's gravitational bulge. This grip is strongest every 1.5 rotations of the planet.

DISCOVERY SCARP (*crack shown in images at right*) is a 500-kilometer-long thrust fault probably created when parts of Mercury's core solidified and shrank. Daybreak seen from inside the scarp is probably a stirring sight (*below, at right*).



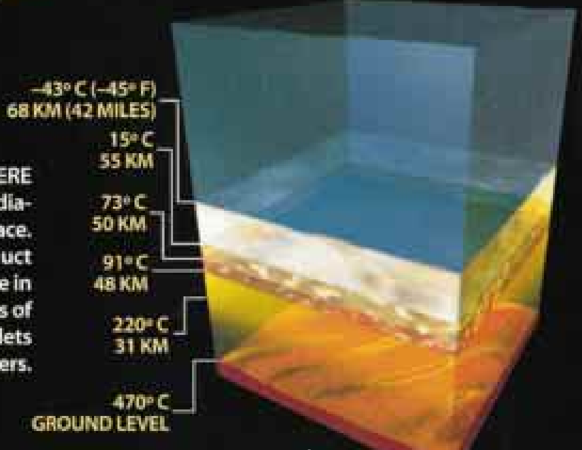


Venus

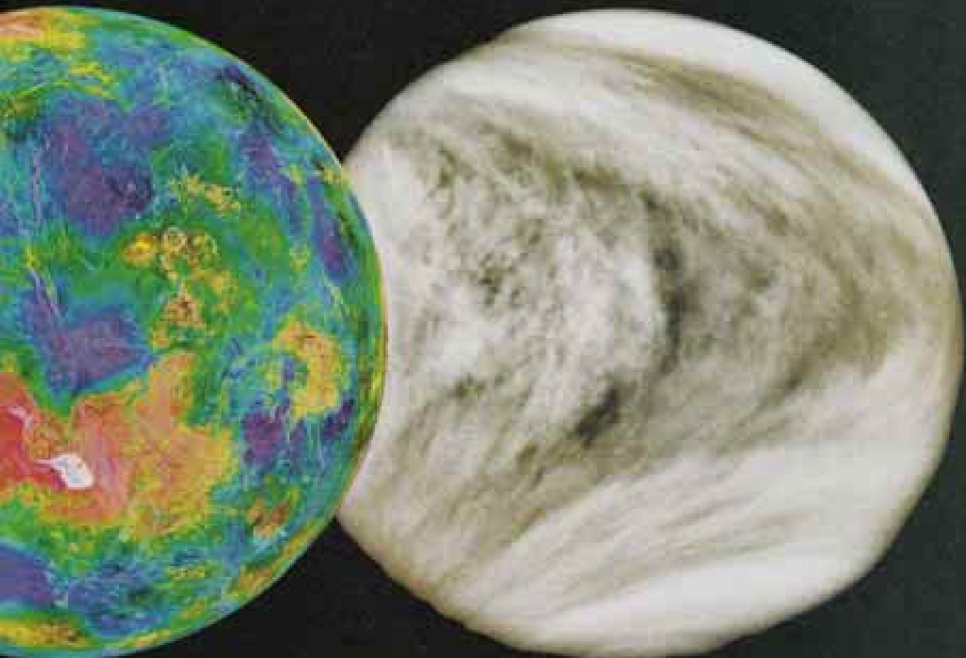


SIZE COMPARED WITH EARTH

THICK CARBON DIOXIDE ATMOSPHERE of Venus is opaque to infrared radiation, so it traps heat at the surface. Three decks of clouds are a by-product of a complex meteorological cycle in which sulfur goes through a series of reactions to form sulfuric acid droplets at altitudes above about 70 kilometers.



JPL/CALTECH/NASA (top); NASA (bottom left); SLIM FILMS (illustration)



VENUS UNVEILED, through the use of radar, was seen in a global view for the first time in an image produced using data from the Magellan orbiter in 1991 (*far left*). More recently, the U.S. Geological Survey used the Magellan data to produce topographic maps of the surface (*above, left*), which is normally obscured by clouds (*above, right*).

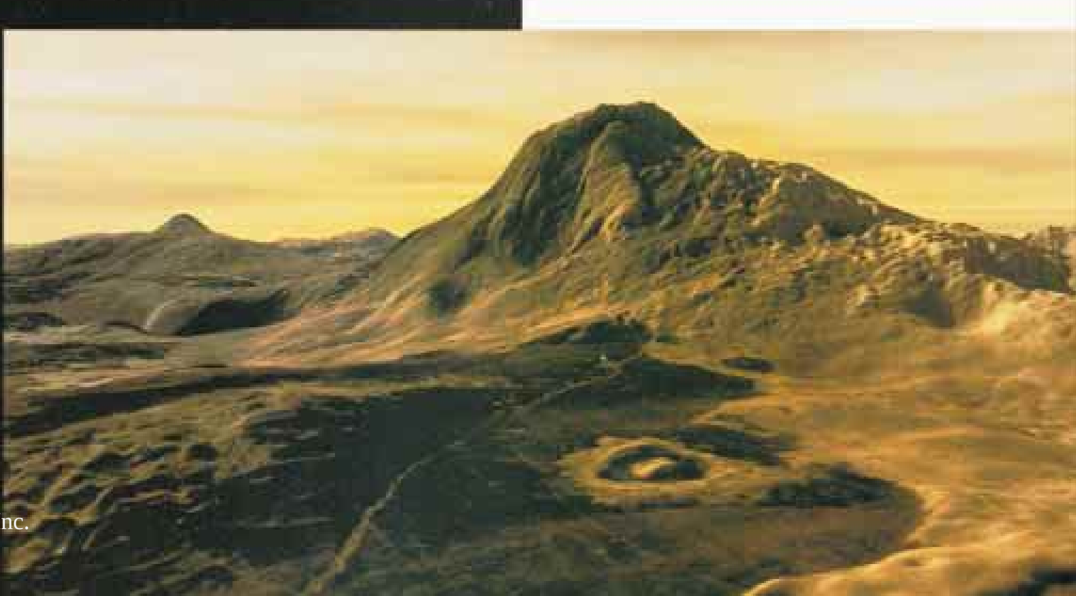
JPL/CALTECH/NASA (top left); NASA (top right and middle); DAVID P. ANDERSON Southern Methodist University (bottom)

Though named for the goddess of love, Venus is more like Earth's ugly sister. The two planets formed from the same general region of the solar nebula, suggesting that their compositions are basically similar. They are of roughly the same size, mass and density, and Venus orbits the sun at an average distance about 70 percent that of Earth's.

But where Earth has temperatures and conditions conducive to life, a variety of environments and a robust magnetic field, Venus is a dry, hellish, high-pressure furnace whose magnetic field is not even strong enough to keep the solar wind from stripping away the upper atmosphere. Below ever present clouds of sulfuric acid and a thick carbon dioxide atmosphere, the Venusian surface hits temperatures up to 450 degrees Celsius (842 degrees Fahrenheit).

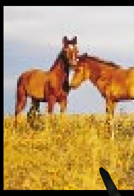
One of the fundamental mysteries about Venus is its relative scarcity of craters. This paucity suggests that the surface of the planet may be a mere 600 million years old. A convincing explanation eludes planetary scientists, although most agree that it must somehow involve volcanism and tectonics.

VENUSIAN SURFACE (*above*) was photographed by the Soviet Venera 13 lander in March 1982. The lander survived on the planet's surface for 127 minutes. The orange color detected in the Venera images was later added to radar images of the planet, such as the large one at the far left and the landscape at the right.

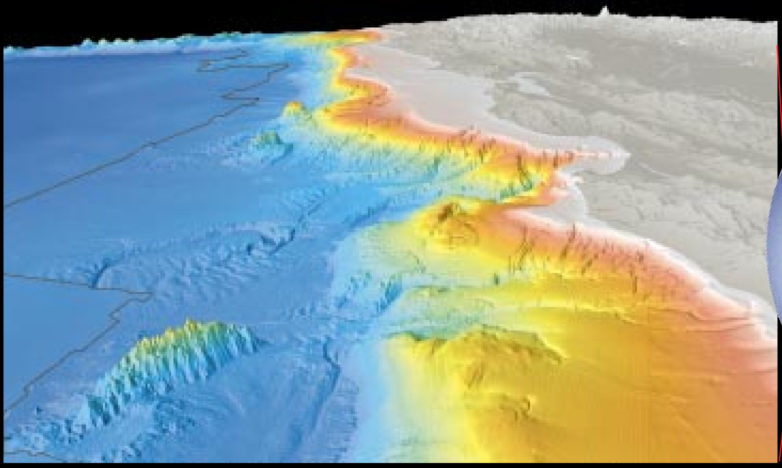


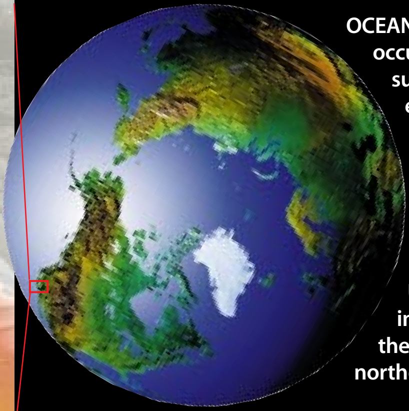
ENORMOUS VOLCANO known as Maat Mons was imaged from radar data collected by the Magellan orbiter. The data were processed to create this perspective, a view from a distance of about 550 kilometers and an altitude of 1.7 kilometers. The volcano itself is about six kilometers tall.

Earth

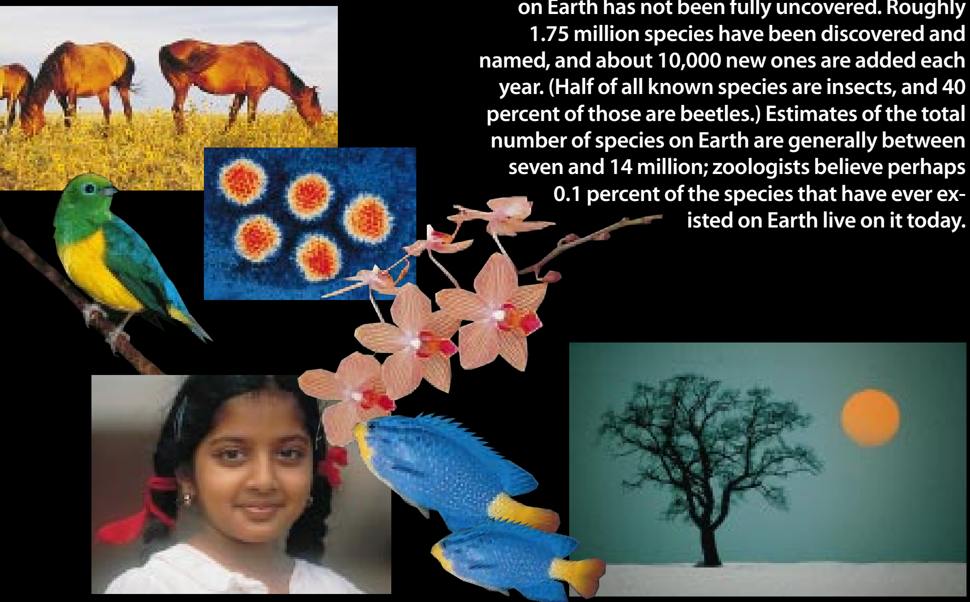


ASTROGEOLOGY TEAM, U.S. GEOLOGICAL SURVEY, FLAGSTAFF, ARIZ. (top); LINCOLN F. PRATSON AND WILLIAM F. HAXBY (bottom left); EDWARD BELL (bottom right)





OCEANS occupy 71 percent of the surface area of the planet but remain largely unexplored. During the 1980s and early 1990s, researchers from the National Science Foundation generated images of the U.S. continental shelf, including this picture of the Monterey Bay area in northern California (*left*).

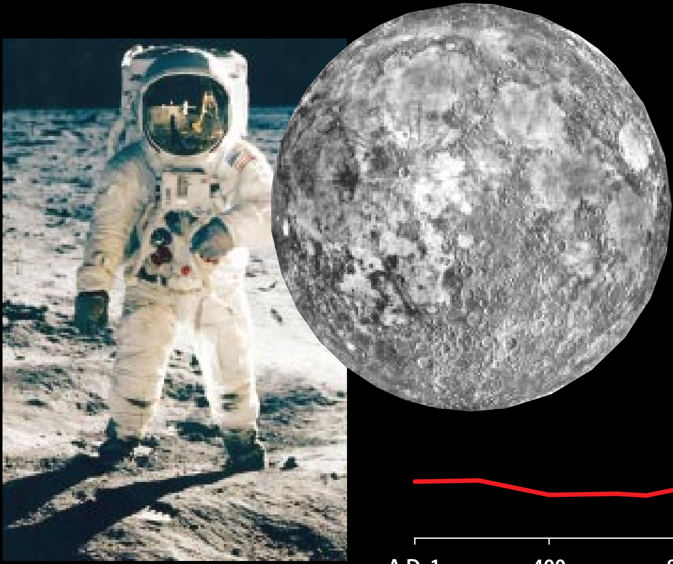


DIVERSITY OF LIFE on Earth has not been fully uncovered. Roughly 1.75 million species have been discovered and named, and about 10,000 new ones are added each year. (Half of all known species are insects, and 40 percent of those are beetles.) Estimates of the total number of species on Earth are generally between seven and 14 million; zoologists believe perhaps 0.1 percent of the species that have ever existed on Earth live on it today.

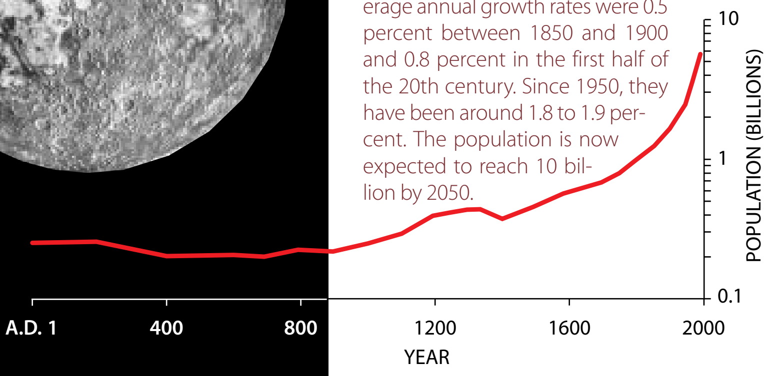
That it teems with life makes Earth a precious oddity among planets—although just how odd, scientists cannot say. Certainly the conditions that made life possible were sensitive to the planet's surface temperature and therefore to its distance from the sun.

Abundant liquid water was critical to the planet's evolution. This water moderated temperatures, eroded rocks, dissolved minerals and supported complex chemical reactions, some of which yielded single-celled life close to four billion years ago. Macroscopic animals started proliferating only around 600 million years ago, eons after photosynthesis enriched the atmosphere with oxygen.

Earth's large moon probably formed from debris after a collision between early Earth and another huge body. Because the moon and sun appear the same size from Earth, our planet is the only one to witness the beauty of the sun's corona during a total eclipse.



EARTH'S MOON has been visited by 12 people; shown here is Edwin P. ("Buzz") Aldrin, Jr., the second person to set foot on its surface. The moon orbits Earth at an average distance of 380,000 kilometers (236,000 miles) and has a diameter about one fourth that of Earth—making it an unusually large natural satellite.



HUMAN POPULATION, currently estimated at 5.8 billion, has surged in recent decades. Average annual growth rates were 0.5 percent between 1850 and 1900 and 0.8 percent in the first half of the 20th century. Since 1950, they have been around 1.8 to 1.9 percent. The population is now expected to reach 10 billion by 2050.

MAJOR ECOSYSTEMS of Earth are varied and include mountain, tropical rain forest, desert and ocean types. Urban areas, which have swelled disproportionately with population growth, are in some ways complex ecosystems in their own right.

LUIZ C. MARIGO *Peter Arnold, Inc.* (*bird*); BONNIE COSGROVE *SPL/Photo Researchers, Inc.* (*adenovirus*); STEVE SOLUM *Bruce Coleman Inc.* (*fish*); MICHEL VIARD *Peter Arnold, Inc.* (*orchids*); IFA/BRUCE COLEMAN INC. (*tree, girl*); JUDD COONEY *Natural Selection* (*horses*); NASA (*astronaut, moon*); SIEGFRIED EIGSTLER *Tony Stone Images* (*Alps*); RIC ERGENBRIGHT *Tony Stone Images* (*waterfall*); GARY YEOWELL *Tony Stone Images* (*desert*); REGIS LEFEBURE *Natural Selection* (*beach*); TIM CROSBY *Liaison International* (*Seattle*); BRYAN CHRISTIE (*chart*)

Mars

JPL/CALTECH/NASA (top); NASA (bottom left); JPL (bottom right)

SIZE COMPARED WITH EARTH

MARTIAN LANDSCAPE, (*right*) was photographed in July 1997 by the Mars Pathfinder lander, part of which is visible at the bottom of this panoramic image. The bumps on the horizon, called Twin Peaks, were about one kilometer south-southwest of the lander. Pathfinder carried a roving vehicle, Sojourner (*left*), which analyzed soil and a group of rocks. In the panorama, Sojourner can be seen in front of one of the rocks, which was dubbed Yogi.

MARTIAN METEORITE ALH84001 (*above*) was found to contain segmented objects, about 380 nanometers long (*right*), which some researchers took to be the fossilized remnants of bacterial life that came into contact with the rock more than 1.3 billion years ago. Other scientists, however, were more skeptical, contending that the formations had nonbiological origins and that the rock was chemically contaminated after it fell to Earth.

Mars's relative nearness, mythological connotations and even its hue have made it the favored planet of popular culture. Countless works of science fiction and science have explored the possibility of life on Mars. In 1976, however, the two U.S. Viking probes found no evidence of life at their landing sites.

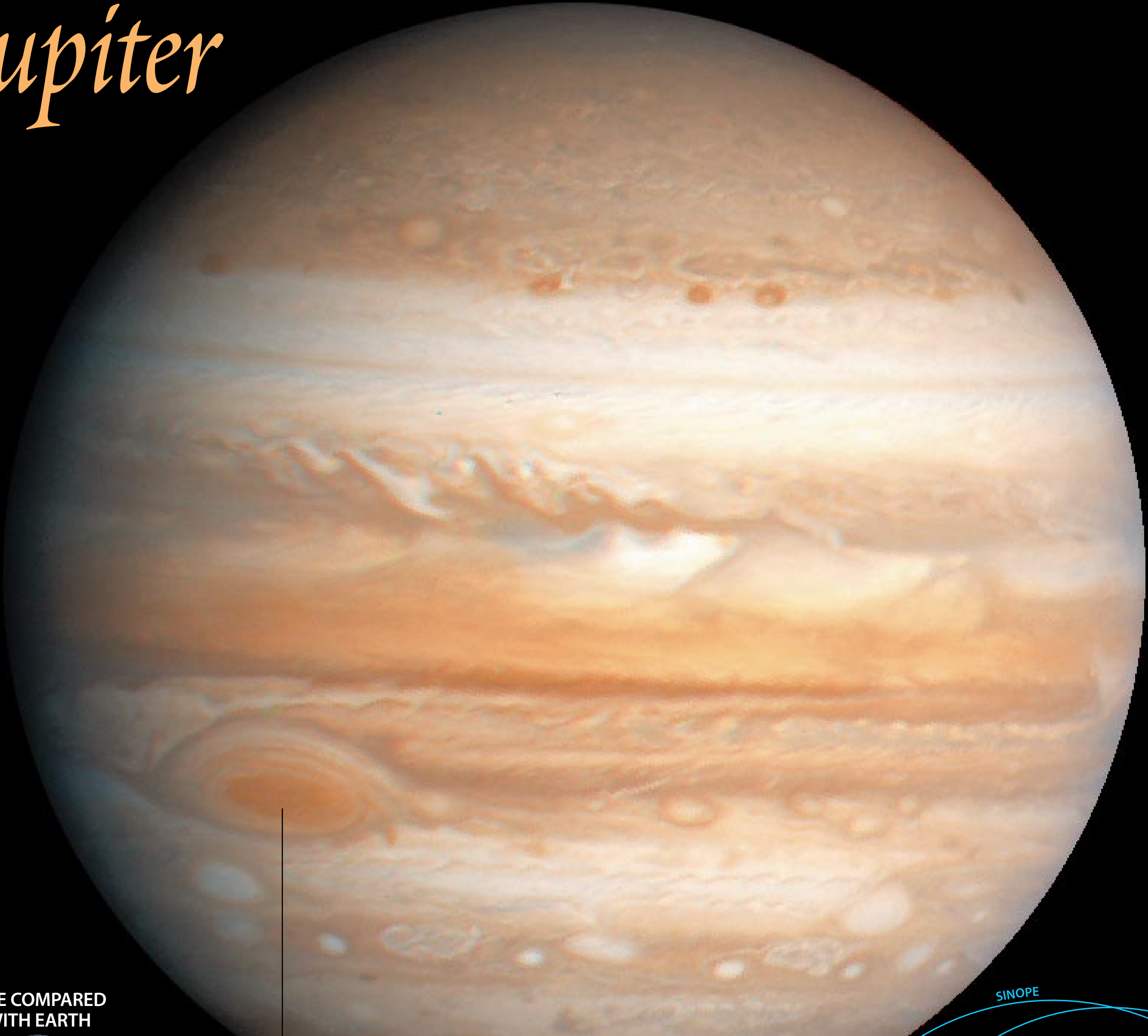
Two events thrust Mars back into the public consciousness lately. In 1996 a team from the National Aeronautics and Space Administration Johnson Space Center and Stanford University announced that unusual characteristics in a meteorite known to have come from Mars could be best interpreted as the vestiges of ancient bacterial life. In the summer of 1997 the Mars Pathfinder lander and its diminutive roving vehicle, Sojourner, analyzed and imaged Martian rocks, atmosphere and soil. Investigators concluded that many of the rocks were deposited by a massive flood at least two billion years ago and that some of them were surprisingly similar to a class of Earth rocks known as andesites.

MINUSCULE MARTIAN MOONS Deimos (*below, top*) and Phobos (*bottom*) are respectively about 15 and 27 kilometers (nine and 17 miles), at their longest. Because both moons are carbon-rich, some planetary scientists have concluded that they are captured asteroids from the relatively nearby asteroid belt.

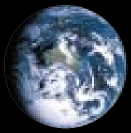
SINUOUS RIDGES known as eskers are made up of soil deposited by streams running under a sheet of ice. They appear to exist on the floor of the Argyre basin (*above, seen from orbit*) on Mars, suggesting that melting glaciers once covered the area. Evidence abounds that the planet was warmer and wetter in the past, although scientists still cannot say how much water there was, how many wet periods there were or how long they lasted.

NASA JOHNSON SPACE CENTER (*top two images*); COURTESY OF NASA/JPL (*Deimos*); NASA/JPL, ©1997 CALVIN J. HAMILTON (*Phobos*); JPL/CALTECH/NASA (*landscapes*)

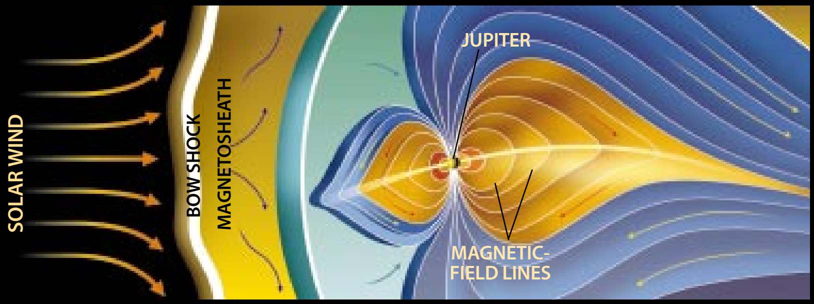
Jupiter



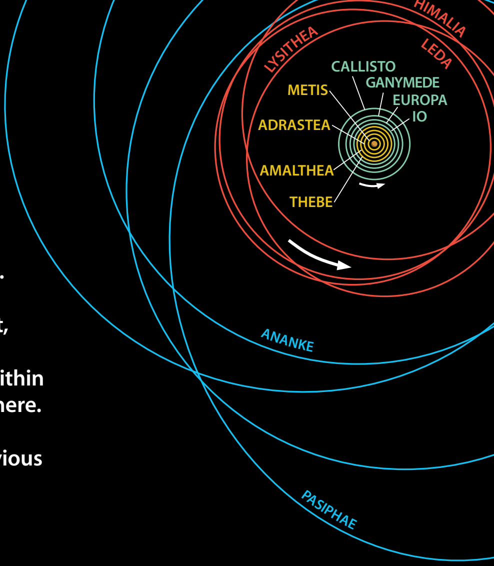
SIZE COMPARED WITH EARTH



JUPITER'S MOST CONSPICUOUS FEATURE, the Great Red Spot has persisted in the atmosphere since the first detailed observations of the planet were made. Two Earths could rest in the region marked by the spot. The material making up the spot appears to complete a counterclockwise rotation in 12 hours. Based on Voyager photographs, the interior of the spot is relatively stable. The Great Red Spot is thus a gigantic vortex, with wind speeds approaching 400 kilometers (250 miles) an hour.



IMMENSE JOVIAN MAGNETOSPHERE is larger than the sun. Its tail spreads out beyond Saturn's orbit, meaning that Saturn finds itself at times within Jupiter's magnetosphere. Solar winds push the field, causing the obvious asymmetry.

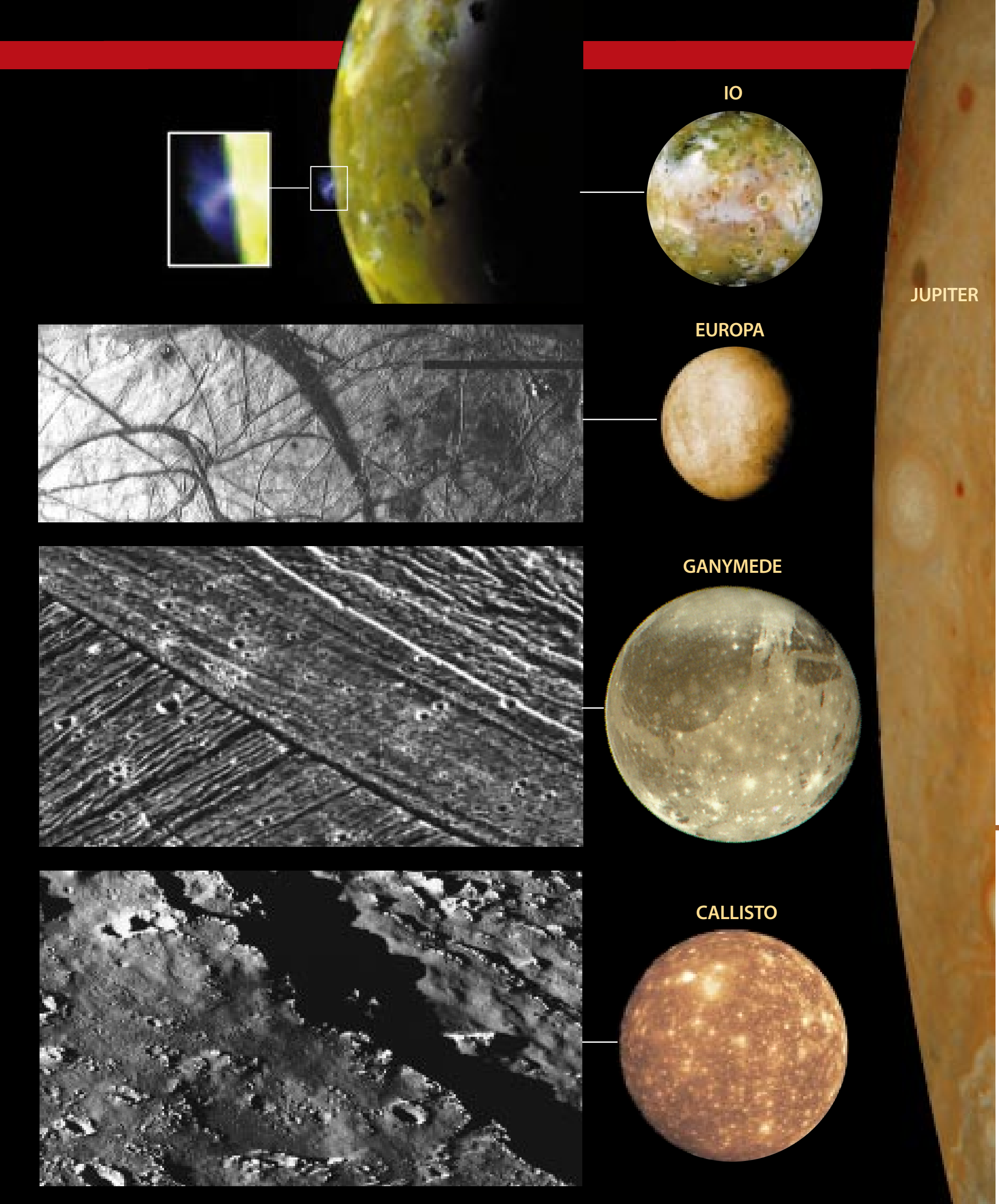


NASA/U.S. GEOLOGICAL SURVEY (*top*); NASA (*middle left*); ANDREW CHRISTIE (*bottom*)

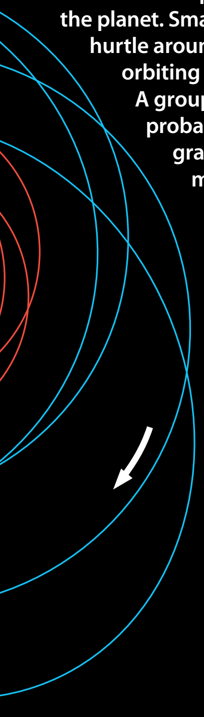
Jupiter represents a departure from the four relatively tiny rock planets that precede it as we travel away from the sun. It is the first of the four "gas giants," planets that dwarf Earth and that have no solid surfaces. Jupiter does everything on a grand scale. It is larger than all the other planets combined, and its moon Ganymede is bigger than Mercury.

Jupiter's hydrogen and helium content once led astronomers to think that the planet formed out of the same gas cloud that gave rise to the sun. More recent analysis of the subtleties in Jupiter's chemistry point to a solid core, with perhaps the mass of 10 Earths, about which the rest of the planet formed. Jupiter also differs in kind from the terrestrial planets by radiating more energy than it receives from the sun. In 1994 fragments of Comet Shoemaker-Levy 9 slammed into Jupiter, thrilling observers.

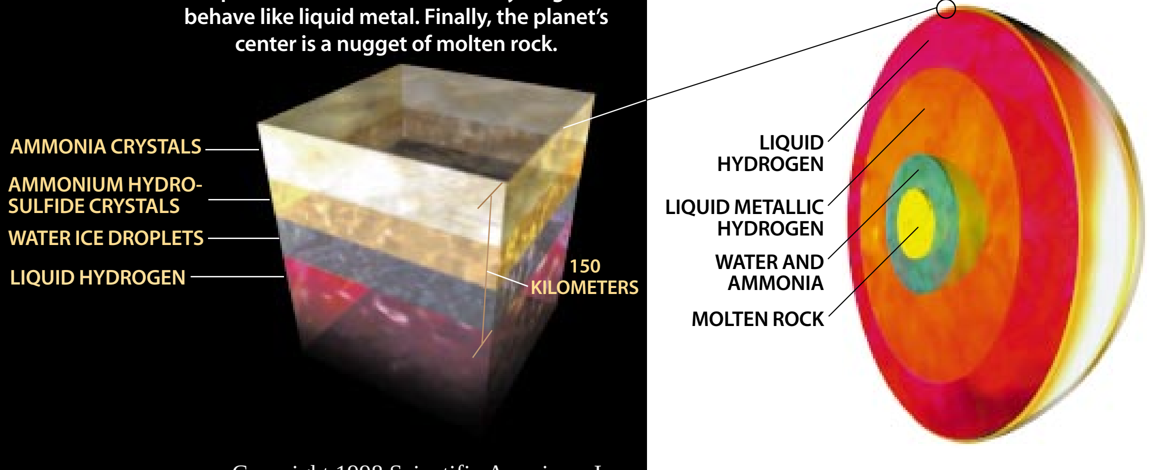
JPL/CALTECH/NASA (*all photographs*); BRYAN CHRISTIE (*bottom left*); ANDREW CHRISTIE (*bottom right*)



FOUR GALILEAN SATELLITES bear the name of their discoverer. Innermost Io suffers massive volcanic activity, caught by Voyager's camera (*top left*), that continually resurfaces the planet. Europa also seems to be continually resurfaced, but based on infrared spectra, this smallest of the Galilean moons appears to be covered with water ice, emerging from the interior and freezing at the surface. This false-color view shows contaminants in the ice (*red*) and vast frozen plains (*blue*). The presence of liquid water under that ice cover, along with organic molecules, has led some scientists to speculate that Europa's ocean may harbor some of the biochemistry necessary for life. The largest Galilean moon, Ganymede, is likely a mostly rocky core with a largely icy surface. That surface is marked by grooves hundreds of meters deep that run for thousands of kilometers, probably the result of early tectonic activity. Kin to the rest of the Galilean satellites but different in kind, Callisto's surface shows no evidence of any resurfacing since its craters were first formed by impacts some four billion years ago. The photographed cliff, causing the shadow (*left*), is part of a ring left by an impact.



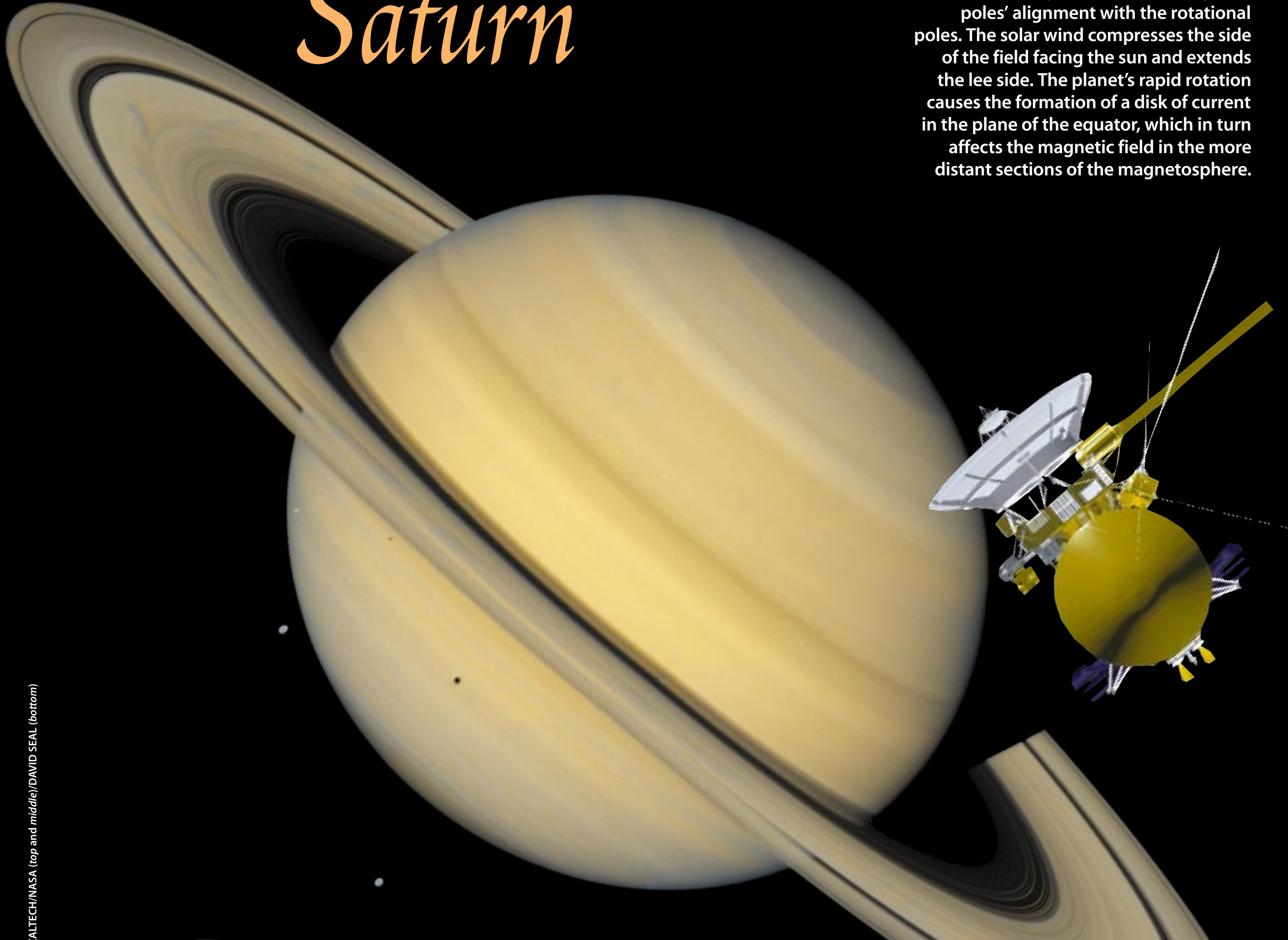
FOUR DISTINCT CLASSES OF SATELLITES orbit giant Jupiter. The Galileans (*green*) travel in almost perfect circles close to the planet. Small nearby moons (*yellow*) hurtle around Jupiter, with two orbiting in just seven hours. A group of small moons (*red*) probably were captured by gravity. Finally, outer moons (*blue*) revolve in the opposite direction in highly elliptical and tilted orbits.



CROSS SECTION OF JUPITER reveals its layers. Cold clouds of ammonia, hydrogen and water rest atop hot liquid hydrogen. Go deeply enough into the planet, and pressure and heat cause the hydrogen to behave like liquid metal. Finally, the planet's center is a nugget of molten rock.

Saturn

MAGNETOSPHERE OF SATURN is quite orderly, thanks to the magnetic poles' alignment with the rotational poles. The solar wind compresses the side of the field facing the sun and extends the lee side. The planet's rapid rotation causes the formation of a disk of current in the plane of the equator, which in turn affects the magnetic field in the more distant sections of the magnetosphere.



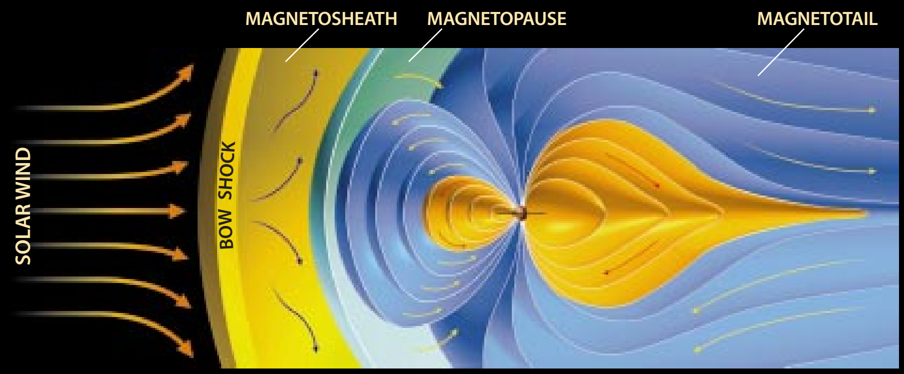
RHEA

IAPETUS

JPL/CALTECH/NASA (*top and middle*)/DAVID SEAL (*bottom*)

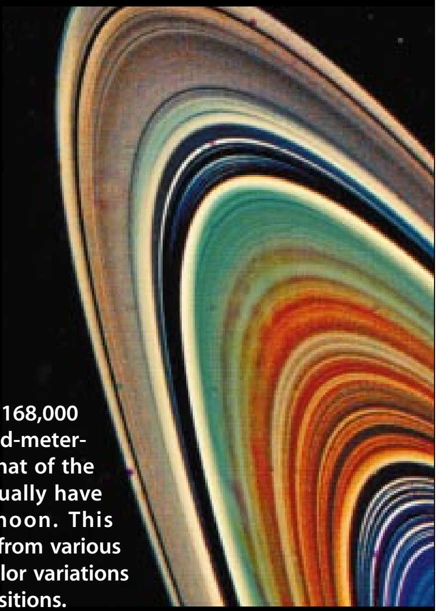
Saturn's rings make it one of the most familiar, and spectacular, images of astronomy, not to mention science fiction. When Galileo trained a primitive telescope on the planet for the first time in 1610, he was misled. From the poorly resolved image in his viewfinder, he believed Saturn to be a triple-system, with a large body in the center and smaller ones on each side. The rings may be much younger than the planet itself, and great mathematicians have found them worthy of contemplation. Laplace and James Clerk Maxwell calculated that Saturn's rings must consist of many smaller objects. Although the planet is almost the size of Jupiter, its mass is but one third as great, giving Saturn the lowest mean density of any solar system object.

As a gas giant, the planet has no single rotation period but rather a variety depending on latitude. Upper atmosphere clouds travel around the equator in as little as 10 hours and 10 minutes; clouds in high latitudes may take half an hour longer to pass across the planet. Based on gravitational field data, Saturn appears to have a solid core with a mass equivalent to up to 20 Earths. As the most oblate planet, the pull of gravity at its equator is less than three quarters of that at the poles.



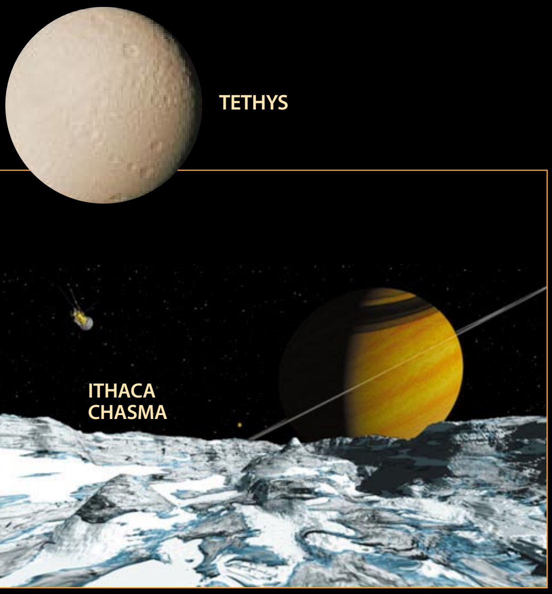
ANDREW CHRISTIE (*top*); JPL/CALTECH/NASA (*upper middle, middle and lower moons*); JPL/NASA/DAVID SEAL (*bottom*)

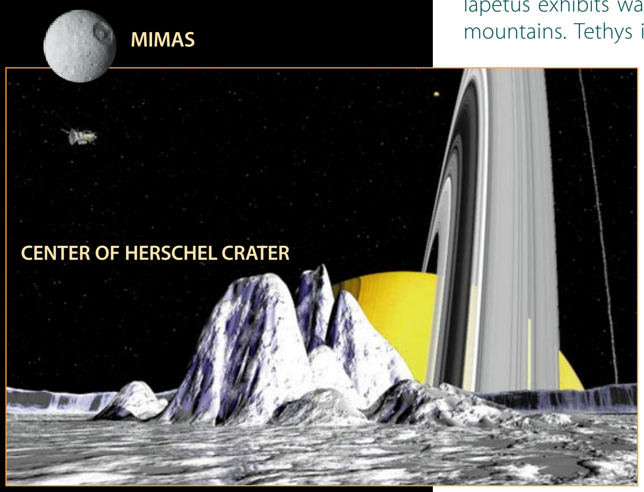
CASSINI SPACECRAFT
left Earth in October 1997 for a Saturn rendezvous in late 2004. The ship is named for Jean-Dominique Cassini, who in 1675 discovered the gap in the rings, known as the Cassini division. Once it arrives at Saturn, Cassini will launch the Huygens probe, which will descend to the surface of the moon Titan. Huygens will chemically sample the thick atmosphere as it falls to the surface and may continue to operate for as long as an hour once it lands—or splashes down in liquid hydrocarbons. Titan's chemistry may be similar to that of early Earth.



SATURN'S RINGS
have a diameter of some 270,000 kilometers (168,000 miles). The total mass of the several-hundred-meter-thick rings, however, is only equivalent to that of the Saturnian moon Mimas. The rings may actually have formed from a shattered Mimas-size moon. This enhanced color photograph was assembled from various filtered views captured by Voyager 2. The color variations may represent differences in chemical compositions.

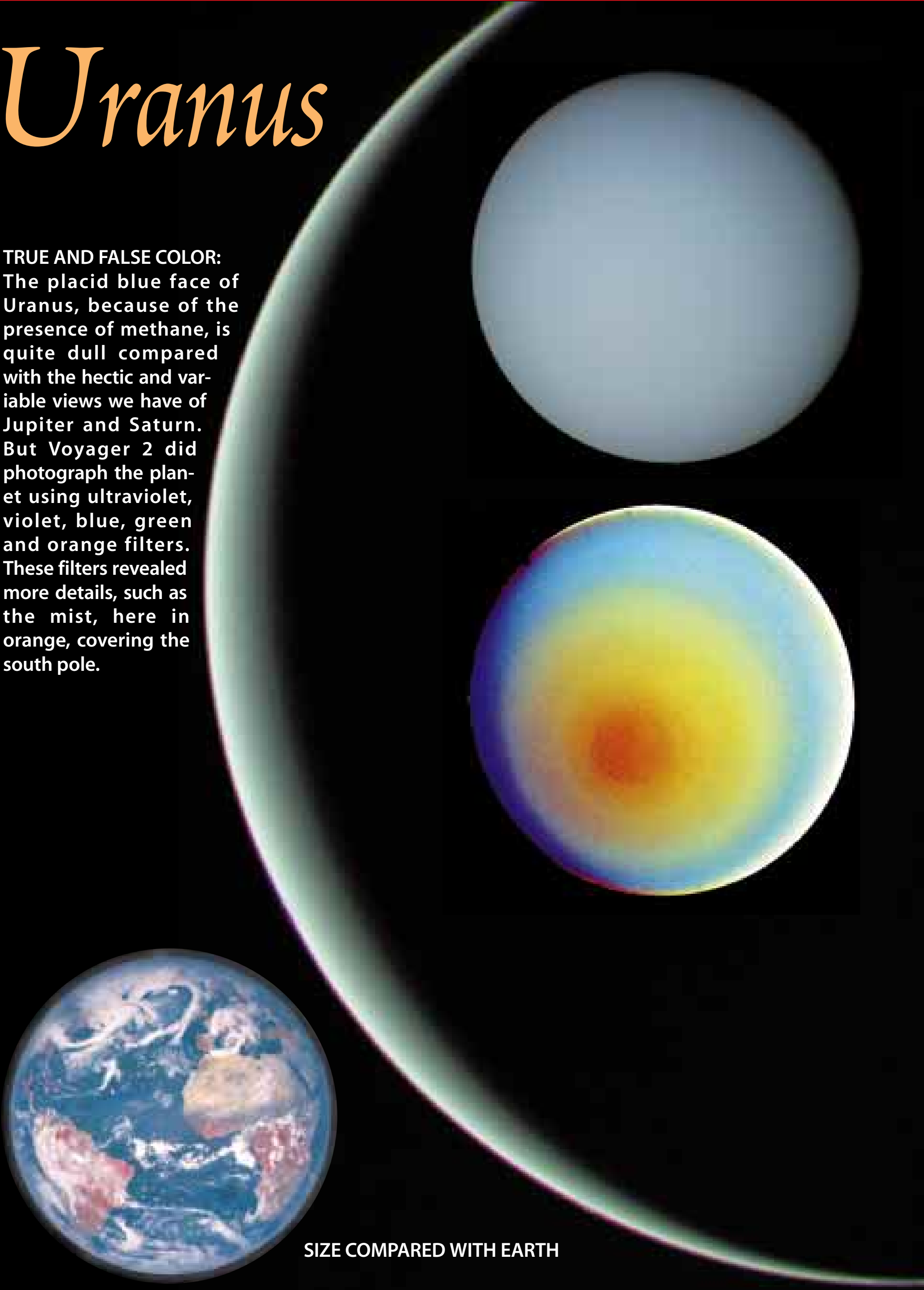
SMALLER MOONS OF SATURN (*in orbital order, outermost at left*) are dwarfed by Titan. Pan, Atlas, Telesto, Calypso and Helene are shown at a five-times-larger scale for visibility. Density measurements indicate that all of the moons are rich in ice, mostly water ice and possibly some ammonia. Many exhibit quirks and oddities: Hyperion has the solar system's only known chaotic orbit. Enceladus may have volcanoes. Rhea is extremely cratered, although brighter regions may be new ice formations. Iapetus exhibits wavy ice structures as well as mountains. Tethys is heavily cratered and features the Ithaca Chasma, a 100-kilometer-wide trench some four to five kilometers deep running almost pole to pole. Mimas is marked by the 10-kilometer-deep Herschel Crater, which has a diameter of 130 kilometers, fully one third that of the entire moon.



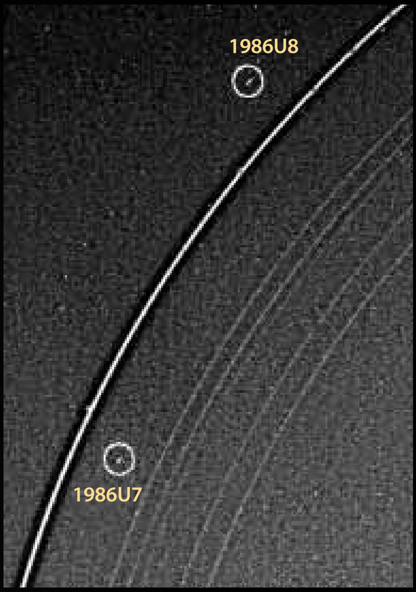


Uranus

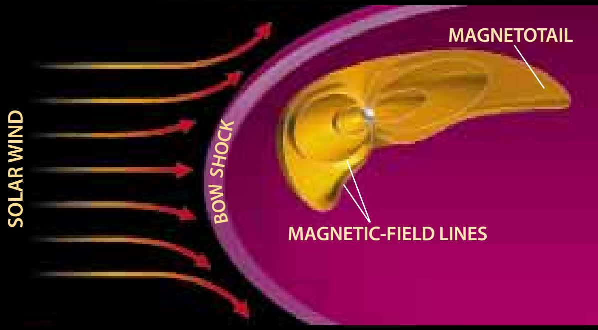
TRUE AND FALSE COLOR: The placid blue face of Uranus, because of the presence of methane, is quite dull compared with the hectic and variable views we have of Jupiter and Saturn. But Voyager 2 did photograph the planet using ultraviolet, violet, blue, green and orange filters. These filters revealed more details, such as the mist, here in orange, covering the south pole.



SIZE COMPARED WITH EARTH

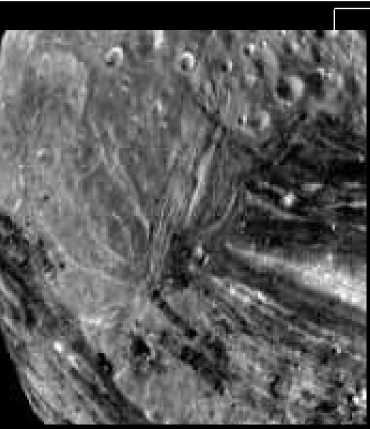


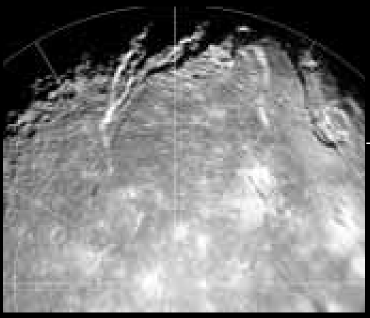
SHEPHERD MOONS hem in the Epsilon ring through gravitational interactions from either side. The shepherds Ophelia (1986U8) and Cordelia (1986U7) were caught in the act by Voyager's camera (*above*). The Epsilon ring is the brightest and broadest of the nine rings, all clearly visible in the image (*right*) captured by Voyager from a distance of more than one million kilometers from the planet.



MAGNETOSPHERE OF URANUS is tilted 59 degrees from the rotation axis. In addition, the field is skewed, perhaps because its dynamo region is well off-center. In general, no planetary dynamo, including Earth's, has been convincingly explained.

JPL/CALTECH/NASA (*top two views and background*); ASTROGEOLOGY TEAM, U.S. GEOLOGICAL SURVEY, FLAGSTAFF, ARIZ. (*middle left*); JPL (*middle right*); ANDREW CHRISTIE (*bottom*)





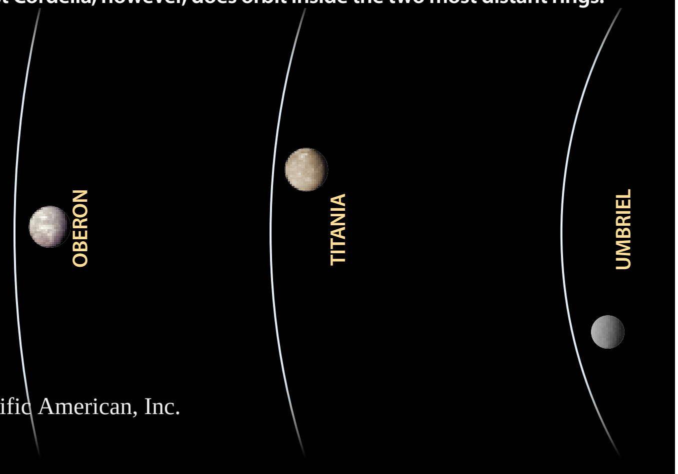
FIVE MAJOR MOONS are mixtures of rock and ice. Ariel, Umbriel, Titania and Oberon have densities that indicate compositions of about three parts ice to two parts rock. Smaller Miranda, as well as the other 10 tiny moons, probably has a greater proportion of ice. The surfaces of Oberon and Umbriel are densely cratered. Titania and Ariel are in keeping with Oberon and Umbriel with respect to density of small craters, but they have far fewer larger craters, in the 50- to 100-kilometer (31- to 62-mile) range. These larger craters are probably older, leading astronomers to believe that Titania and Ariel have younger surfaces than Oberon and Umbriel, for reasons as yet unclear. All the moons have canyons that seem to reveal ancient spreading and fracturing of their surfaces because of expansions of 1 to 2 percent, with the exception of Miranda, which probably expanded more on the order of 6 percent. The expansions could be the result of the freezing of what was originally liquid water, but the presence of liquid water at any time on these moons still requires an explanation. Miranda's expansion scarred the surface with extensive networks of grooves and troughs (*above*) as well as deep canyons that reach widths of 80 kilometers and depths of perhaps 20 kilometers. The large trenches on Titania (*immediately above*) suggest that the moon had at least one period of severe tectonic activity.

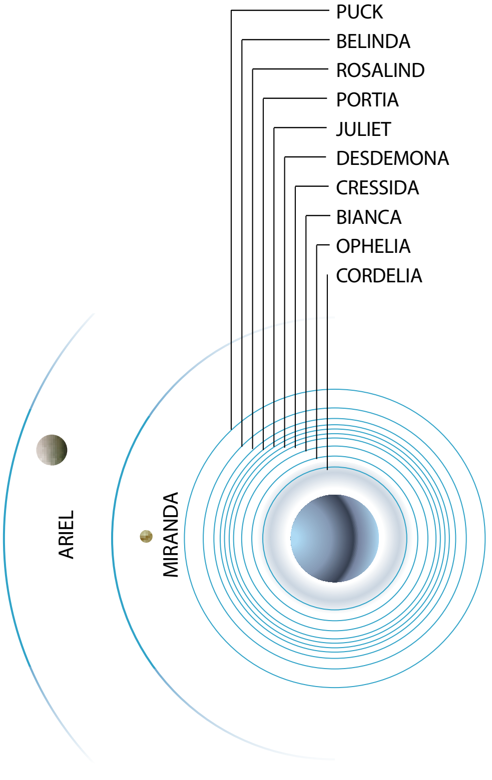
Strange even by the standards of the far reaches of the solar system, Uranus is an almost featureless, blue-green planet that has the distinction of being knocked on its side. Its axis of rotation points 98 degrees away from its orbital axis. This unique tilt most likely testifies to a massive collision while the planet was still forming. Adding to its peculiarity, Uranus's magnetic field is also tilted, 59 degrees from the rotation axis. Finally, the planet rotates in the opposite direction that Earth does. Although greatly enhanced images from Voyager 2's visit in 1986 reveal bands like those on Saturn and Jupiter, the planet seems to be far more placid than its stormy gas giant comrades. Uranus maintains their custom, however, of accompaniment by rings and numerous satellites.

Ten small moons were discovered by Voyager in 1986. Nine rings were found in 1977 during stellar occultations; two more have been found since.

JPL/CALTECH/NASA (*Ariel surface, Miranda, Ariel, Umbriel and Oberon*); COURTESY OF A. TAYFUN ONER (*Titania surface*); NASA, VOYAGER 2 AND CALVIN J. HAMILTON (*Titania*); BRYAN CHRISTIE (*diagram*); JPL (*bottom left*)

FIFTEEN OF THE MOONS OF URANUS orbit in near-perfect circles. Although the planet was discovered in 1781, it would be more than two centuries before Voyager found the 10 smaller moons. In the fall of 1997 astronomers found two more very small moons (*not shown*) in relatively eccentric orbits. In general, the rings orbit nearest the planet, followed by the smaller moons, with the large moons farthest away. Innermost Cordelia, however, does orbit inside the two most distant rings.



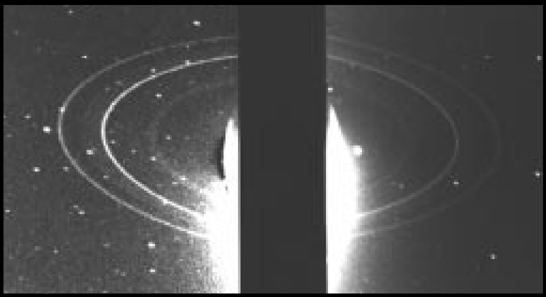


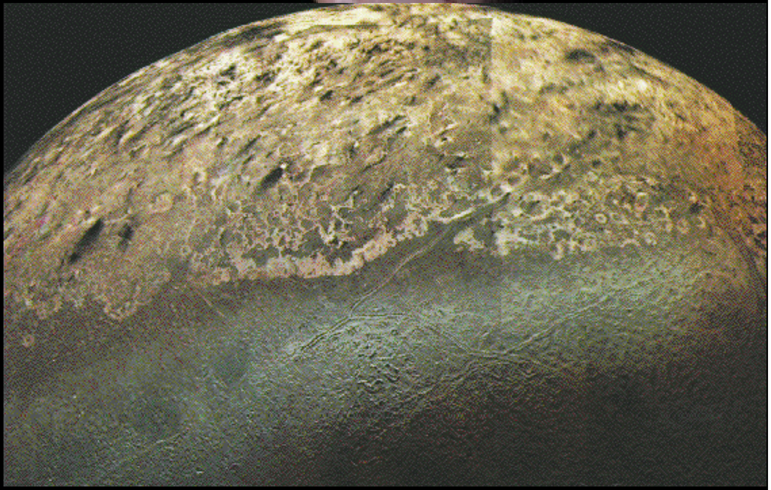
Neptune

JPL/CALTECH/NASA

GREAT DARK SPOT AND CLOUD LAYERS are clearly visible in these Voyager images. The dark spot (*left*) is probably a vast storm system rotating counterclockwise. Patterns in the white clouds accompanying the dark spot change greatly from one dark spot rotation to the next. Linear strips of clouds (*right*) stretch almost exactly along latitude lines.

NEPTUNE'S FAINT RINGS
(*right*) are ordinarily overwhelmed by the brightness of the planet, but this split image blocks the overexposed Neptune. Two sharply defined rings are clearly visible in these Voyager images. A third, diffuse ring is closer to the planet. The braided appearance of part of the outer ring (*left*) may be from clumping in the original ring material when it first began orbiting. Voyager's own motion, smearing the image slightly, may also be contributing to the unusual scene.



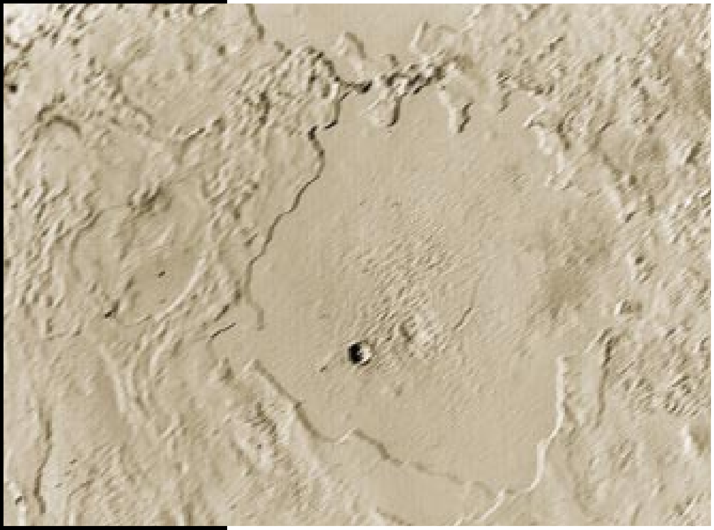


Astronomers searched for an eighth planet when Uranus's observed orbit disagreed with its calculated one, leading to suspicions of a large body exerting gravitational forces. In 1846 they confirmed the existence of Neptune, a planet so far from the sun that it will take another 13 years before it completes its first full orbit since discovery. The planet is the eighth from the sun in average distance, but it ends a two-decade tenure as the outermost planet in 1999, when Pluto again moves beyond it. The atmosphere of deep-blue Neptune is raked by winds moving at up to 700 meters (2,300 feet) per second, the fastest found on any planet. Denser than the other gas giants, Neptune probably has ice and molten rock in its interior, although rotational data imply that these heavy materials are spread out rather than concentrated in a tidy core.

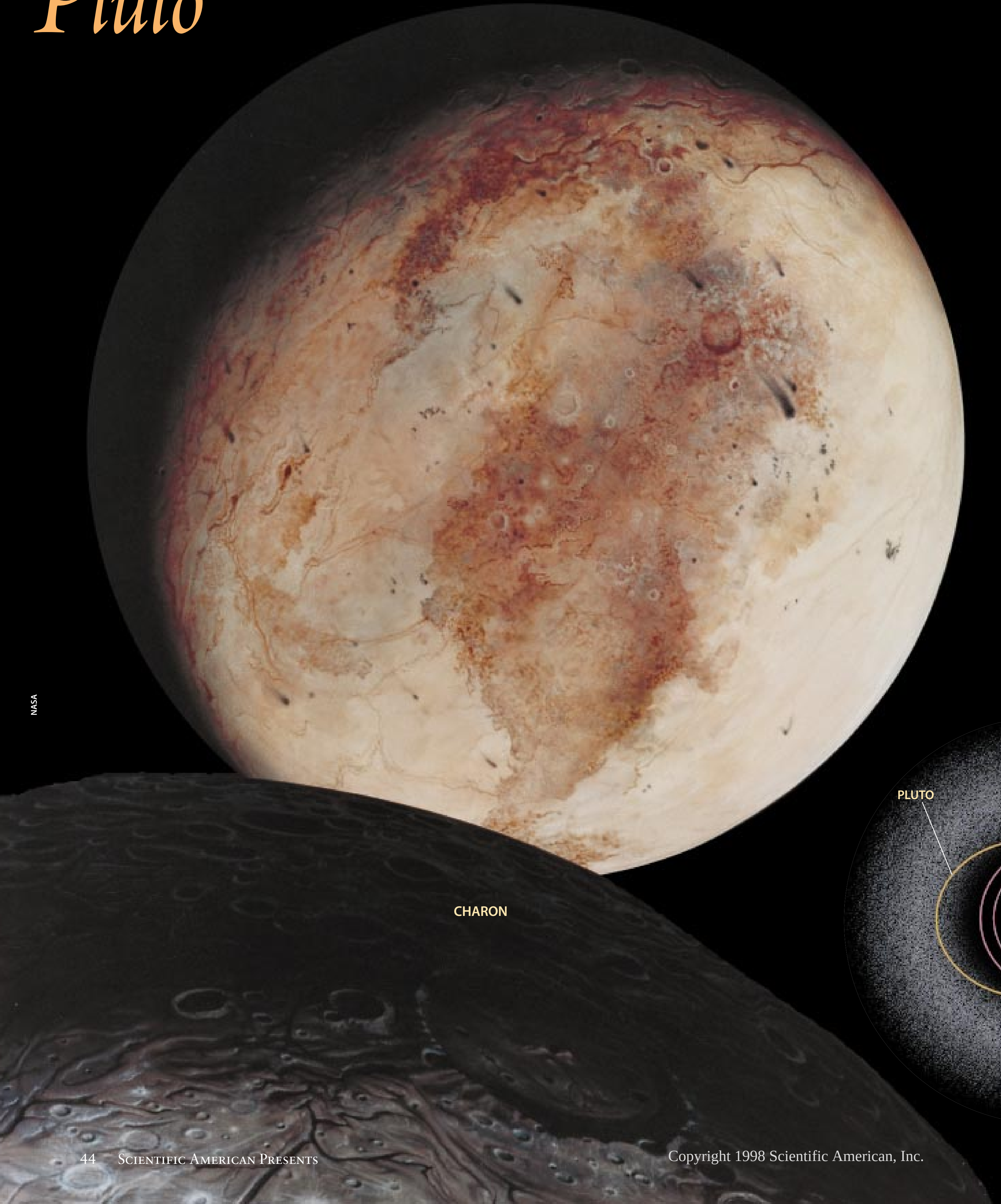
Like Uranus, Neptune has a magnetic field off kilter with its rotational axis, the latter's being tilted by 47 percent. The source of the field seems to be well outward from the planet's center. Its rings may have formed long after the planet itself, and the outermost ring's odd assortment of particle sizes may be the result of a satellite breakup within the past few thousand years. Neptune's defiant moons include Nereid, with the most eccentric orbit of any planetary satellite, seven times as distant from the planet at its farthest compared with its closest approach; and Triton, whose orbit opposes Neptune's rotation and is tilted 157 degrees from the planet's equator.

JPL/NASA (*top*); JPL/CALTECH/NASA (*middle and bottom*)

CONTRARY TRITON
is the only large moon known to travel in the direction opposite to its planet's rotation. Adding to its oddity is its rotation, tilted from Neptune's by 157 degrees. Triton may well have been an independent body later captured by Neptune's gravity. Voyager observations greatly improved our understanding of this moon. It probably has a rocky interior surrounded by water ice. The pink hue (*top*) may be caused by evaporation of a surface layer of nitrogen ice. Dark streaks across the south polar cap (*bottom*) may be from eruptions of ice volcanoes, a kind of frigid geyser. The ejecta is probably liquid nitrogen, dust or methane. Icy plains look suspiciously like lakes (*right*), suggesting that regions of the surface were once fluid.



Pluto

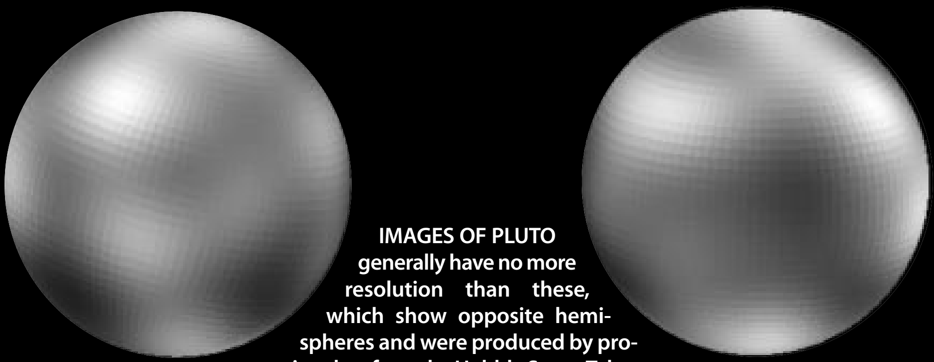


NASA

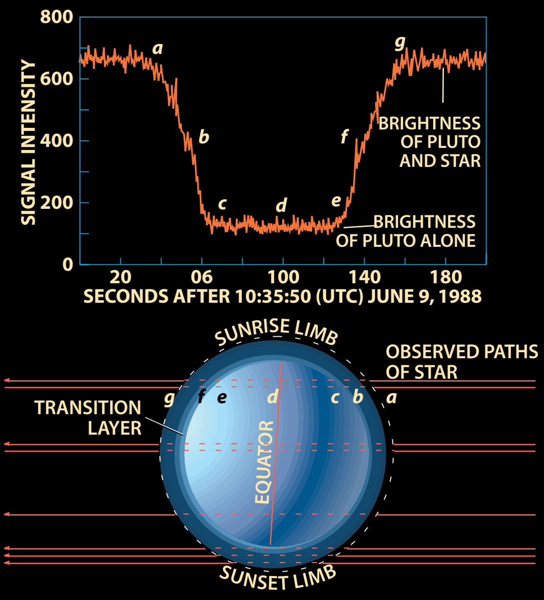
Is Pluto really a planet? Until about six years ago, the question would have seemed silly. But in the early 1990s, astronomers found a region of orbiting bodies just beyond Neptune. The region, which was dubbed the Kuiper belt, is populated mostly by bodies too small to be planets and also by comets with relatively short periods, meaning that they approach the sun at least once every couple of centuries.

Most astronomers still consider Pluto a planet. Although its mass is only $^{1}/_{400}$ that of Earth, it is still easily the largest Kuiper-like object. Also, Pluto seems to be more reflective than the other bodies in the Kuiper belt. Tradition may also have something to do with it; Pluto has been regarded as a planet since Clyde Tombaugh discovered it in 1930.

Pluto has never been photographed with high resolution; the best photographs that exist were made with the Hubble Space Telescope (*above left*). While studying much coarser images in 1978, James W. Christy, an astronomer at the U.S. Naval Observatory, noticed a bump in Pluto's disk. The bump turned out to be a satellite, which was named Charon after the mythological oarsman who ferried passengers across the river Styx to Pluto's realm.



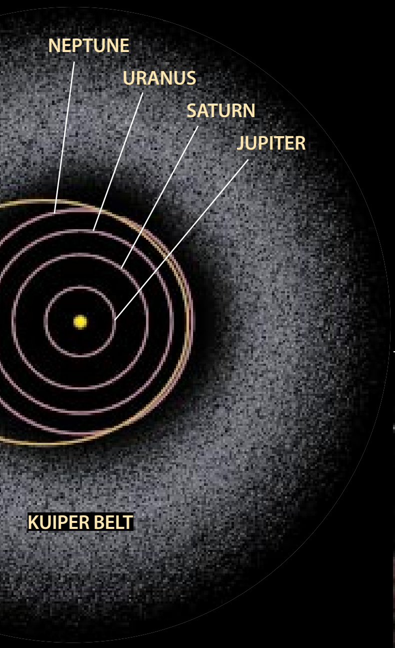
IMAGES OF PLUTO generally have no more resolution than these, which show opposite hemispheres and were produced by processing data from the Hubble Space Telescope. The data suggest that the face of Pluto has more large-scale contrast than any other planet, except possibly Earth. This fact and other information about the planet were used to create the artist's conception of Pluto at the far left.



TWINKLING STARLIGHT demonstrated that Pluto has an atmosphere. Astronomers at eight sites watched as Pluto passed in front of a distant star on June 9, 1988. The star dimmed gradually as its light entered the atmosphere. A sharp drop in the light curve (*b* and *f*) indicated a transition layer in Pluto's atmosphere. This layer may be haze or a region of rapid temperature change.

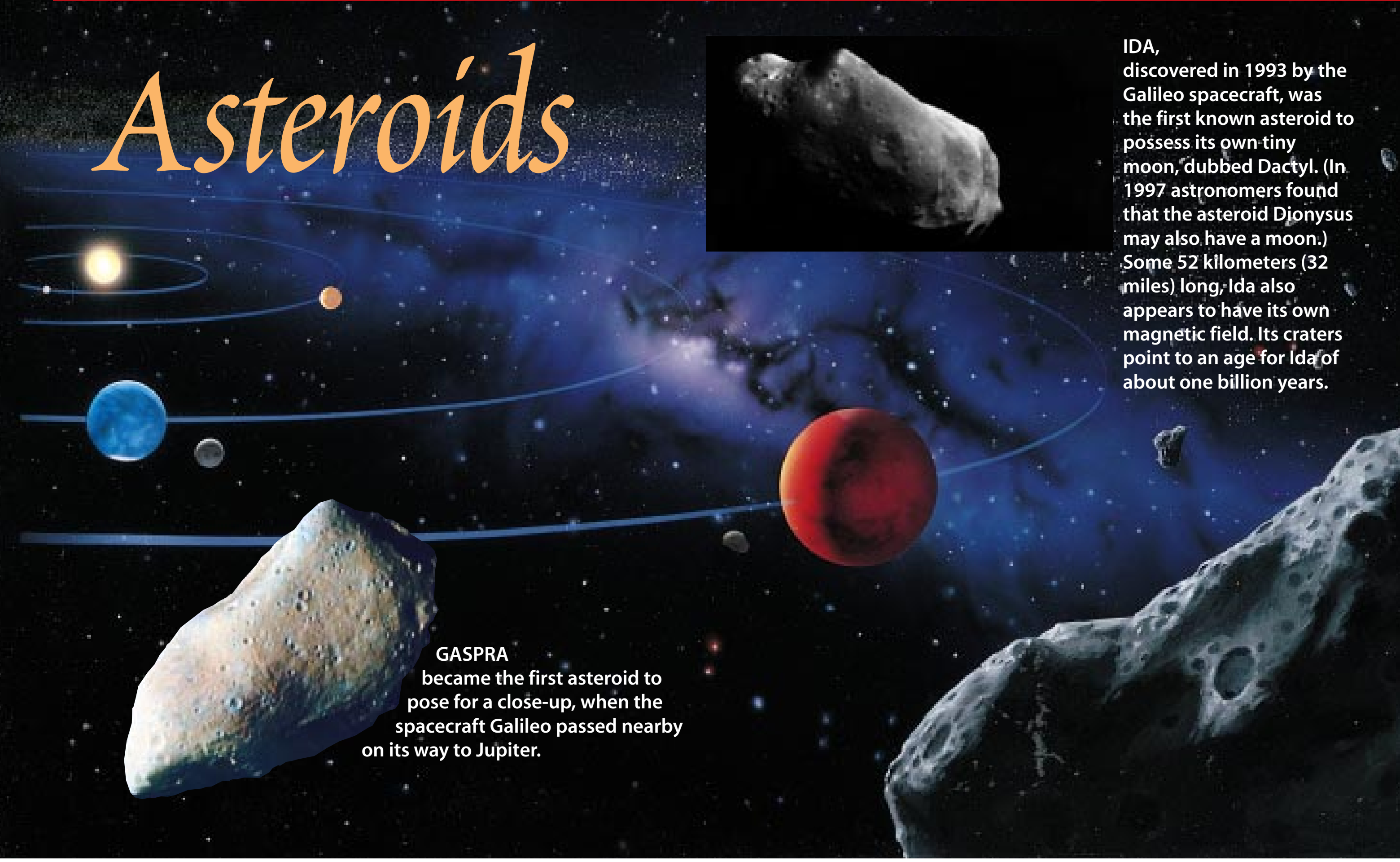
NASA AND ESA (top); BRYAN CHRISTIE (middle); SLIM FILMS (lower left); EDWARD BELL (bottom)

KUIPER BELT consists of incipient comets and objects too small to be considered planets. Astronomers estimate that the total mass of all the items in the belt is about one quarter to one half that of Earth. Other than Pluto, the largest objects are only hundreds of kilometers across.



PLUTONIAN PANORAMA could include a brilliant starry sky and a view of Charon over jagged terrain, tinged pink by complex photochemistry, with patches of frozen methane, carbon monoxide and nitrogen. The planet's atmosphere is so thin that the sky probably looks black even in the daytime.

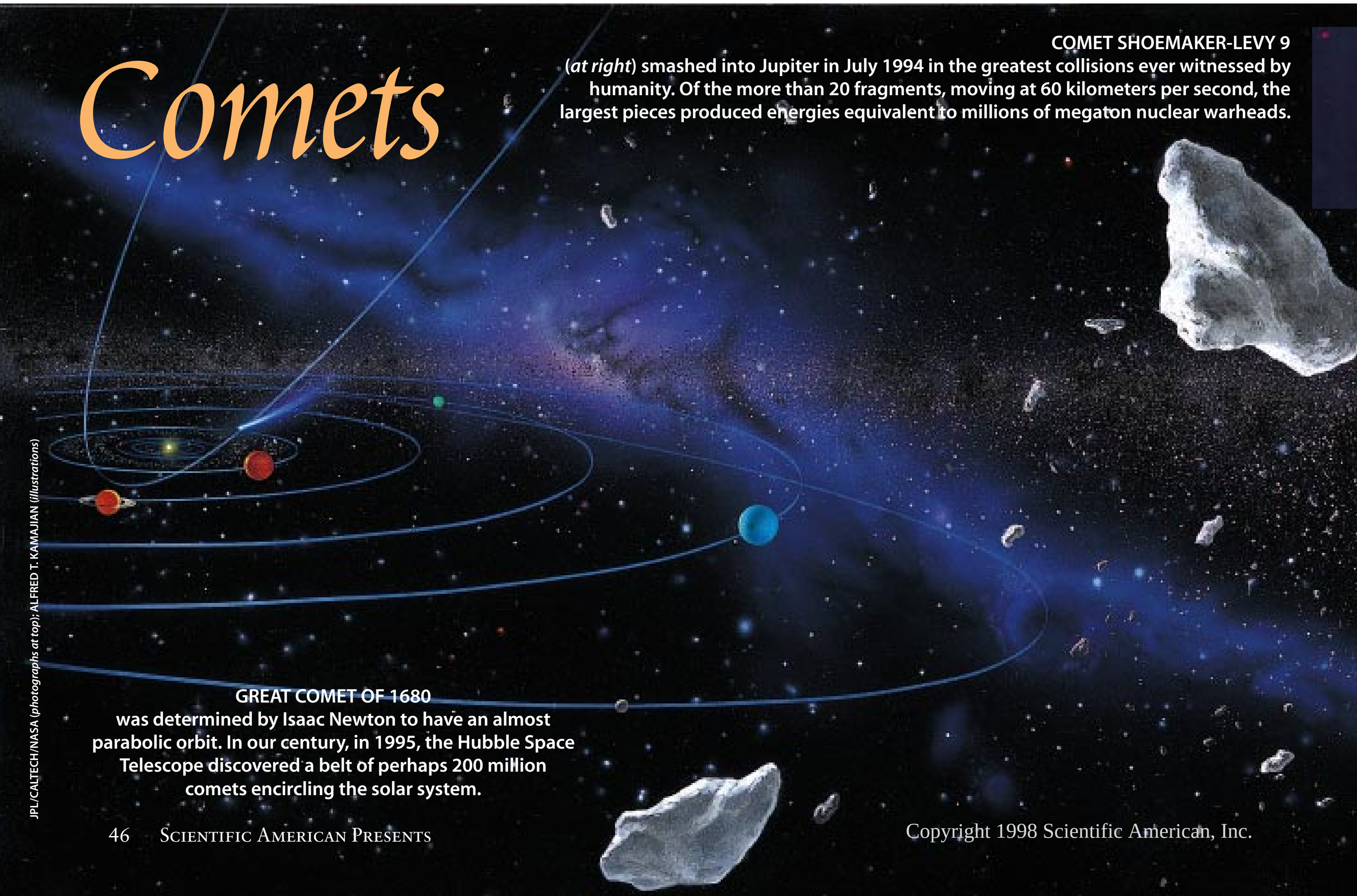
Asteroids



IDA,
discovered in 1993 by the Galileo spacecraft, was the first known asteroid to possess its own tiny moon, dubbed Dactyl. (In 1997 astronomers found that the asteroid Dionysus may also have a moon.) Some 52 kilometers (32 miles) long, Ida also appears to have its own magnetic field. Its craters point to an age for Ida of about one billion years.

GASPRA
became the first asteroid to pose for a close-up, when the spacecraft Galileo passed nearby on its way to Jupiter.

Comets

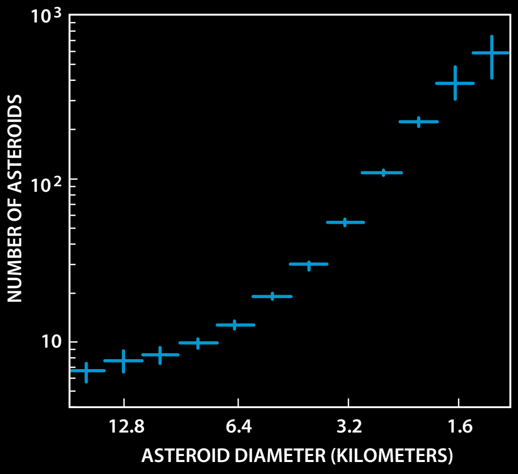


COMET SHOEMAKER-LEVY 9
(*at right*) smashed into Jupiter in July 1994 in the greatest collisions ever witnessed by humanity. Of the more than 20 fragments, moving at 60 kilometers per second, the largest pieces produced energies equivalent to millions of megaton nuclear warheads.

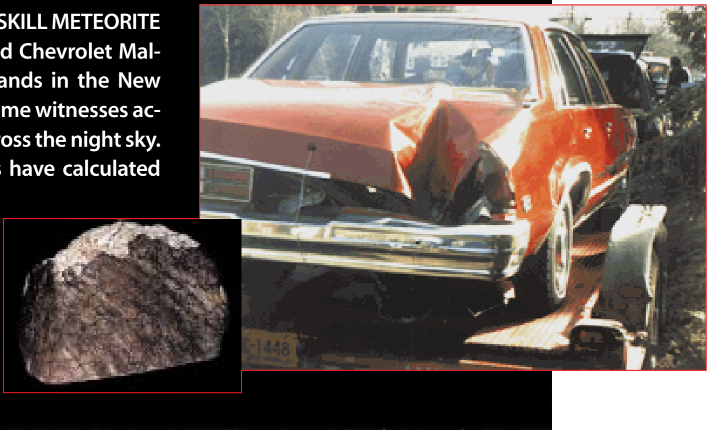
GREAT COMET OF 1680
was determined by Isaac Newton to have an almost parabolic orbit. In our century, in 1995, the Hubble Space Telescope discovered a belt of perhaps 200 million comets encircling the solar system.

JPL/CALTECH/NASA (*photographs at top*); ALFRED T. KAMAJIAN (*illustrations*)

EXTREMELY LARGE ASTEROIDS, with diameters greater than 10 kilometers, are rare. The graph illustrates the relative numbers of small asteroids compared with large ones.



PEEKSKILL METEORITE (*below*) smashed into this parked Chevrolet Malibu on October 9, 1992. Thousands in the New York area saw the fireball, and some witnesses actually videotaped it streaking across the night sky. Using those tapes, astronomers have calculated the trajectory and original orbit of the meteorite. Some even speculate that the meteorite's orbit and composition point to its having originated as part of the asteroid 6 Hebe.



Concentrated between the orbits of Mars and Jupiter float thousands of what astronomers often call minor planets, or asteroids. These might have coalesced to form a small planet had they not been under the immense gravitational influence of Jupiter, which accelerated them. Low-velocity collisions of small bodies can build a planet, but bodies moving at five kilometers per second, the average for asteroids, collide violently. Such collisions can send chunks of asteroids out of their typical orbit between Mars and Jupiter. Some fragments take up stable orbits, part of which brings them closer to Earth or, on occasion, to the surface of our planet as meteorites. Our knowledge of asteroids should increase significantly early in 1999, when a probe called Near Earth Asteroid Rendezvous approaches within 48 kilometers of the asteroid Eros.

HALE-BOPP, the brightest comet since 1811, was clearly visible to the naked eye even in large cities flooded with artificial light. It featured three distinct tails—of dust, ionized gas and sodium atoms.

HALLEY'S COMET visits on regular intervals of about 75 years. Its orbit and the time between visits are slightly variable because of perturbations by the planets Jupiter and Saturn. In the 17th century Edmund Halley analyzed known comet data and discovered the repeat visitor, which now bears his name.

JOHNNY JOHNSON (graph), AFTER ROBERT JEDICKE *University of Arizona*; WALT RADOMSKI *R. A. Langheinreich Meteorites (car, meteorite)*; HUBBLE SPACE TELESCOPE COMET TEAM AND NASA *(Shoemaker-Levy 9)*; BILL WHIDDON AND NINA WHIDDON *(Hale-Bopp)*; ROYAL GREENWICH OBSERVATORY *SPL/Photo Researchers, Inc. (Halley)*

The word "comet," from the Greek, means "long-haired," an apt description for what may appear to be a blur or smudge in the heavens. Visitors from the farthest reaches of the solar system, comets consist of a solid nucleus of dust and ice, which has led them to be called "dirty snowballs." Interactions with the sun produce the nebulous coma and one or more tails that smear the comet against the sky. It was most likely a comet (although an asteroid remains a candidate) that smashed into Earth 65 million years ago, causing the mass extinction that killed the dinosaurs and paved the way for our own evolution.

Calculations by Dutch astronomer Jan Hendrick Oort in the 1950s showed that there must be a huge swarm of comets, since then dubbed the Oort Cloud, some 40,000 to 50,000 times farther away from the sun than is Earth.

II

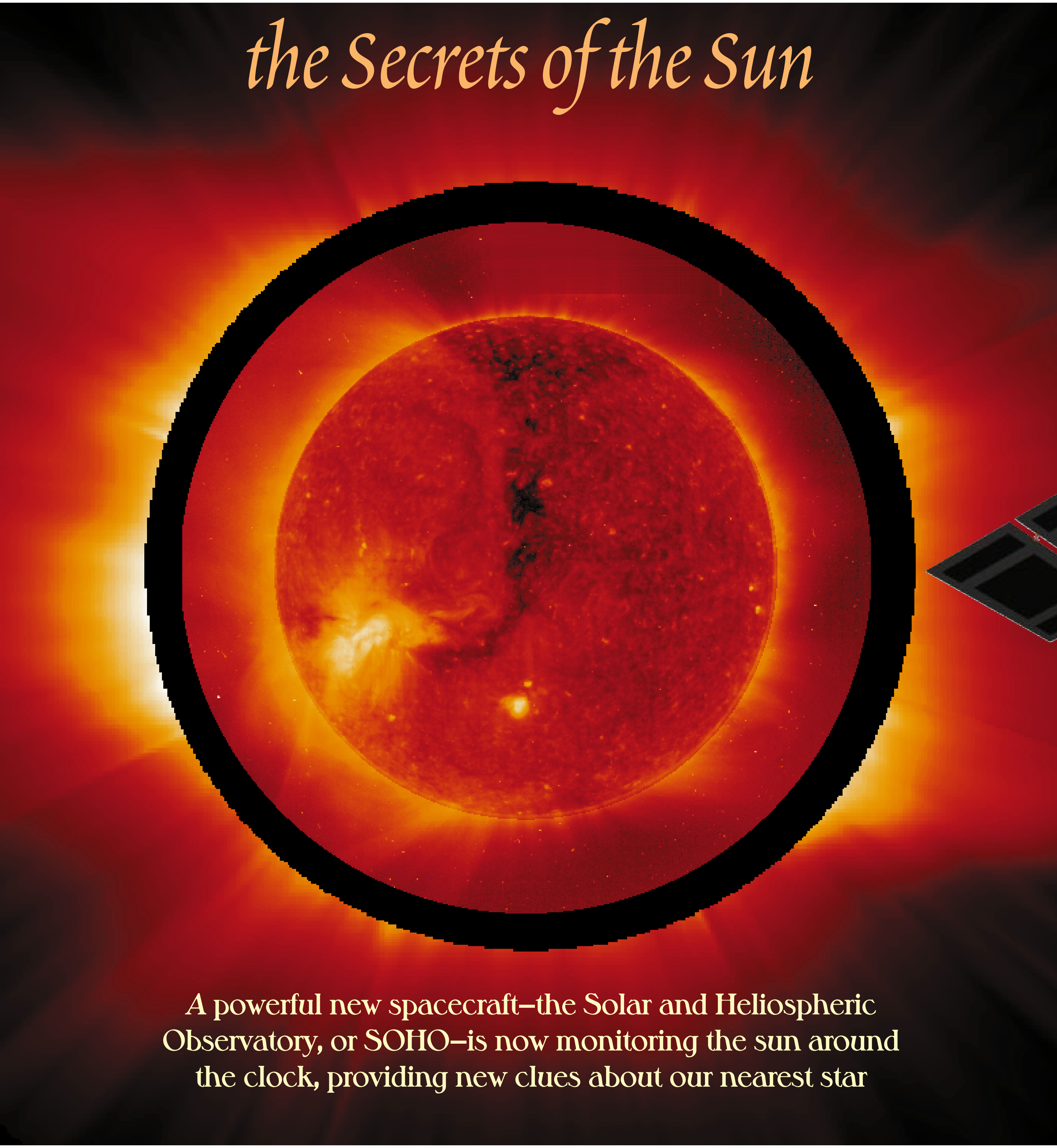
Fire and Light

- SOHO REVEALS THE SECRETS OF THE SUN
- V1974 CYGNI 1992: THE MOST IMPORTANT NOVA OF THE CENTURY
- COSMIC RAYS AT THE ENERGY FRONTIER
- GAMMA-RAY BURSTS
- COLOSSAL GALACTIC EXPLOSIONS
- THE GHOSTLIEST GALAXIES

GIANT TWISTERS
of interstellar gas and dust
in the Lagoon Nebula

A. CAULET (ST-ECF, ESA) AND NASA

SOHO Reveals the Secrets of the Sun



A powerful new spacecraft—the Solar and Heliospheric Observatory, or SOHO—is now monitoring the sun around the clock, providing new clues about our nearest star by Kenneth R. Lang

From afar, the sun does not look very complex. To the casual observer, it is just a smooth, uniform ball of gas. Close inspection, however, shows that the star is in constant turmoil—a fact that fuels many fundamental mysteries. For instance, scientists do not understand how the sun generates its magnetic fields, which are responsible for most solar activity, including unpredictable explosions that cause magnetic storms and power blackouts here on Earth. Nor do they know why this magnetism is concentrated into so-called sunspots, dark islands on the sun's surface that are as large as Earth and thousands of times more magnetic. Furthermore, physicists cannot explain why the sun's magnetic activity varies dramatically, waning and intensifying again every 11 years or so.


ESA AND NASA

To solve such puzzles—and better predict the sun's impact on our planet—the European Space Agency (ESA) and the National Aeronautics and Space Administration launched the two-ton Solar and Heliospheric Observatory (SOHO, for short) on December 2, 1995. The spacecraft reached its permanent strategic position—which is called the inner Lagrangian point and is about 1 percent of the way to the sun—on February 14, 1996. There SOHO is balanced between the pull of Earth's gravity and the sun's gravity and so orbits the sun together with Earth. Earlier spacecraft studying the sun orbited Earth, which would regularly obstruct their view. In contrast, SOHO monitors the sun continuously: 12 instruments examine the sun in unprecedented detail. They downlink several thousand images a day through NASA's Deep Space Network antennae to SOHO's Experimenters' Operations Facility at the NASA Goddard Space Flight Center located in Greenbelt, Md.

At the Experimenters' Operations Facility, solar physicists from around the world work together, watching the sun night and day from a room without windows. Many of the unique images they receive move nearly instantaneously to the SOHO home page on the World Wide Web (located at http://sohowww.nascom.nasa.gov). When these pictures first began to arrive, the sun was at the very bottom of its 11-year activity cycle. But SOHO carries enough fuel to continue operating for a decade or more. Thus, it will keep watch over the sun through all its tempestuous seasons—from the recent lull in magnetic activity to its next maximum, which should take place at the end of the century. Already, though, SOHO has offered some astounding findings.

Exploring Unseen Depths

To understand the sun's cycles, we must look deep inside the star, to where its magnetism is generated. One way to explore these unseen depths is by tracing the in-and-out, heaving motions of the sun's outermost visible surface, named the photosphere from the Greek word *photos*, meaning "light." These oscillations, which can be tens of kilometers high and travel a few hundred meters per second, arise from sounds that course through the solar interior. The sounds are trapped inside the sun; they cannot propagate through the near vacuum of space. (Even if they could reach Earth, they are too low for human hearing.) Nevertheless, when these sounds strike the sun's surface and rebound back down, they disturb the gases there, causing them to rise and fall, slowly and rhythmically, with a period of about five minutes. The throbbing motions these sounds create are imperceptible to the naked eye, but SOHO instruments routinely pick them out.

The surface oscillations are the combined effect of about 10 million separate notes—each of which has a unique path of propagation and samples a well-defined section inside the sun. So to trace the star's physical landscape all the way through—from its churning convection zone, the outer 28.7 percent (by radius), into its radiative zone and core—we must determine the precise pitch of all the notes.

The dominant factor affecting each sound is its speed, which in turn depends on the temperature and composition of the solar regions through which it passes. SOHO scientists compute the expected sound speed using a numerical model. They then use relatively small discrepancies between their computer calculations and the observed sound speed to fine-tune the model and establish the sun's radial variation in temperature, density and composition.

At present, theoretical expectations and observations made with SOHO's Michelson Doppler Imager (MDI) telescope are in close agreement, showing a maximum difference of only 0.2 percent. Where these discrepancies occur is, in fact, significant. They suggest that turbulent material is moving in and out just below the convection zone and hint that such mixing motions might occur at the boundary of the energy-generating core—concepts that could be very important for studies of stellar evolution.

For more than three centuries, astronomers have known from watching sunspots that the photosphere rotates faster at

COMPOSITE IMAGE (*left*), taken by two instruments on board SOHO (*above*) and joined at the black circle, reveals the sun's outer atmosphere from the base of the corona to millions of kilometers above the solar surface. Raylike structures appear in the ultraviolet light emitted by oxygen ions flowing away from the sun to form the solar wind (*outside black circle*). The solar wind with the highest speed originates in coronal holes, which appear as dark regions at the north pole (*top*) and across the solar disk (*inside black circle*).

opposite page: COURTESY OF SOHO EIT CONSORTIUM (*inner region*) AND SOHO UVCS CONSORTIUM (*outer region*)

the equator than at higher latitudes and that the speed decreases evenly toward each pole. SOHO data confirm that this differential pattern persists through the convection zone. Furthermore, the rotation speed becomes uniform from pole to pole about a third of the way down. Thus, the rotation velocity changes sharply at the base of the convection zone. There the outer parts of the radiative interior, which rotates at one speed, meet the overlying convection zone, which spins faster in its equatorial middle. We now suspect that this thin base layer of rotational shear may be the source of the sun's magnetism.

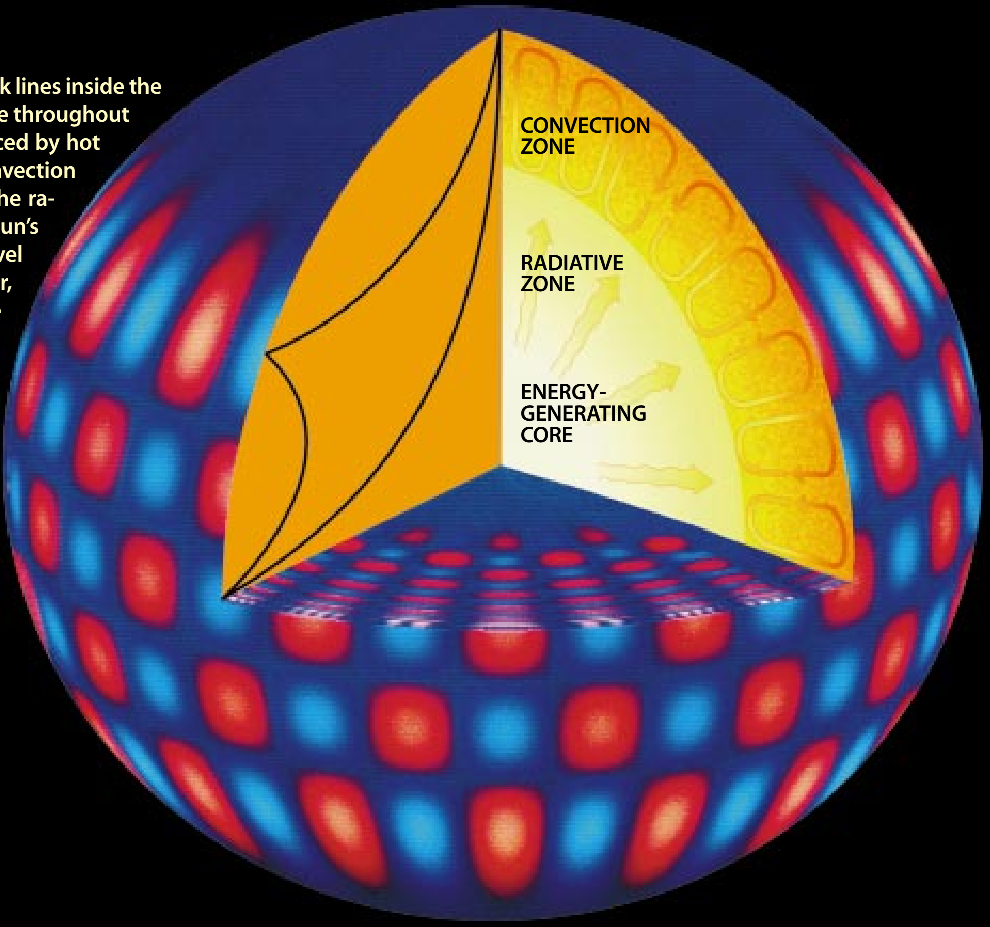
The MDI telescope on board SOHO has also helped probe the sun's outer shells. Because its lenses are positioned well above Earth's obscuring atmosphere, it can continuously resolve fine detail that cannot always be seen from the ground. For this reason, it has proved particularly useful in time-distance helioseismology, a new technique for revealing the motion of gases just below the photosphere. The method is quite straightforward: the telescope records small periodic changes in the wavelength of light emitted from a million points across the sun every minute. By keeping track of them, it is possible to determine how long it takes for sound waves to skim through the sun's outer layers. This travel time tells of both the temperature and gas flows along the internal path connecting two points on the visible solar surface. If the local temperature is high, sound waves move more quickly—as they do if they travel with the flow of gas.

The MDI has provided travel times for sounds crossing thousands of paths, linking myriad surface points. And SOHO scientists have used these data to chart the three-dimensional internal structure and dynamics of the sun, much in the same way that a computed tomographic (CT) scan creates an image of the inside of the brain. They fed the SOHO data to supercomputers to work out temperatures and flow directions along these intersecting paths. Using these techniques during two years of nearly continuous observations, SOHO scientists have discovered vast rivers of hot gas that circulate within the sun.

Completely unexpected currents circle the polar regions of the sun just below the photosphere. They seem to resemble the jet streams high in the atmosphere of Earth, which have a major influence on terrestrial weather. Ringing the sun at about 75 degrees latitude, the solar jet streams are totally inside the sun, 40,000 kilometers (25,000 miles) below the photosphere, and cannot be seen at the visible surface. They move about 10 percent faster than the surrounding gas—about 130 kilometers per hour faster—and they are wide enough to engulf two planet Earths.

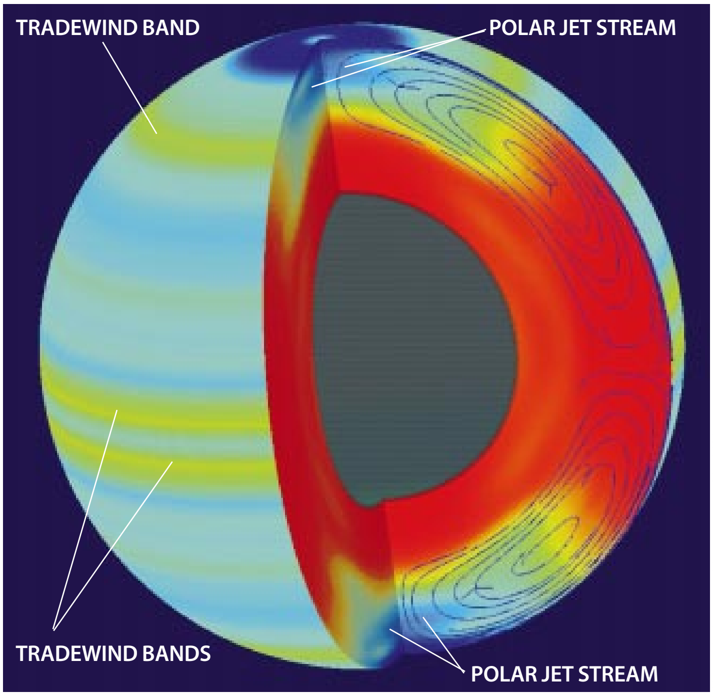
The outer layer of the sun, to a depth of at least 25,000 kilometers, is also slowly flowing from the equator to the poles, at a speed of about 90 kilometers per hour. At this rate, an object would be transported from the equator to the pole in little more than a year. Of course, the sun rotates at a much faster rate of about 7,000 kilometers per hour, completing one revolution at the equator in 25.7 days. The combination of differential rotation and poleward flow has been the explanation for the stretched-out shapes of magnetic regions that have migrated toward the poles. The new SOHO MDI observations demonstrate for the first time that the poleward flow reaches deeply into the sun, penetrating at least 12 percent of the convection zone.

Researchers have also identified internal rivers of gas mov-



SOUND WAVES, represented here by black lines inside the cutaway section, resonate throughout the sun. They are produced by hot gas churning in the convection zone, which lies above the radiative zone and the sun's core. As sound waves travel toward the sun's center, they gain speed and are refracted back out. At the same time, the sun's surface reflects waves traveling outward back in. Thus, the entire star throbs, with regions pulsing in (*red spots*) and out (*blue spots*).

COURTESY OF JACK HARVEY *National Optical Astronomy Observatories*; CROSS SECTIONS BY MICHAEL GOODMAN



SOLAR METEOROLOGY can be seen by looking at internal large-scale flows measured with the MDI instrument on board the SOHO spacecraft from May 1996 to May 1997. Red represents faster than average flows, yellow slower than average and blue slower yet. Yellow bands are deeply rooted zones that move slightly faster than their surroundings; sunspots tend to form at the edges of these zones. The poleward flow is shown as streamlines within the cutaway section. The newly discovered "jet streams" move approximately 10 percent faster than their surroundings.

COURTESY OF PHILIP SCHERRER AND SOHO/SOI TEAM

ing in bands near the equator at different speeds relative to each other in both the northern and southern hemispheres. The solar belts are more than 64,000 kilometers in width and move about 16 kilometers per hour faster than the gases to either side. These broad belts of higher-velocity currents remind one of Earth's equatorial tradewinds and also of Jupiter's colorful, banded atmosphere. The bands are deeply rooted, extending down approximately 19,000 kilometers into the sun. The full extent of the newfound solar meteorology could never have been seen by looking at the visible layer of the solar atmosphere.

The MDI team also investigated horizontal motions at a depth of about 1,400 kilometers and compared them with an overlying magnetic image, also taken by the MDI instrument. They found that strong magnetic concentrations tend to lie in regions where the subsurface gas flow converges. Thus, the churning gas probably forces magnetic fields together and concentrates them, thereby overcoming the outward magnetic pressure that ought to make such localized concentrations expand and disperse.

The Million-Degree Corona

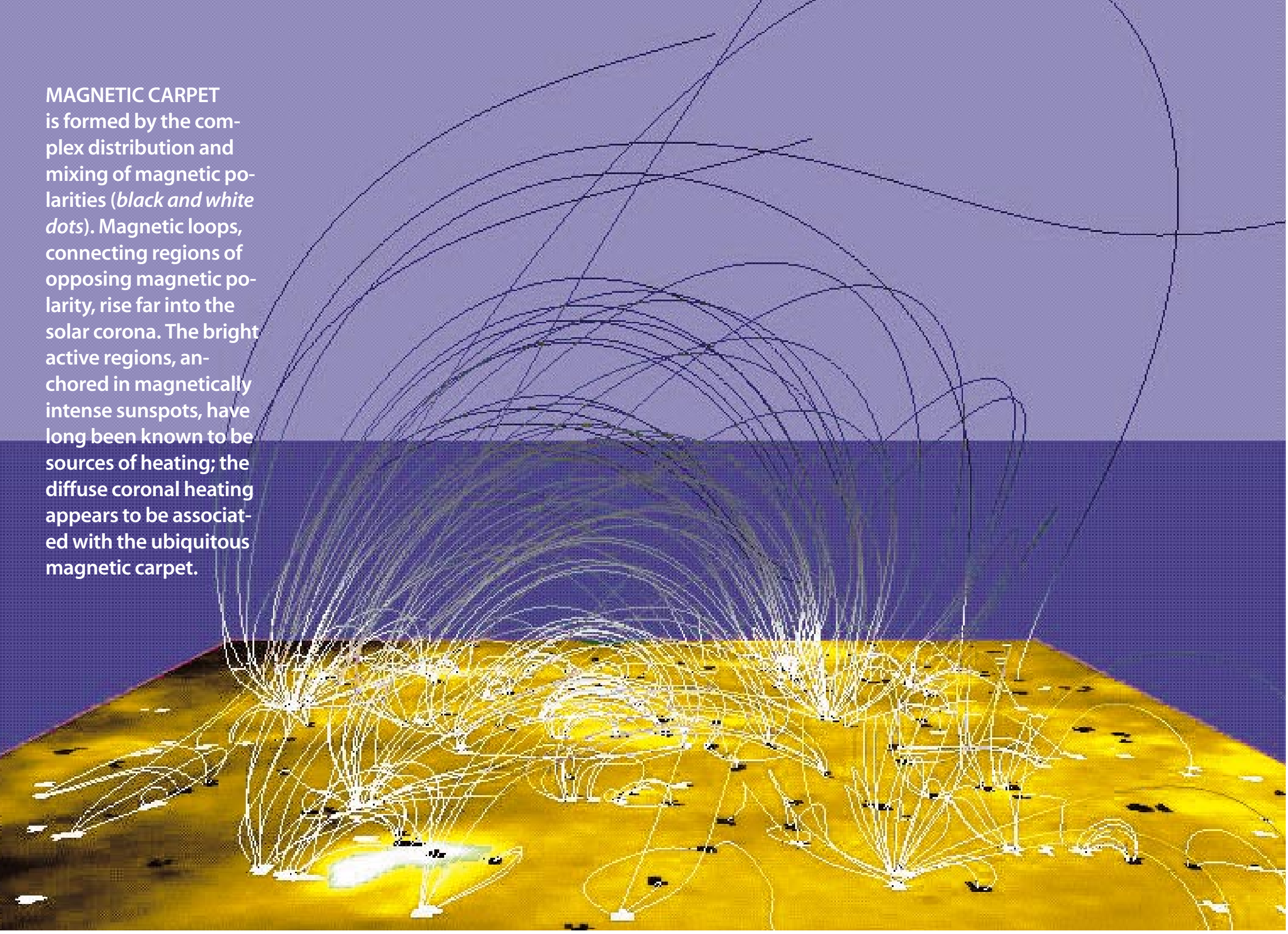
SOHO is also helping scientists explain the solar atmosphere, or corona. The sun's sharp outer rim is illusory. It merely marks the level beyond which solar gas becomes transparent. The invisible corona extends beyond the planets and presents one of the most puzzling paradoxes of solar physics: it is unexpectedly hot, reaching temperatures of more than one million kelvins just above the photosphere; the sun's visible surface is only 5,780 kelvins. Heat simply should not flow outward from a cooler to a hotter region. It violates the second law of thermodynamics and all common sense as well. Thus, there must be some mechanism transporting energy from the photosphere, or below, out to the corona. Both kinetic and magnetic energy can flow from cold to hot regions. So writhing gases and shifting magnetic fields may be accountable.

For studying the corona and identifying its elusive heating mechanism, physicists look at ultraviolet (UV), extreme ultraviolet (EUV) and x-ray radiation. This is because hot material—such as that within the corona—emits most of its energy at these wavelengths. Also, the photosphere is too cool to emit intense radiation at these wavelengths, so it appears dark under the hot gas. Unfortunately, UV, EUV and x-rays are partially or totally absorbed by Earth's atmosphere, and so they must be observed through telescopes in space. SOHO is now measuring radiation at UV and EUV wavelengths using four instruments: the Extreme-ultraviolet Imaging Telescope (EIT), the Solar Ultraviolet Measurements of Emitted Radiation (SUMER), the Coronal Diagnostic Spectrometer (CDS) and the UltraViolet Coronagraph Spectrometer (UVCS).

To map out structures across the solar disk, ranging in temperature from 6,000 to two million kelvins, SOHO makes use of spectral lines. These lines appear when the sun's radiation intensity is displayed as a function of wavelength. The various SOHO instruments locate regions having a specific temperature by tuning into spectral lines emitted by the ions formed there. Atoms in a hotter gas lose more electrons through collisions, and so they become more highly ionized. Because these different ions emit spectral lines at different wavelengths, they serve as a kind of thermometer. We can also infer the speed of the material moving in these regions from the Doppler wavelength changes of the spectral lines that SOHO records.

Ultraviolet radiation has recently revealed that the sun is a vigorous, violent place even when its 11-year activity cycle is in an apparent slump—and this fact may help explain why the corona is so hot. The whole sun seems to sparkle in the UV light emitted by localized bright spots. According to SOHO measurements, these ubiquitous hot spots are formed at a temperature of a million kelvins, and they seem to originate in small, magnetic loops of hot gas found all over the sun, including both its north and south poles. Some of these spots explode and hurl material outward at speeds of hundreds of kilometers per second. SOHO scientists are now studying these bright spots to see if they play an important role in the elusive coronal heating mechanism.

SOHO has provided direct evidence for the transfer of magnetic energy from the sun's visible surface toward the corona above. Images of the photosphere's magnetism, taken with SOHO's MDI, reveal ubiquitous pairs of opposite magnetic polarity, each joined by a magnetic arch that rises above them, like bridges that connect two magnetic islands. Energy flows from these magnetic loops when they interact, producing electrical and magnetic "short circuits." The very strong electric currents in these short circuits can heat the corona to a temperature of several million degrees. Images from the EIT



MAGNETIC CARPET is formed by the complex distribution and mixing of magnetic polarities (*black and white dots*). Magnetic loops, connecting regions of opposing magnetic polarity, rise far into the solar corona. The bright active regions, anchored in magnetically intense sunspots, have long been known to be sources of heating; the diffuse coronal heating appears to be associated with the ubiquitous magnetic carpet.

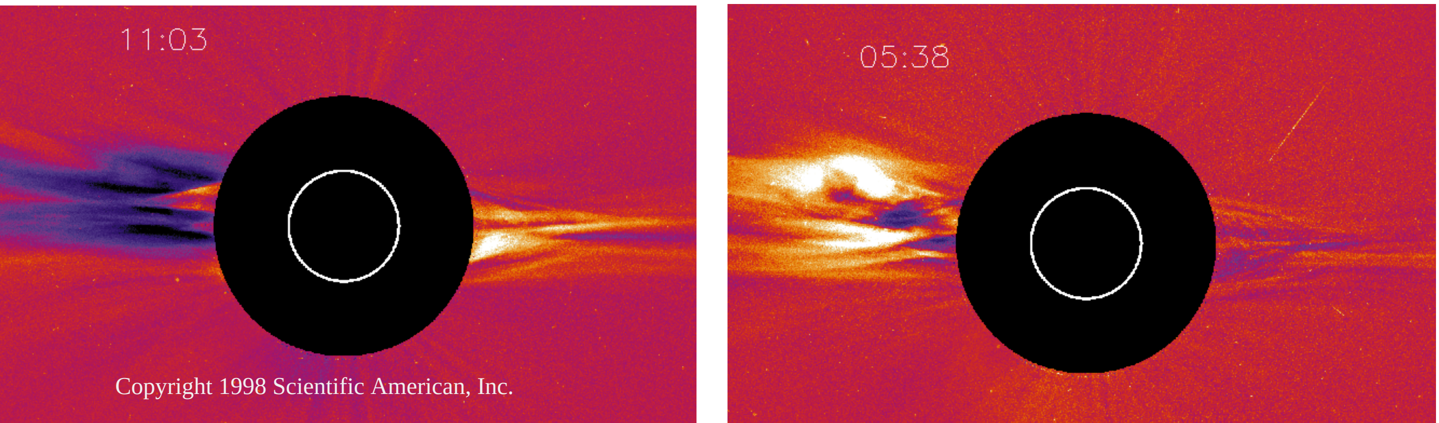
STANFORD-LOCKHEED INSTITUTE FOR SPACE RESEARCH AND NASA GODDARD SPACE FLIGHT CENTER

and CDS instruments on SOHO show the hot gases of the ever-changing corona reacting to the evolving magnetic fields rooted in the solar surface.

To explore changes at higher levels in the sun's atmosphere, SOHO relies on its UVCS and its Large Angle Spectroscopic COronagraph (LASCO). Both instruments use occulting disks to block the photosphere's underlying glare. LASCO detects visible sunlight scattered by electrons in the corona. Initially it revealed a simple corona—one that was highly symmetrical and stable. This corona, viewed during the sun's magnetic lull, exhibited pronounced holes in the north and south. (Coronal holes are extended, low-density, low-temperature regions where EUV and x-ray emissions are abnormally low or absent.)

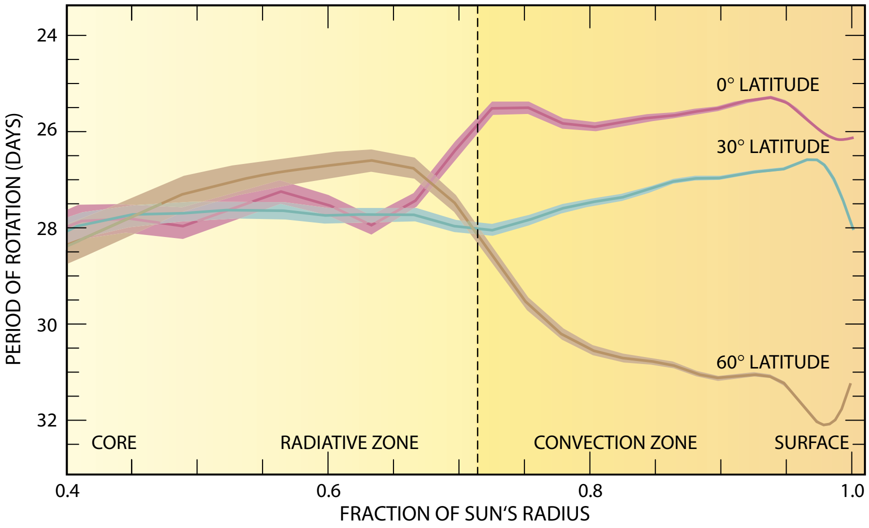
In contrast, the equatorial regions were ringed by straight, flat streamers of outflowing matter. The sun's magnetic field shapes these streamers. At their base, electrified matter is densely concentrated within magnetized loops rooted in the photosphere. Farther out in the corona, the streamers narrow into long stalks that stretch tens of millions of kilometers into space. These extensions confine material at temperatures of about two million kelvins within their elongated magnetic boundaries, creating a belt of hot gas that extends around the sun.

The streamers live up to their name: material seems to flow continuously along their open magnetic fields. Occasionally the coronagraphs record dense concentrations of material moving through an otherwise unchanging streamer—like seeing leaves floating on a moving stream. And sometimes tremendous eruptions, called coronal mass ejections, punctuate the



CORONAL MASS EJECTIONS (*white*), occurring on the east and west sides of the sun, were recorded within hours on the same day by one of SOHO's coronagraphs. The black occulting disk blocks the glare of the sun, whose visible edge is represented here by the white circle.

COURTESY OF SOHO LASCO CONSORTIUM AND GUENTER E. BRUECKNER



MICHAEL GOODMAN, AFTER SOHO SOI/MDI CONSORTIUM AND ALEXANDER G. KOSOVICHEV

INTERNAL ROTATION RATE OF THE SUN at latitudes of zero, 30 and 60 degrees has been inferred using data from the Michelson Doppler Imager. Down to the base of the convection zone, the polar regions spin more slowly than the equatorial ones do. Beyond that, uniform rotation appears to be the norm, although scientists have not yet determined rotation rates within the sun's core.

steady outward flow. These ejections hurl billions of tons of million-degree gases into interplanetary space at speeds of hundreds of kilometers per second. This material often reaches Earth in only two or three days. To almost everyone's astonishment, LASCO found equatorial ejections emitted within hours of each other from opposite sides of the sun.

The coronagraphs have only a side view of the sun and so can barely see material moving to or from Earth. But based on what we can see, we guess that these ejections are global disturbances, extending all the way around the sun. In fact, unexpectedly wide regions of the sun seem to convulse when the star releases coronal mass ejections, at least during the minimum in the 11-year activity cycle. And the coronagraphs have detected that a few days before the ejections, the streamer belt gets brighter, suggesting that more material is accruing there. The pressure and tension of this added material probably build until the streamer belt blows open in the form of an ejection. The entire process is most likely related to a large-scale global reorganization of the sun's magnetic field.

Solar Winds and Beyond

The sun's hot and stormy atmosphere is forever expanding in all directions, filling the solar system with a ceaseless flow—called the solar wind—that contains electrons, ions and magnetic fields. The million-degree corona creates an outward pressure that overcomes the sun's gravitational attraction, enabling this perpetual outward flow. The wind accelerates as it moves away from the sun, like water overflowing a dam. As the corona disperses, it must be replaced by gases welling up from below to feed the wind. Earlier spacecraft measurements, as well as those from Ulysses (launched in 1990), showed that the wind has a fast and a slow component. The fast one moves at about 800 kilometers per second; the slow one travels at half that speed.

The slow component is associated with equatorial regions of the sun, now being scrutinized by LASCO and UVCS. These instruments suggest that the slow component of the solar wind flows out along the stalklike axes of equatorial coronal streamers. The high-speed component pours forth from the polar coronal holes. (Open magnetic fields there allow charged particles to escape the sun's gravitational and magnetic grasp.) SOHO is now investigating whether polar plumes—tall structures rooted in the photosphere that extend into the coronal holes—help to generate this high-speed solar wind.

SOHO's UVCS has examined the spectral emission of hydrogen and heavily charged oxygen ions in the regions where the corona is heated and the solar wind accelerates. And these spectral-line profiles have produced surprising results, revealing a marked difference in the agitation speeds at which hydrogen and oxygen ions move. In polar coronal holes, where the fast solar wind originates, the heavier oxygen is far more agitated, with about 60 times more energy of motion; above two solar radii from the sun's center, oxygen has the higher agitation speed, approaching 500 kilometers per second. Hydrogen, on the other hand, moves at only 250 kilometers per second. In contrast, within equatorial regions, where the slow-speed wind begins, the lighter hydrogen moves faster than the oxygen, as one would expect from a heat-driven wind.

Researchers are now trying to determine why the more massive oxygen ions move at greater speeds in coronal holes. One possibility is that the ions are whirling around magnetic-field lines that stretch from the sun. Information about the heating and acceleration processes is probably retained within the low-density coronal holes, wherein ions rarely collide with electrons. Frequent collisions in high-density streamers might erase any signature of the relevant processes.

SOHO has obtained marvelous results to date. It has revealed features on the mysterious sun never seen before or never seen so clearly. It has provided new insights into fundamental unsolved problems, all the way from the sun's interior to Earth and out to the farthest reaches of the solar wind. Some of its instruments are now poised to resolve several other mysteries. Two of them will soon have looked at the solar oscillations long enough, and deep enough, to determine the temperature and rotation at the sun's center. Moreover, during the next few years, our home star's inner turmoil and related magnetic activity—which can directly affect our daily lives—will increase. SOHO should then offer even greater scientific returns, determining how its threatening eruptions and hot, gusty winds originate and perhaps predicting conditions in the sun's atmosphere.

The Author

KENNETH R. LANG is professor of astronomy at Tufts University. His recent illustrated book, *Sun, Earth and Sky,* describes all aspects of the sun and its interactions with Earth. Lang has also written more than 150 professional articles and four additional books, which have been translated into seven languages. Among them is the classic reference *Astrophysical Formulae.* This article updates a version that appeared in *Scientific American* in March 1997.

V1974 Cygni 1992: The Most Important Nova of the Century

by Sumner Starrfield and Steven N. Shore

Never has a nova been watched by so many astronomers with so many instruments. Since its discovery by Peter Collins, an amateur astronomer in Boulder, Colo., in the early morning of February 19, 1992, nova V1974 Cygni has been recorded in x-rays through radio waves and from the ground, the air, Earth orbit and beyond.

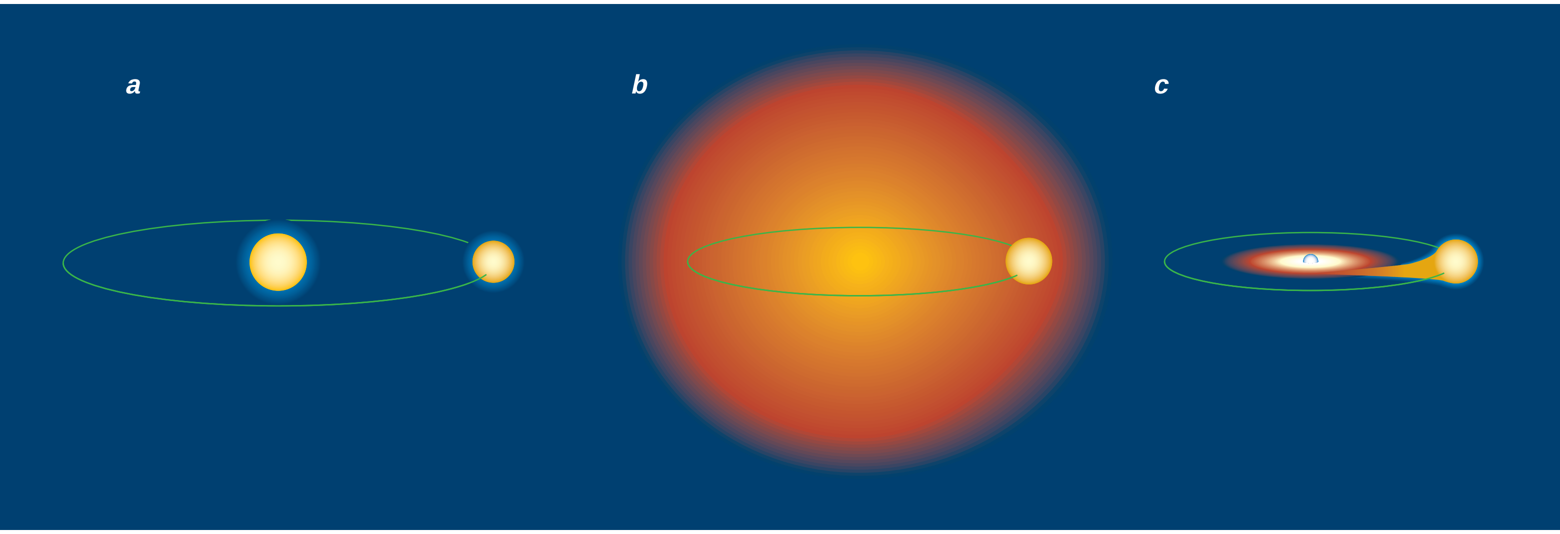
Within hours of his report, we looked at the nova with the International Ultraviolet Explorer (IUE) satellite. We caught it in the "fireball" stage—familiar from photographs of hydrogen bomb explosions, when the gases are first expanding. Before long, it became the only nova to be seen both in birth and in death. In late 1993 the low-energy x-rays coming from the nova's core ceased, indicating to us that the nuclear explosion had run out of fuel.

This nova answered many questions during its life and raised more in death

V1974 Cygni 1992 confirmed many of our ideas about novae—such as how the ejected gases evolve—but also presented new challenges. It threw off about 10 times more matter than was expected, part of it in the form of dense knots and filaments. The knots may hold the key to the cause of the excess mass. They point to turbulent processes that dredged up material from the nova's core. Although we have been forced to rethink many details of how novae evolve, the most essential elements of the original picture remain intact.

In 1892 the nova T Aurigae became the first to be recognized as an explosion, from the peculiarities in its spectrum as compared with normal stars. Since then, scientists have found and studied one or two novae each year. A "naked eye" nova, such as V1974 Cygni, bright enough to be easily visible to the unaided eye, appears perhaps once in a decade.

About 40 years ago a picture of how novae occur began to



fall into place. In 1954 Merle F. Walker, then at the Mount Wilson and Palomar Observatories, discovered that the old nova DQ Herculis (which exploded in 1934) is a system of two orbiting stars. One of the stars in the binary system conveniently passes in front of the other, allowing astronomers to measure the time the two stars take to orbit each other. The period turns out to be extraordinarily short—four hours and 39 minutes. One star is also very small; we now know it to be a white dwarf.

White dwarfs, the end product of stellar evolution, have as much matter as the sun within a volume no larger than Earth's. Robert P. Kraft, also then at the Mount Wilson and Palomar Observatories, showed that other old novae are closely orbiting binary systems. In all these novae, one star was relatively large and unevolved, and the other was a white dwarf. But how can a white dwarf that has no remaining nuclear fuel, along with a stable companion star, give rise to an explosion 10,000 times brighter than the sun? It turns out that each star inexorably alters the other's development.

Calamitous Company

A nova system begins as a widely separated binary, in which one star is more massive than the other. The massive star evolves faster, fusing its hydrogen into helium through the "CNO" cycle of nuclear reactions, which involves carbon, nitrogen and oxygen. At the end of this stage the star becomes a red giant. Its surface swells, engulfing the smaller star. Meanwhile the more massive star fuses the helium in its core to carbon and oxygen.

The stars continue to orbit each other within the common gaseous envelope, losing orbital energy and angular momentum to the gas. As a result, gas is expelled from the system, and the two stars spiral in toward each other. Eventually all the material extending from the massive star past the smaller star is lost. At the end of this "common envelope" evolution, the distantly orbiting stars have become a close binary system. The massive star, having used up all its fuel, has transformed into a compact white dwarf. Its companion has remained relatively unchanged.

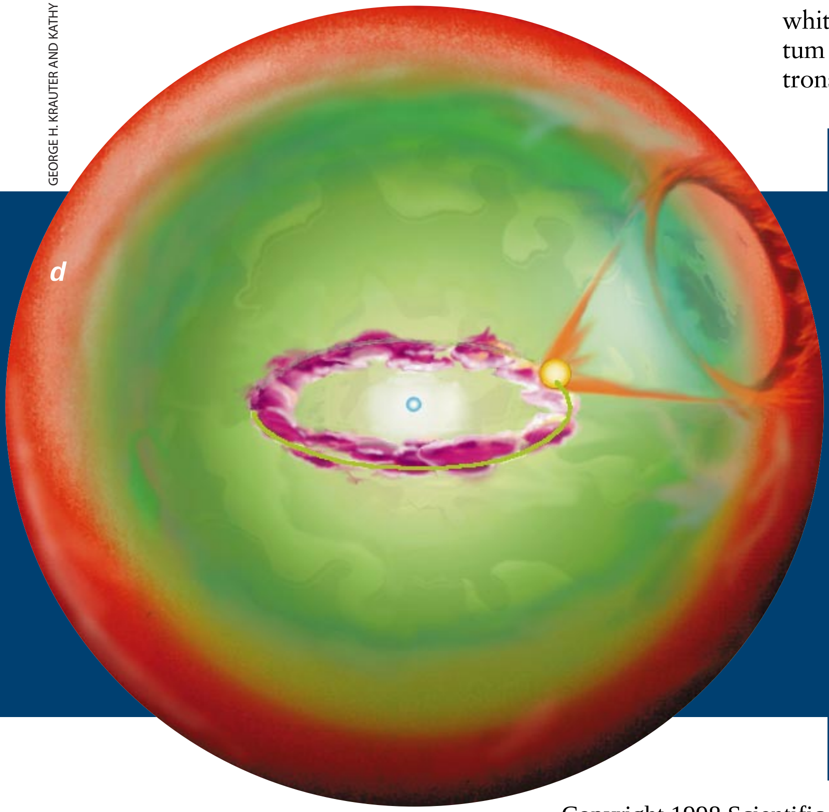
Suppose the stars are initially even more widely separated, and the more massive star began its life with about eight to 12 solar masses. Then the latter star can further fuse its core carbon into magnesium and neon. The white dwarf it ultimately becomes is made of these heavier elements, rather than just carbon and oxygen (a CO nova), and is called an ONeMg white dwarf.

Kraft also made the crucial discovery that the companion star is losing gas. After swirling around in an accretion disk, the hydrogen-rich gas falls onto the surface of the white dwarf. In 1972 one of us (Starrfield, then at the IBM Thomas J. Watson Research Center), along with Warren M. Sparks, then at the National Aeronautics and Space Administration Goddard Space Flight Center, James W. Truran, then at Yeshiva University, and G. Siegfried Kutter, then at the University of Virginia, developed computer simulations that showed how the accreted gas triggers the subsequent explosion.

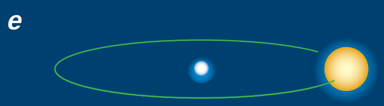
The intense gravity on the white-dwarf surface compresses the gas as it falls in. If an amount of gas 100 times more massive than Earth accumulates on the white dwarf's surface, then the density in the bottom layer becomes more than 10,000 grams per cubic centimeter. (The density of water is one gram per cubic centimeter.) Because the gas is compressed, its temperature rises to a few million kelvins. The process of accumulation also mixes material from the core of the white dwarf into the overlying and infalling layers, thereby changing their composition.

Under these conditions, the hydrogen nuclei fuse into helium and release energy, by the same CNO nuclear reactions that power normal stars more massive than the sun. The material becomes even hotter, so that the fusion proceeds faster, creating runaway thermonuclear reactions like those in a hydrogen bomb.

If the gas were normal, then it would now expand and cool, thereby shutting off the fusion reactions. But the material on a white dwarf behaves in a peculiar manner described by quantum mechanics. It is packed together so tightly that the electrons, which are unable to interpenetrate, become the source

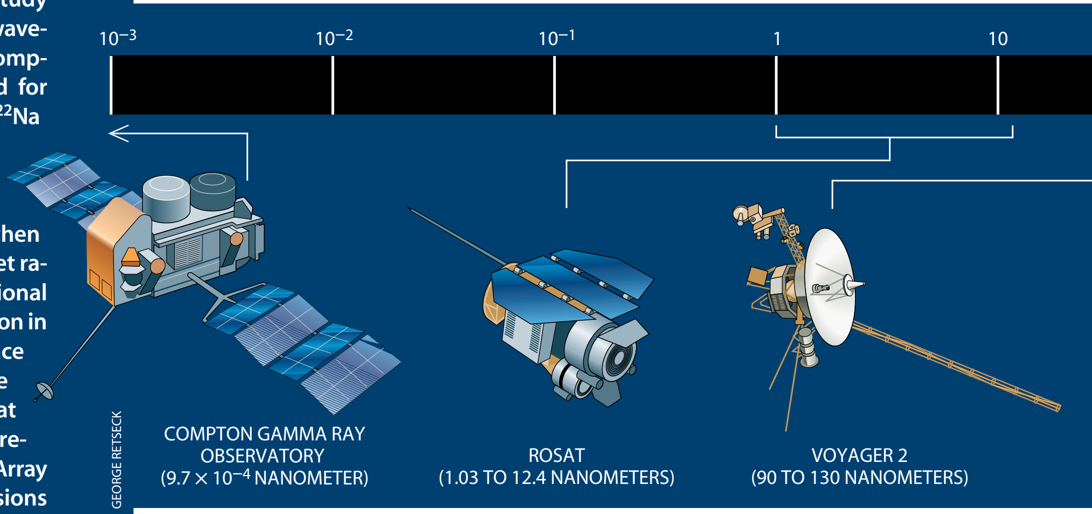


GEORGE H. KRAUTER AND KATHY KONKLE



NOVA SYSTEM begins as a pair of widely separated orbiting stars (*a*). The more massive star evolves faster, becoming a red giant and enveloping the smaller star (*b*). The stars lose angular momentum to the gas and spiral in toward each other while the gas is expelled. Eventually they form a close binary system in which the remaining core of the red giant, having used up all its fuel, has become a white dwarf. The less massive star now sheds matter, which first forms an accretion disk (*c*). Falling onto the white dwarf's surface, this material is compressed by the high gravitational field. Then a runaway thermonuclear reaction—a nova explosion—takes place (*d*), stripping most of the accreted material off the white dwarf (*e*). It can, however, accrete fuel from its neighbor again, cycling through the steps from *c* to *e* many times.

DIVERSE INSTRUMENTS were used to study electromagnetic radiation of different wavelengths emitted by V1974 Cygni. The Compton Gamma Ray Observatory searched for photons emitted by the sodium isotope ^{22}Na (and found none). The ROSAT satellite detected x-rays coming from the burning core; the cessation of these rays signaled the nova's demise. Voyager 2, then beyond Neptune, observed far ultraviolet radiation, the first for a nova. The International Ultraviolet Explorer captured the explosion in its early fireball stage. The Hubble Space Telescope revealed clumps within the ejected gases. The 1.8-meter telescope at Lowell Observatory in Flagstaff, Ariz., recorded optical light, and the Very Large Array in Socorro, N.M., detected radio emissions that confirmed the presence of clumps.



of pressure. Unlike an ordinary gas, the material heats up but cannot expand and cool. Nor can radiation carry away the heat fast enough.

The carbon and oxygen mixed in from the core catalyze the CNO cycle and thus speed up the fusion reactions. The rates of the nuclear reactions also depend very sensitively on the temperature, becoming 10^{16} to 10^{18} times faster when the temperature increases by a factor of 10. As the temperature deep within the accreted layers grows to more than 30 million kelvins, the material starts mixing turbulently with the zones above. The mixed region grows toward the surface, carrying with it both heat and nuclei from the interior. Within minutes the surface layers explode into space. They carry along with them fusion products and elements from the dwarf's core, accompanied by a tremendous increase in brightness.

Burning Out

The first few minutes of a nova explosion have never been observed. Our simulations predict that the surface temperature can exceed one million kelvins and that the hot gases are blown away at more than 5,000 kilometers (3,000 miles) per second. Because its volume increases suddenly, the gas cools. In a few hours the radiation it emits shifts from being primarily in x-rays to the lower-energy ultraviolet. At the same time, the surface area of the gas increases, making the nova brighter even as it becomes cooler. A spectacular transformation ensues.

Initially, the expanding shell consists of a hot, dense gas of electrons and ions—atoms missing one or more electrons. This gas is reasonably transparent. But as it expands, its temperature drops below 10,000 kelvins. The electrons start to recombine with the ions to form atoms that are missing only a few electrons. These atoms have many energy levels and can absorb tens of millions of individual wavelengths of light.

The most important absorbers have atomic numbers around 26, that of iron. The spectrum of light that they can absorb is extremely complex. These ions and atoms block most of the energy being radiated in the ultraviolet, which is where most of the energy is emitted at this phase. When we first studied this phase, with Peter H. Hauschildt, then at Arizona State University, and other collaborators, we called it the iron curtain. The energy absorbed by the curtain is reemitted at longer—optical and infrared—wavelengths.

The iron curtain was vividly confirmed by our first observations of V1974 Cygni. Within hours of its discovery George Sonneborn of the NASA Goddard Space Flight Center activated our Target of Opportunity program, which allows us to observe immediately with the IUE satellite when a bright nova occurs. Pointing the wonderfully maneuverable satellite at the nova, he obtained a series of ultraviolet spectra.

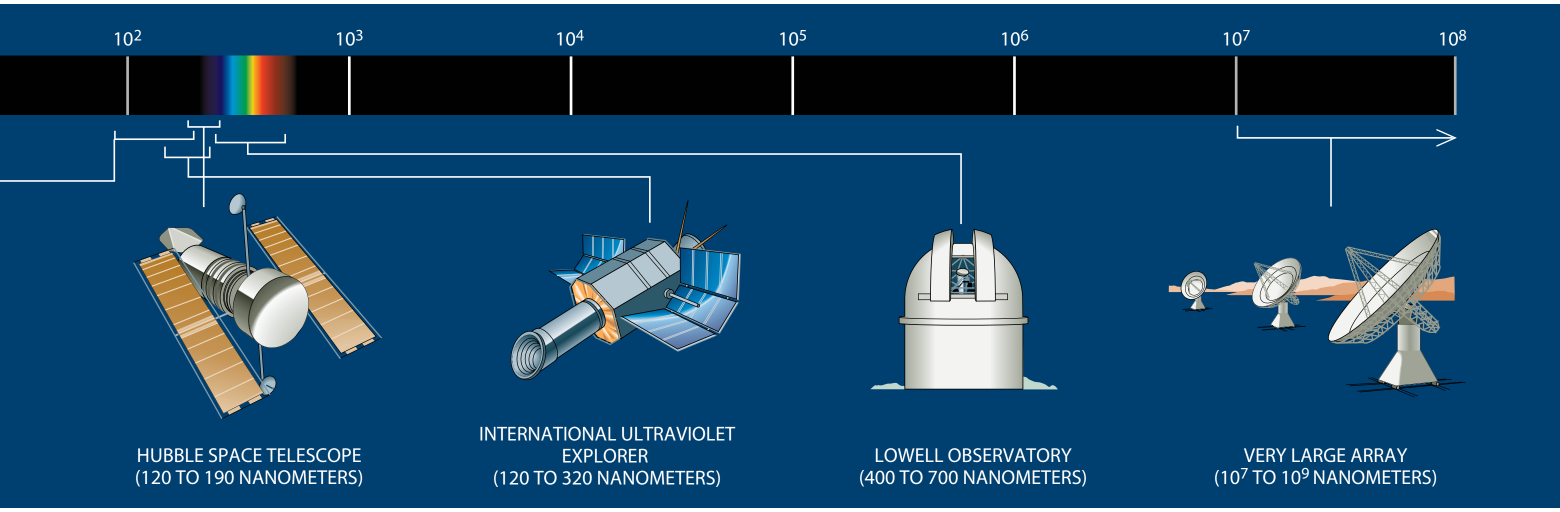
Within an hour we were able to observe that the nova's ultraviolet brightness had dropped slightly and that its optical brightness had risen. Astronomical change is measured, as a rule, in billions of years; it is rare to see evolution on such a short timescale. During the next two days, the ultraviolet radiation dropped to 3 percent of its original value. All the while the nova became optically brighter. As soon as the visual brightness peaked, the ultraviolet emissions bottomed out and began to climb.

The recovery comes from a second change in ionization. As the gas expands, its density drops. Then the iron group elements once again become ionized and hence transparent. Radiation now flows from the interior, enhancing the ionization and in turn the transparency. In effect, the iron curtain lifts, and ultraviolet light from the hot, deep layers penetrates through the outer layers. Within two months the ultraviolet brightness had climbed back up to its original value.

At the same time as the ultraviolet brightness increased, the visual brightness of the nova declined. The total (bolometric) brightness of the underlying star remained, however, virtually unchanged. This constant bolometric luminosity phase, predicted by our 1972 simulations, was finally confirmed in detail by observations of V1974 Cygni.

Anticipating that the radiation peak would continue toward shorter wavelengths, Ronald S. Polidan of the NASA Goddard Space Flight Center requested that Voyager 2, then flying beyond the orbit of Neptune, observe the spectra of V1974 Cygni. On April 27, 1992, the spacecraft detected the nova—the first to be seen in the far ultraviolet. Its brightness in this wavelength range increased during the observations.

The radiation peak continued to shift into shorter wavelengths. Using the ROSAT satellite, Joachim Krautter of Heidelberg Observatory, Hakki Ögelman of the University of Wisconsin and Starrfield had started observing the nova on April 22, 1992. The x-ray spectrum was very faint but included very high



energy photons. (We do not as yet know where the highest-energy photons come from.) Over the next year the x-ray brightness of V1974 Cygni steadily increased, mainly at low energies.

It seemed that a new source of x-rays had appeared, and it was steadily brightening. We realized that we were seeing through the thinning shell of ejected gas to the hot underlying white dwarf. Within three months the nova had become the brightest source of low-energy x-rays in the sky.

Such x-ray sources (called SSS, for supersoft sources) probably stay on for decades. To our surprise, the nova rapidly began to fade during the summer of 1993 and by December had become undetectable with ROSAT.

Fortunately, we were able to keep observing with the IUE. We found that the amount of highly ionized nitrogen was declining, which meant that the ions were recombining with electrons to form less ionized atoms. Furthermore, nitrogen ions that were missing four electrons were recombining faster than were the ions missing three electrons. Apparently the intense radiation that had been stripping the nitrogen of its electrons had vanished: the x-rays were indeed gone. To us, this absence could mean only that the white dwarf had consumed all its fuel and that the nuclear fusion on its surface had ceased.

The nova outburst had lasted about 18 months. The life span of a nova depends on the mass of the white dwarf that hosts it. A massive white dwarf compresses the accumulated gases more intensely. In that case, fusion starts early, and the fuel runs out quickly, causing the nova's life to be brief. Also, the explosion ejects much less matter than does one on a low-mass white dwarf. According to our models, the short life of V1974 Cygni implies that its mass was 20 to 30 percent greater than that of the sun. Theory predicts that about 10^{-5} solar mass should have been ejected from this kind of white dwarf. The observations imply that V1974 Cygni lost about five times this amount.

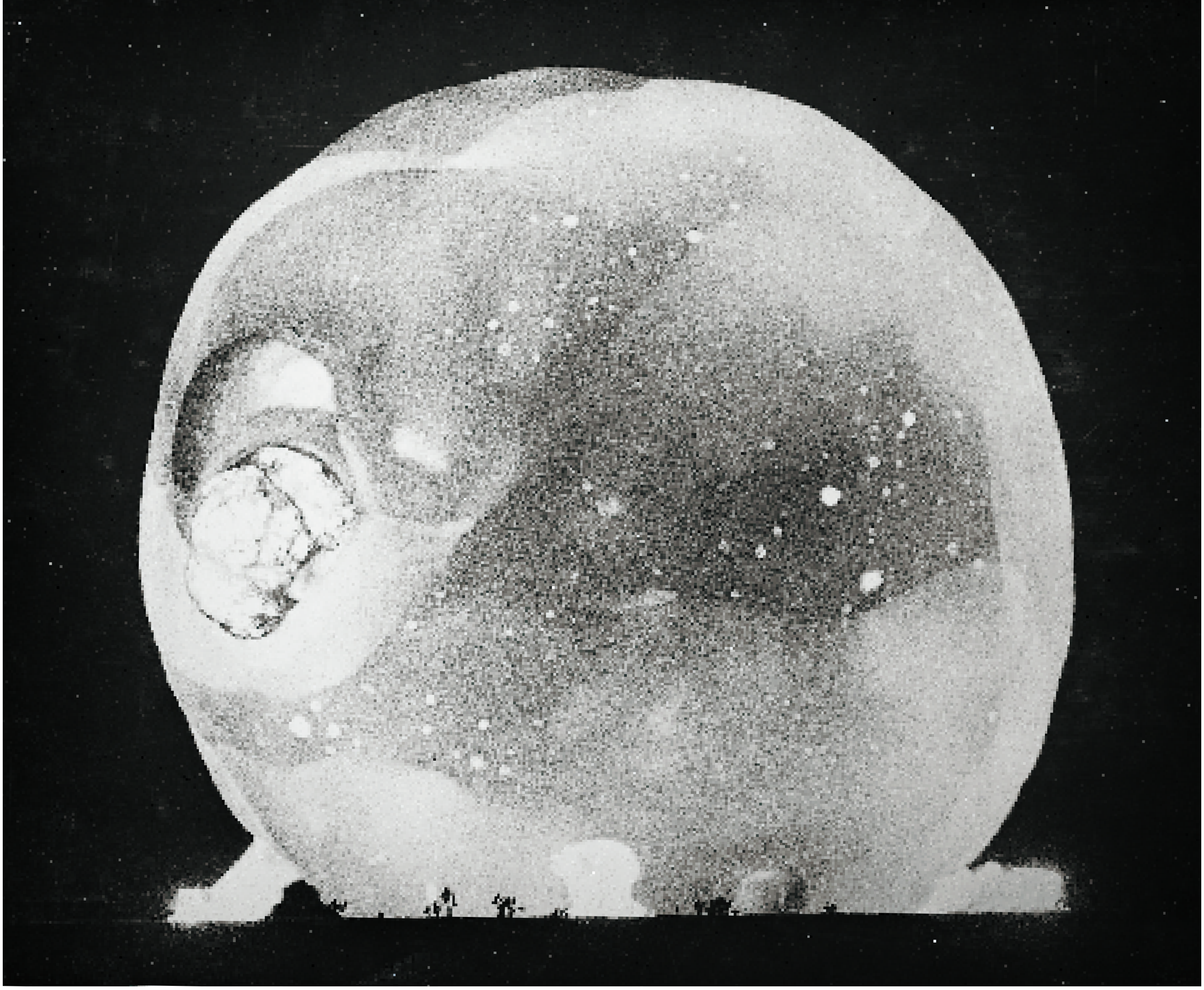
Clumpy Clues

Some hints to the explanation for this discrepancy may possibly be found in the knots. Our first clear view of the knots was on September 7, 1992, when we observed the nova with the Goddard High Resolution Spectrograph (GHRS) on the Hubble Space Telescope. With this powerful instrument we obtained the highest-quality ultraviolet spectra ever for a nova. Each emission line showed evidence that the gas had been expelled in two stages. There was high-velocity gas that had been ejected uniformly and denser, slower-moving clumps.

Armed with the high-quality GHRS spectra, we reexamined the earlier data from the IUE. The spectra we had taken just after the iron curtain cleared also displayed the knots. This evidence indicates that the structures had been formed during the explosion. Looking again with the GHRS on April 1, 1993, we found the same clumps we had identified earlier, moving at the same speeds. The faster material had largely vanished, so we were now seeing completely through the ejected gas. This is the first time we have ever had such a clear view so early of the debris from a nova explosion. The knots appear to be deeply embedded within the ejecta. Now we need to understand what caused them and what they are made of.

The first indications of the composition of the ejecta had come around April 1, 1992, when the iron curtain finally lifted, leaving an intense spectrum with bright emission lines from carbon, nitrogen, oxygen and other abundant elements. Previously, we had encountered emission lines of this kind only in novae that took place on massive ONeMg white dwarfs. We conjectured that V1974 Cygni, too, belonged to this class. The idea also occurred to Thomas L. Hayward of Cornell University and Robert D. Gehrz of the University of Minnesota and their collaborators, who had just obtained infrared spectra of the nova using the five-meter telescope on Mount Palomar. They found the characteristic 12-micron line emitted by ionized neon. This line is normally very weak or absent in CO novae but is strong in ONeMg novae.

In the fall of 1993 the gases thinned enough so that Scott Austin of Arizona State University, R. Mark Wagner of Ohio State University and the two of us could finally use the optical and ultraviolet spectra to determine the chemical abundances of the debris. (While the gas was dense, the atoms collided with one another, thus complicating the spectra greatly.) We found large quantities of elements from the core, such as oxygen and neon. In our last observations with the GHRS, in September 1995, we directly observed one of the knots and confirmed that it contained at least 15 times more neon than solar material. We were also able to directly observe the white dwarf and confirm that it had turned off. These results demonstrate that we were indeed seeing core material from a white dwarf. In no other astronomical object can we see core gases blown into space where they can be studied and provide data on stellar death.



HAROLD EDGERTON

FIREBALL billows out a fraction of a second after an atomic bomb explosion at a Nevada test site. Its structure is very similar to the fireball from a nova. This photograph from the 1950s was taken by automatic instruments situated 32 kilometers away. In the foreground are Joshua trees, about to be incinerated. The intense heat melted desert sand into glass.

Another, related mystery pertains to the elements synthesized during the explosion. Achim Weiss of the Max Planck Institute for Astrophysics in Garching, Irit Idan and Giora Shaviv of the Technion University in Israel, Truran and Starrfield have calculated that ^{22}Na, an isotope of sodium with mass number 22, should be produced in an ONeMg nova. This isotope is radioactive, with a distinct pattern of gamma-ray emissions.

Our calculations indicate that V1974 Cygni produced large amounts of ^{22}Na. With the Compton Gamma Ray Observatory, we searched for the appropriate gamma rays in September 1993—but found none.

All these anomalies tell us that although we have come a long way in understanding nova explosions, we still have much to learn. We understand the thermonuclear reactions that produced the explosion. What is not so clear is the dynamics. Do the shell and the core mix while material is being accreted or during the last stages of the explosion?

Another mystery is the long-term effect of repeated nova outbursts on the evolution of the white dwarf. All nova binary systems go through the cycle of accretion and explosion many times. If parts of the core are shed during each outburst, then the mass of the white dwarf must be decreasing with repeated explosions. Does its mass become ultimately very small, or does something happen to stop any further outbursts?

Because of the brightness and slow evolution of its debris, we will be observing nova V1974 Cygni well into the 21st century. We hope the nova will supply some answers to the questions it has raised. SA

The Authors

SUMNER STARRFIELD and STEVEN N. SHORE have a particular fascination with stellar explosions. Starrfield obtained his Ph.D. at the University of California, Los Angeles, and since 1972 has been a professor at Arizona State University. Starrfield was the principal investigator for observing novae with the International Ultraviolet Explorer and the Compton Gamma Ray Observatory satellites. Shore received his Ph.D. in 1978 from the University of Toronto and now chairs the department of physics and astronomy at Indiana University, South Bend. He observed V1974 Cygni with the Goddard High Resolution Spectrograph, which was on board the Hubble Space Telescope. Shore serves on the editorial board of the *Encyclopedia of Physical Science and Technology* and is a scientific editor of the *Astrophysical Journal*. This article updates a version that appeared in *Scientific American* in January 1995.

Cosmic Rays at the Energy Frontier

These particles carry more energy than any others in the universe. Their origin is unknown but may be relatively nearby

by James W. Cronin, Thomas K. Gaisser and Simon P. Swordy

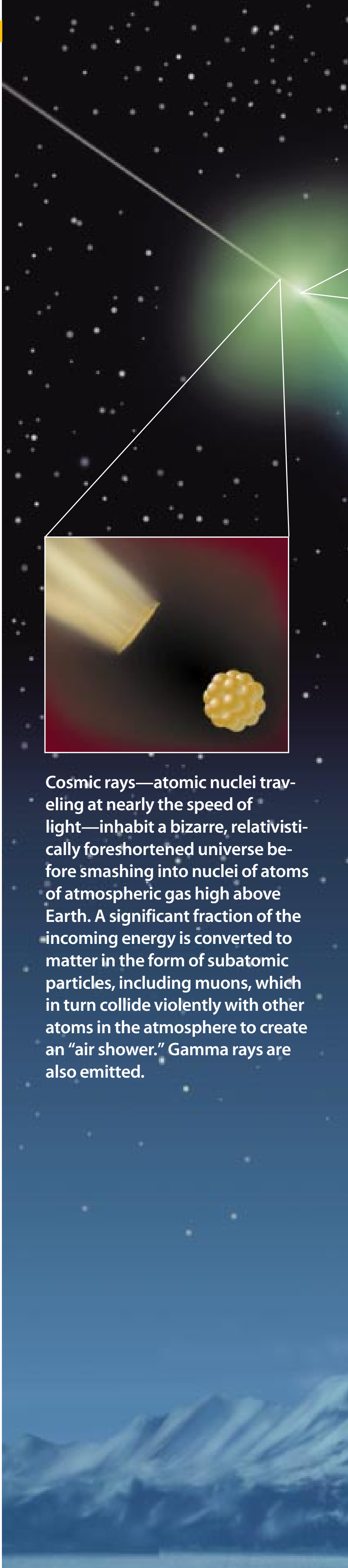
Roughly once a second, a subatomic particle enters Earth's atmosphere carrying as much energy as a well-thrown rock. Somewhere in the universe, that fact implies, there are forces that can impart to a single proton 100 million times the energy achievable by the most powerful Earthbound accelerators. Where and how?

Those questions have occupied physicists since cosmic rays were first discovered in 1912 (although the entities in question are now known to be particles, the name "ray" persists). The interstellar medium contains atomic nuclei of every element in the periodic table, all moving under the influence of electrical and magnetic fields. Without the screening effect of Earth's atmosphere, cosmic rays would pose a significant health threat; indeed, people living in mountainous regions or making frequent airplane trips pick up a measurable extra radiation dose.

Perhaps the most remarkable feature of this radiation is that investigators have not yet found a natural end to the cosmic-ray spectrum. Most well-known sources of charged particles—such as the sun, with its solar wind—have a characteristic energy limit; they simply do not produce particles with energies above this limit. In contrast, cosmic rays appear, albeit in decreasing numbers, at energies as high as astrophysicists can measure. The data run out at levels around 300 billion times the rest-mass energy of a proton because there is no detector large enough to sample the very low number of incoming particles predicted.

Nevertheless, evidence of ultrahigh-energy cosmic rays has been seen at intervals of several years as particles hitting the atmosphere create myriad secondary particles (which are easier to detect). On October 15, 1991, for example, a cosmic-ray observatory in the Utah desert registered a shower of secondary particles from a 50-joule (3×10^{20} electron volts) cosmic ray. Although the cosmic-ray flux decreases with higher energy, this decline levels off somewhat above about 10^{18} eV, suggesting that the mechanisms responsible for ultrahigh-energy cosmic rays are different from those for rays of more moderate energy.

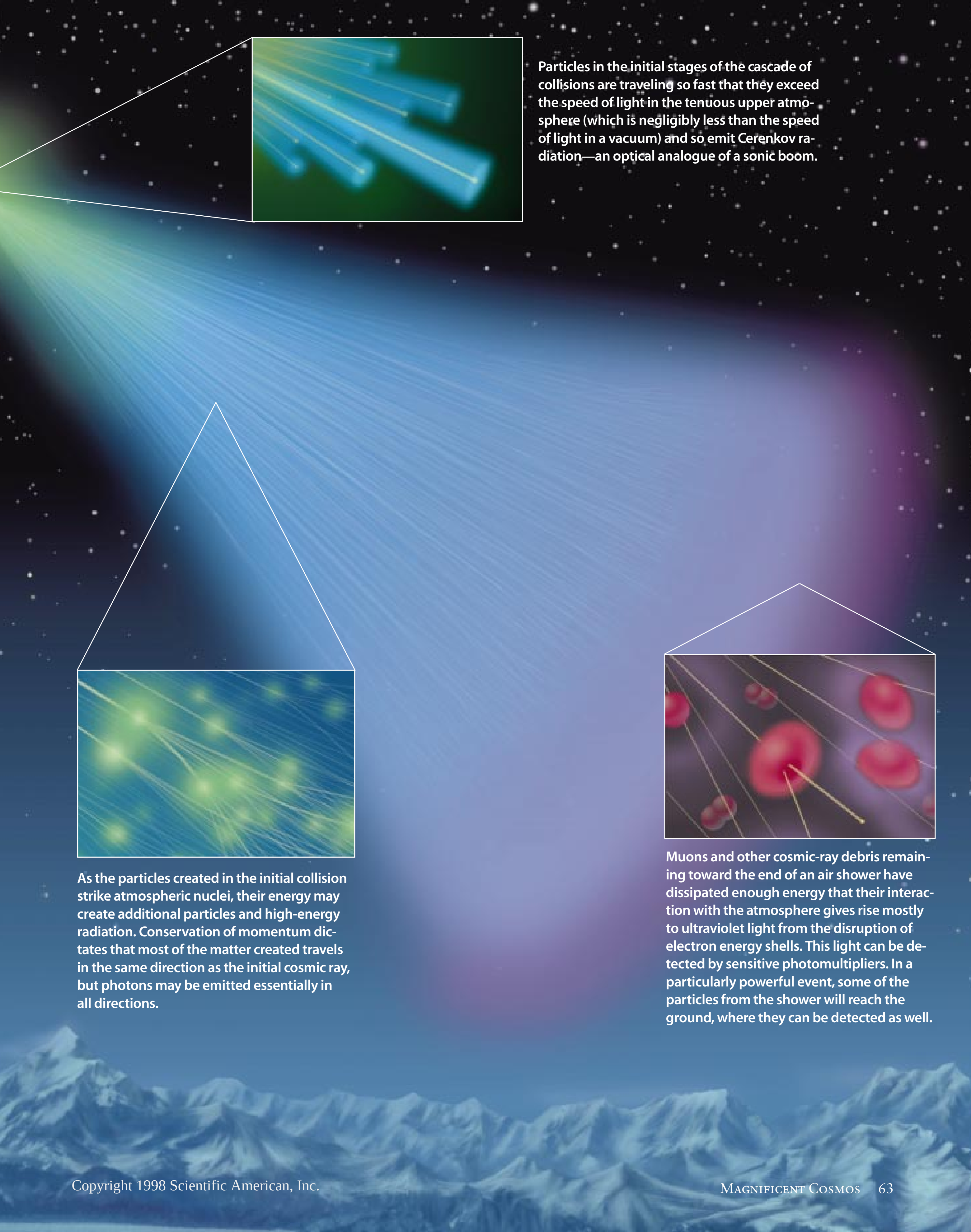
In 1960 Bernard Peters of the Tata Institute in Bombay suggested that lower-energy cosmic rays are produced predominantly inside our own galaxy, whereas those of higher energy come from more distant sources. One reason to think so is that a cosmic-ray proton carrying more than 10^{19} eV, for example, would not be deflected significantly by any of the magnetic fields typically generated by a galaxy, so it would travel more or less straight. If such particles came from inside our galaxy, we might expect to see different numbers coming from various directions because the galaxy is not arranged symmetrically around us. Instead the distribution is essentially isotropic, as is that of the lower-energy rays, whose directions are scattered.



Cosmic rays—atomic nuclei traveling at nearly the speed of light—inhabit a bizarre, relativistically foreshortened universe before smashing into nuclei of atoms of atmospheric gas high above Earth. A significant fraction of the incoming energy is converted to matter in the form of subatomic particles, including muons, which in turn collide violently with other atoms in the atmosphere to create an "air shower." Gamma rays are also emitted.

MICHAEL GOODMAN

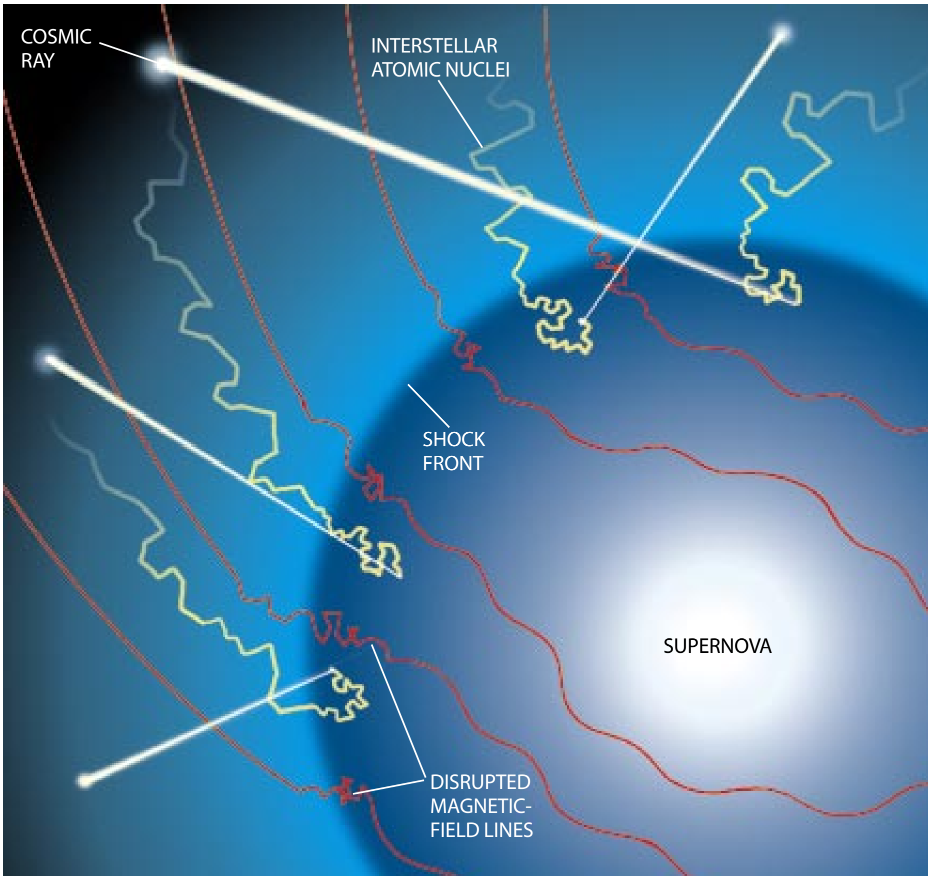
The Life of a Cosmic Ray



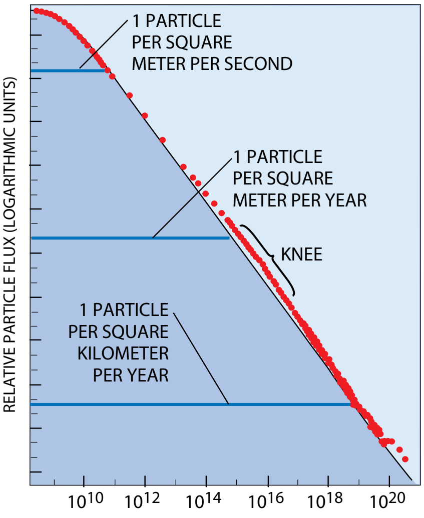
Particles in the initial stages of the cascade of collisions are traveling so fast that they exceed the speed of light in the tenuous upper atmosphere (which is negligibly less than the speed of light in a vacuum) and so emit Cerenkov radiation—an optical analogue of a sonic boom.

As the particles created in the initial collision strike atmospheric nuclei, their energy may create additional particles and high-energy radiation. Conservation of momentum dictates that most of the matter created travels in the same direction as the initial cosmic ray, but photons may be emitted essentially in all directions.

Muons and other cosmic-ray debris remaining toward the end of an air shower have dissipated enough energy that their interaction with the atmosphere gives rise mostly to ultraviolet light from the disruption of electron energy shells. This light can be detected by sensitive photomultipliers. In a particularly powerful event, some of the particles from the shower will reach the ground, where they can be detected as well.



GEORGE KELVIN



JENNIFER C. CHRISTIANSEN

Such tenuous inferences reveal how little is known for certain about the origin of cosmic rays. Astrophysicists have plausible models for how they might be produced but have no definitive answers. This state of affairs may be the result of the almost unimaginable difference between conditions on Earth and in the regions where cosmic rays are born. The space between the stars contains only about one atom per cubic centimeter, a far lower density than the best artificial vacuums we can create. Furthermore, these volumes are filled with vast electrical and magnetic fields, intimately connected to a diffuse population of charged particles even less numerous than the neutral atoms.

Supernova Pumps

This environment is far from the peaceful place one might expect: the low densities allow electrical and magnetic forces to operate over large distances and timescales in a manner that would be quickly damped out in material of terrestrial densities. Galactic space is therefore filled with an energetic and turbulent plasma of partially ionized gas in a state of violent activity. The motion is often hard to observe on human timescales because astronomical distances are so large; nevertheless, those same distances allow even moderate forces to achieve impressive results. A particle might zip through a terrestrial accelerator in a few microseconds, but it could spend years or even millennia in the accelerator's cosmic counterpart. (The timescales are further complicated by the strange, relativity-distorted framework that ultrahigh-energy cosmic rays inhabit. If we could observe such a particle for 10,000 years, that period would correspond to only a single second as far as the particle is concerned.)

Astronomers have long speculated that the bulk of galactic cosmic rays—those with energies below about 10^{16} eV—originate with supernovae. A compelling reason for this theory is that the power required to maintain the observed supply of cosmic-ray nuclei in our Milky Way galaxy is only slightly less than the average kinetic energy delivered to the galactic medium by the three supernova explosions that occur every century. There are few, if any, other sources of this amount of power in our galaxy.

When a massive star collapses, the outer parts of the star explode at speeds of up to 10,000 kilometers (6,000 miles) per second and more. A similar amount of energy is released when a white dwarf star undergoes complete disintegration in a thermonuclear detonation. In both types of supernovae the ejected matter expands at supersonic velocities, driving a strong shock into the surrounding medium. Such shocks are expected to accelerate nuclei from the material they pass through, turning them into cosmic rays. Because cosmic rays are charged, they follow complicated paths through interstellar magnetic fields. As a result, their directions as observed from Earth yield no information about the location of their original source.

By looking at the synchrotron radia-

AIR-SHOWER DETECTOR watches for traces of cosmic rays entering the upper atmosphere. Photodetectors can track flashes of light caused by particles interacting with air molecules and determine the energy and probable identity of the incoming rays. The Fly's Eye detector (*close-up at far right*) is located in Utah.

B. BOLA *U.S. Army*

COSMIC-RAY ACCELERATOR
is believed to arise from a supernova explosion. Astrophysicists hypothesize that atomic nuclei crossing the supernova shock front will pick up energy from the turbulent magnetic fields embedded in the shock. A particle may be deflected in such a way that it crosses the boundary of the shock hundreds or even thousands of times, picking up more energy on each passage, until it escapes as a cosmic ray. Most of the particles travel on paths that result in relatively small accelerations, accounting for the general shape of the cosmic-ray energy spectrum (*far right*), which falls off at higher energies. The "knee," or bend, in the curve suggests that most of the particles are accelerated by a mechanism incapable of imparting more than about 10^{15} electron volts. The relative excess of ultrahigh-energy particles indicates an additional source of acceleration whose nature is as yet unknown.

tion sometimes associated with supernova remnants, researchers have found more direct evidence that supernovae can act as accelerators. Synchrotron radiation is characteristic of high-energy electrons moving in an intense magnetic field of the kind that might act as a cosmic-ray accelerator, and the presence of synchrotron x-rays in some supernova remnants suggests particularly high energies. (In Earthbound devices, synchrotron emission limits a particle's energy because the emission rate increases as a particle goes faster; at some point, the radiation bleeds energy out of an accelerating particle as fast as it can be pumped in.) Recently the Japanese x-ray satellite Asca made images of the shell of Supernova 1006, which exploded 990 years ago. Unlike the radiation from the interior of the remnant, the x-radiation from the shell has the features characteristic of synchrotron radiation. Astrophysicists have deduced that electrons are being accelerated there at up to 10^{14} eV.

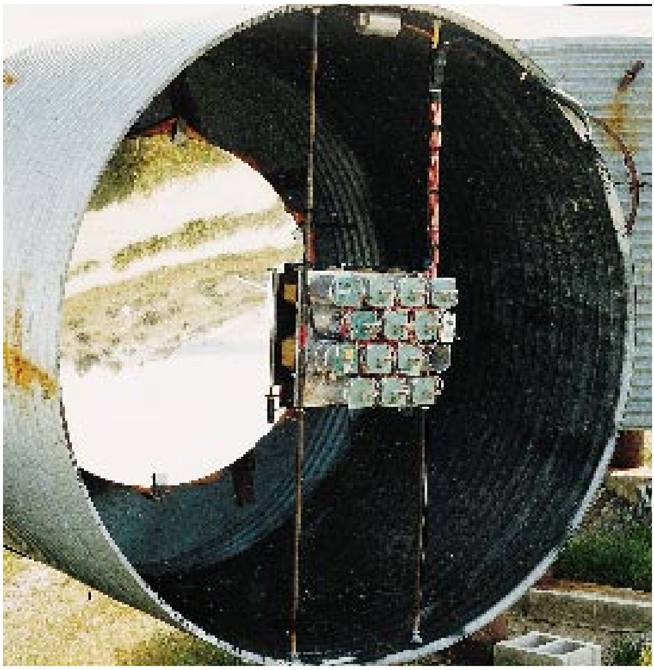
The EGRET detector on the Compton Gamma Ray Observatory has also been used to study point sources of gamma rays identified with supernova remnants. The observed intensities and spectra (up to a billion electron volts) are consistent with an origin from the decay of particles called neutral pions, which could be produced by cosmic rays from the exploding star's remnants colliding with nearby interstellar gas. Interestingly, however, searches made by the ground-based Whipple Observatory for gamma rays of much higher energies from some of the same remnants have not seen signals at the levels that would be expected if the supernovae were accelerating protons to 10^{14} eV or more.

A complementary method for testing the association of high-energy cosmic rays with supernovae involves the elemental composition of cosmic-ray nuclei. The size of the orbit of a charged particle in a magnetic field is proportional to its total momentum per unit charge, so heavier nuclei have greater total energy for a given orbit size. Any process that limits the particle acceleration on the basis of orbit size (such as an accelerating region of limited extent) will thus lead to an excess of heavier nuclei at high energies.

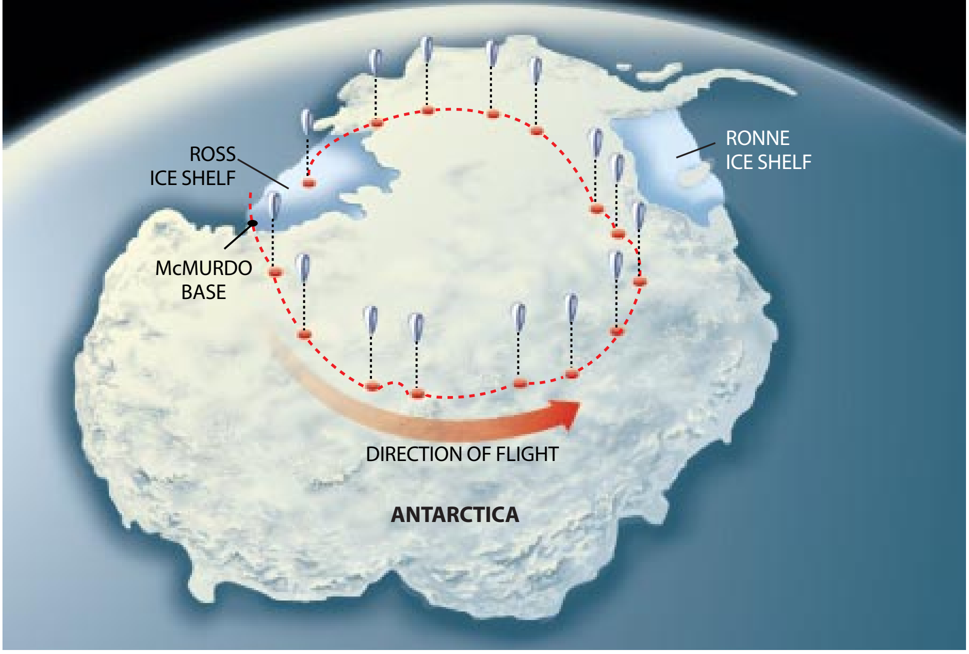
Eventually we would like to be able to go further and look for elemental signatures of acceleration in specific types of supernovae. For example, the supernova of a white dwarf detonation would accelerate whatever nuclei populate the local interstellar medium. A supernova that followed the collapse of a massive star, in contrast, would accelerate the surrounding stellar wind, which is characteristic of the outer layers of the progenitor star at earlier stages of its evolution. In some cases, the wind could include an increased fraction of helium, carbon or even heavier nuclei.

The identity of high-energy cosmic rays is all but lost when they interact with atoms in Earth's atmosphere and form a shower of secondary particles. Hence, to be absolutely sure of the nuclear composition, measurements must be made before the cosmic rays reach dense atmosphere. Unfortunately, to collect 100 cosmic rays of energies near 10^{15} eV, a one-square-meter detector would have to be in orbit for three years. Typical exposures at present are more like the equivalent of one square meter for three days.

Researchers are attacking this problem with some ingenious experiments. For example, the National Aeronautics and Space Administration has developed techniques to loft large payloads (about three metric tons) with high-altitude bal-



UNIVERSITY OF UTAH



HIGH-ALTITUDE BALLOON launched near McMurdo Base in Antarctica carries cosmic-ray detectors above most of the atmosphere. Winds 40 kilometers above the ice cap blow in a circle around the Pole, returning the balloon to the vicinity of its starting point after about 10 days. Balloon detectors are not as sensitive as those placed on board satellites, but they can be made much larger and lofted much more cheaply.

GEORGE KELVIN

loons for many days. These experiments cost a tiny fraction of what an equivalent satellite detector would. The most successful flights of this type have taken place in Antarctica, where the upper atmosphere winds blow in an almost constant circle around the South Pole.

A payload launched at McMurdo Sound on the coast of Antarctica will travel at a nearly constant radius from the Pole and return eventually to near the launch site. Some balloons have circled the continent for 10 days. One of us (Swordy) is collaborating with Dietrich Müller and Peter Meyer of the University of Chicago on a 10-square-meter detector that could measure heavy cosmic rays of up to 10^{15} eV on such a flight. There are efforts to extend the exposure times to roughly 100 days with similar flights nearer the equator.

Across Intergalactic Space

Studying even higher-energy cosmic rays—those produced by sources as yet unknown—requires large ground-based detectors, which overcome the problem of low flux by watching enormous areas for months or years. The information, however, must be extracted from cascades of secondary particles—electrons, muons and gamma rays—initiated high in the atmosphere by an incoming cosmic-ray nucleus. Such indirect methods can only suggest general features of the composition of a cosmic ray on a statistical basis, rather than identifying the atomic number of each incoming nucleus.

At ground level, the millions of secondary particles unleashed by one cosmic ray are spread over a radius of hundreds of meters. Because it is impractical to blanket such a large area with detectors, the detectors typically sample these air showers at a few hundred or so discrete locations.

Technical improvements have enabled such devices to collect increasingly sophisticated data sets, thus refining the conclusions we can draw from each shower. For example, the CASA-MIA-DICE experiment in Utah, in which two of us (Cronin and Swordy) are involved, measures the distributions of electrons and muons at ground level. It also detects Cerenkov light (a type of optical shock wave produced by particles moving faster than the speed of light in their surrounding medium) generated by the shower particles at various levels in the atmosphere. These data enable us to reconstruct the shape of the shower more reliably and thus take a better guess at the energy and identity of the cosmic ray that initiated it.

The third one of us (Gaisser) is working with an array that measures showers reaching the surface at the South Pole. This experiment works in conjunction with AMANDA, which detects energetic muons produced in the same showers by observing Cerenkov radiation produced deep in the ice cap. The primary goal of AMANDA is to catch traces of neutrinos produced in cosmic accelerators, which may generate upward-streaming showers after passing through Earth.

Cosmic rays with energies above 10^{20} eV strike Earth's atmosphere at a rate of only about one per square kilometer a century. As a result, studying them requires an air-shower detector of truly gigantic proportions. In addition to the 1991 event in Utah, particles with energies above 10^{20} eV have been seen by groups elsewhere in the U.S., in Akeno, Japan, in Haverah Park, U.K., and in Yakutsk, Siberia.

Particles of such high energy pose a conundrum. On the one hand, they are likely to come from outside our galaxy because no known acceleration mechanism could produce them and because they approach from all directions even though a galactic magnetic field is insufficient to bend their path. On the other hand, their source cannot be more than about 30 million light-years away, because the particles would otherwise lose energy by interaction with the universal microwave background—radiation left over from the birth of the cosmos in the big bang. In the relativistic universe that the highest-energy cosmic rays inhabit, even a single radio-frequency photon packs enough punch to rob a particle of much of its energy.

If the sources of such high-energy particles were distributed uniformly throughout the cosmos, interaction with the microwave background would cause a sharp cutoff in the number of particles with energy above 5×10^{19} eV, but that is not the case. There are as yet too few events above this nominal threshold for us to know for certain what is going on, but even the few we have seen provide us with a unique opportunity for theorizing. Because these rays are essentially undeflected by the weak intergalactic magnetic fields, measuring the direction of travel of a large enough sample should yield unambiguous clues to the locations of their sources.

It is interesting to speculate what the sources might be. Three recent hypotheses suggest the range of possibilities: galactic black-hole accretion disks, gamma-ray bursts and topological defects in the fabric of the universe.

Astrophysicists have predicted that black holes of a billion solar masses or more, accreting matter in the nuclei of active galaxies, are needed to drive relativistic jets of matter far into intergalactic space at speeds approaching that of light; such

STEVEN PETERZEN *National Scientific Balloon Facility*

jets have been mapped with radio telescopes. Peter L. Biermann of the Max Planck Institute for Radioastronomy in Bonn and his collaborators suggest that the hot spots seen in these radio lobes are shock fronts that accelerate cosmic rays to ultrahigh energy. There are some indications that the directions of the highest-energy cosmic rays to some extent follow the distribution of radio galaxies in the sky.

The speculation about gamma-ray bursts takes off from the theory that the bursts are created by relativistic explosions, perhaps resulting from the coalescence of neutron stars. Mario Vietri of the Astronomical Observatory of Rome and Eli Waxman of Princeton University independently noted a rough match between the energy available in such cataclysms and that needed to supply the observed flux of the highest-energy cosmic rays. They argue that the ultrahigh-speed shocks driven by these explosions act as cosmic accelerators.

Rare Giants

Perhaps most intriguing is the notion that ultrahigh-energy particles owe their existence to the decay of monopoles, strings, domain walls and other topological defects that might have formed in the early universe. These hypothetical objects are believed to harbor remnants of an earlier, more symmetrical phase of the fundamental fields in nature, when gravity, electromagnetism and the weak and strong nuclear forces were merged. They can be thought of, in a sense, as infinitesimal pockets preserving bits of the universe as it existed in the fractional instants after the big bang.

As these pockets collapse, and the symmetry of the forces within them breaks, the energy stored in them is released in the form of supermassive particles that immediately decay into jets of particles with energies up to 100,000 times greater than those of the known ultrahigh-energy cosmic rays. In this scenario the ultrahigh-energy cosmic rays we observe are the comparatively sluggish products of cosmological particle cascades.

Whatever the source of these cosmic rays, the challenge is to collect enough of them to search for detailed correlations with extragalactic objects. The AGASA array in Japan currently has an effective area of 100 square kilometers and can capture only a few ultrahigh-energy events a year. The new Fly's Eye High Resolution experiment in Utah can see out over a much larger area, but only on clear, moonless nights.

For the past few years, Cronin and Alan A. Watson of the University of Leeds have spearheaded an initiative to gather an even larger sample of ultrahigh-energy cosmic rays. This development is named the Auger Project, after Pierre Auger, the French scientist who first investigated the phenomenon of correlated showers of particles from cosmic rays.

The plan is to provide a detection area of 6,000 square kilometers with a 100 percent duty cycle that is capable of measuring hundreds of high-energy events a year. A detector field would consist of many stations on a 1.5-kilometer grid; a single event might trigger dozens of stations. To cover the entire sky, two such detectors are planned, one each for the Northern and Southern hemispheres.

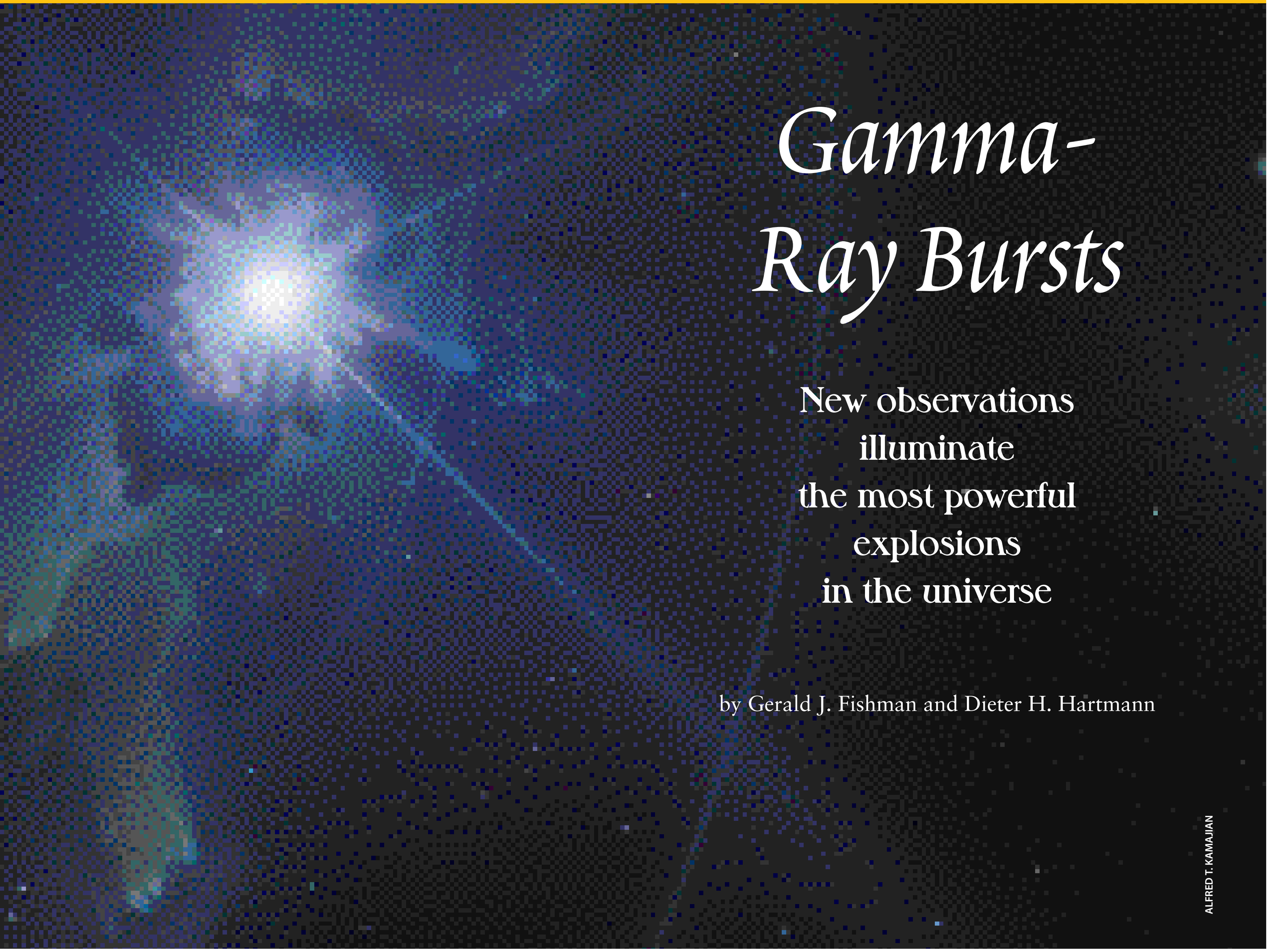
An Auger Project design workshop held at the Fermi National Accelerator Laboratory in 1995 has shown how modern off-the-shelf technology such as solar cells, cellular telephones and Global Positioning System receivers can make such a system far easier to construct. A detector the size of Rhode Island could be built for about $50 million.

Plans exist to cover even larger areas. Detectors in space could view millions of square kilometers of the atmosphere from above, looking for flashes of light signaling the passage of ultrahigh-energy particles. This idea, which goes by the name of OWL (Orbiting Wide-angle Light collectors) in the U.S. and by Airwatch in Europe, was first suggested by John Linsley of the University of New Mexico. To succeed, the project requires developing new technology for large, sensitive, finely segmented optics in space to provide the resolution needed. This development is under way by the U.S. National Aeronautics and Space Administration and in Italy.

As researchers confront the problem of building and operating such gigantic detector networks, the fundamental question remains: Can nature produce even more energetic particles than those we have seen? Could there be still higher-energy cosmic rays, or are we already beginning to detect the highest-energy particles our universe can create?

The Authors

JAMES W. CRONIN, THOMAS K. GAISSER and SIMON P. SWORDY work on both the theoretical questions of how cosmic rays are created and the practical problems inherent in detecting and analyzing them. Cronin, a professor of physics at the University of Chicago since 1971, earned his master's degree from the university in 1953 and his doctorate in 1955. In 1980 he shared the Nobel Prize with Val L. Fitch for work on symmetry violations in the decay of mesons. Gaisser, a professor of physics at the University of Delaware, has concentrated on the interpretation of atmospheric cosmic-ray cascades; he earned his doctorate from Brown University in 1967. In 1995 Gaisser spent two months in Antarctica setting up cosmic-ray detectors. Swordy, an associate professor at Chicago, has been active in cosmic-ray measurement since 1976. He earned his Ph.D. from the University of Bristol in 1979. This article updates a version that appeared in *Scientific American* in January 1997.



Gamma-Ray Bursts

New observations illuminate the most powerful explosions in the universe

by Gerald J. Fishman and Dieter H. Hartmann

ALFRED T. KAMAJIAN

About three times a day our sky flashes with a powerful pulse of gamma rays, invisible to human eyes but not to astronomers' instruments. The sources of this intense radiation are likely to be emitting, within the span of seconds or minutes, more energy than the sun will in its entire 10 billion years of life. Where these bursts originate, and how they come to have such incredible energies, is a mystery that scientists have been attacking for three decades. The phenomenon has resisted study—the flashes come from random directions in space and vanish without trace—until very recently.

On February 28, 1997, we were lucky. One such burst hit the Italian-Dutch Beppo-SAX satellite for about 80 seconds. Its gamma-ray monitor established the position of the burst—prosaically labeled GRB 970228—to within a few arc minutes in the Orion constellation, about halfway between the stars Alpha Tauri and Gamma Orionis. Within eight hours, operators in Rome had turned the spacecraft around to look in the same region with an x-ray telescope. They found a source of x-rays (radiation of somewhat lower frequency than gamma rays) that was fading fast, and they fixed its location to within an arc minute.

Never before has a burst been pinpointed so accurately and so quickly, allowing powerful optical telescopes, which have narrow fields of view of a few arc minutes, to look for it. Astronomers on the Canary Islands, part of an international team led by Jan van Paradijs of the University of Amsterdam and the University of Alabama in Huntsville, learned of the finding by electronic mail. They had some time available on the 4.2-meter William Herschel Telescope, which they had been using to study the locations of other bursts. They took a picture of the area 21 hours after GRB 970228. Eight days later they looked again and found that a spot of light seen in the earlier photograph had disappeared.

On March 13 the New Technology Telescope in La Silla, Chile, took a long, close look at those coordinates and discerned a diffuse, uneven glow. The Hubble Space Telescope later resolved it to be a bright point surrounded by a somewhat elongated background object. In a few days the Hubble reexamined the position and still found the point—now very faint—as well as the fuzzy glow, unaltered. Many of us believe the latter to be a galaxy, but its true identity remains unknown.

Even better, on the night of May 8, Beppo-SAX operators located a 15-second burst, designated GRB 970508. Soon after, Howard E. Bond of the Space Telescope Science Institute in Baltimore photographed the region with the 0.9-meter optical telescope on Kitt Peak in Arizona; the next night a point of light in the field had actually brightened. Other telescopes confirm that after becoming most brilliant on May 10, the source began to fade. This is the first time that a burst has been observed reaching its optical peak—which, astonishingly, lagged its gamma-ray peak by a few days.

Also for the first time, on May 13 Dale Frail, using the Very Large Array of radio telescopes in New Mexico, detected radio emissions from the burst remnant. Even more exciting, the primarily blue spectrum of this burst, taken on May 11 with the Keck II telescope on Hawaii, showed a few dark lines, apparently caused by iron and magnesium in an intervening cloud. Astronomers at the California Institute of Technology find that the displacement of these absorption lines indicates a distance of more than seven billion light-years. If this interpretation holds up, it will establish that bursts occur at cosmological distances. In that case, gamma-ray bursts must represent the most powerful explosions in the universe.

Confounding Expectations

For those of us studying gamma-ray bursts, this discovery salves two recent wounds. In November 1996 the Pegasus XL launch vehicle failed to release the High Energy Transient Explorer (HETE) spacecraft equipped with very accurate instruments for locating gamma-ray bursts. And in December the Russian Mars ’96 spacecraft, with several gamma-ray detectors, fell into the Pacific Ocean after a rocket malfunction. These payloads were part of a set designed to launch an attack on the origins of gamma-ray bursts. Of the newer satellites equipped with gamma-ray instruments, only BeppoSAX—whose principal scientists include Luigi Piro, Enrico Costa and John Heise—made it into space, on April 20, 1996.

Gamma-ray bursts were first discovered by accident, in the late 1960s, by the Vela series of spacecraft of the U.S. Department of Defense. These satellites were designed to ferret out the U.S.S.R.’s clandestine nuclear detonations in outer space—perhaps hidden behind the moon. Instead they came across spasms of radiation that did not originate from near Earth. In 1973 scientists concluded that a new astronomical phenomenon had been discovered.

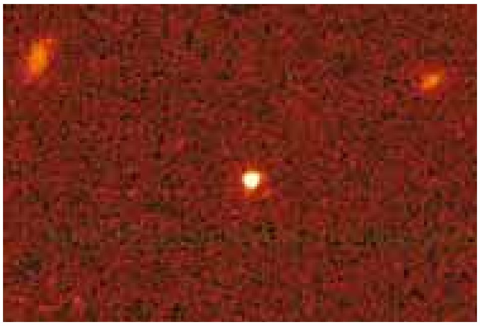
These initial observations resulted in a flurry of speculation about the origins of gamma-ray bursts—involving black holes, supernovae or the dense, dark star remnants called neutron stars. There were, and still are, some critical unknowns. No one knew whether the bursts were coming from a mere 100 light-years away or a few billion. As a result, the energy of the original events could only be guessed at.

By the mid-1980s the consensus was that the bursts originated on nearby neutron stars in our galaxy. In particular, theorists were intrigued by dark lines in the spectra (component wavelengths spread out, as light is by a prism) of some bursts, which suggested the presence of intense magnetic fields. The gamma rays, they postulated, are emitted by electrons accelerated to relativistic speeds when magnetic-field lines from a neutron star reconnect. A similar phenomenon on the sun—but at far lower energies—leads to flares.

In April 1991 the space shuttle *Atlantis* launched the Compton Gamma Ray Observatory, a satellite that carried the Burst And Transient Source Experiment (BATSE). Within a year BATSE had confounded all expectations. The distribution of gamma-ray bursts did not trace out the Milky Way, nor were the bursts associated with nearby galaxies or clusters of galaxies. Instead they were distributed isotropically, with any direction in the sky having roughly the same number. Theorists soon refined the galactic model: the bursts were now said to come from neutron stars in an extended spherical halo surrounding the galaxy.

One problem with this scenario is that Earth lies in the suburbs of the Milky Way, about 25,000 light-years from the core. For us to find ourselves near the center of a galactic halo, the latter must be truly enormous, almost 800,000 light-years in outer radius. If so, the halo of the neighboring Andromeda galaxy should be as extended and should start to appear in the distribution of gamma-ray bursts. But it does not.

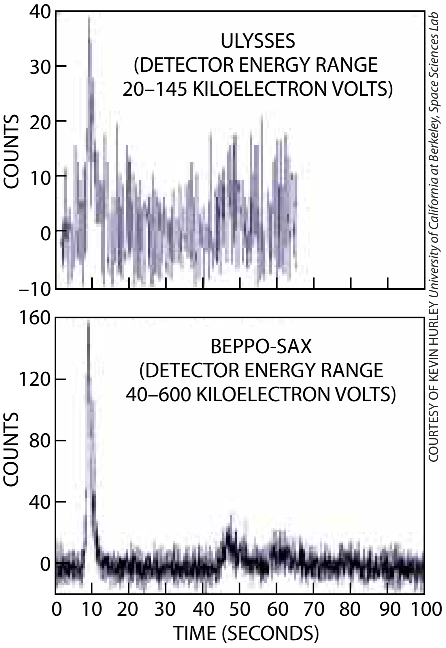
This uniformity, combined with the data from GRB 970508, has convinced most astrophysicists that the bursts come from cosmological distances, on the order of three billion to 10 billion light-years away. At such a distance, though, the bursts



HUBBLE SPACE TELESCOPE

ROGER RESSMEYER *Corbis*

VERY LARGE ARRAY of radio telescopes (*right*) discovered radio waves from a burst (GRB 970508) for the first time in May 1997. The burst (*above, at center*) had a cosmological origin but showed no underlying galaxy, confounding theorists.



COURTESY OF KEVIN HURLEY *University of California at Berkeley, Space Sciences Lab*

TIME PROFILE of GRB 970228 taken by the Ulysses spacecraft (*top*) and by Beppo-SAX (*bottom*) shows a brief, brilliant flash of gamma rays.

should show the effects of the expansion of the universe. Galaxies that are very distant are moving away from Earth at great speeds; we know this because the light they emit shifts to lower, or redder, frequencies. Likewise, gamma-ray bursts should also show a "redshift," as well as an increase in duration.

Unfortunately, BATSE does not see, in the spectrum of gamma rays, bright or dark lines characterizing specific elements whose displacements would betray a shift to the red. (Nor does it detect the dark lines found by earlier satellites.) In April astronomers using the Keck II telescope in Hawaii obtained an optical spectrum of the afterglow of GRB 970228—smooth and red, with no telltale lines. Still, Jay Norris of the National Aeronautics and Space Administration Goddard Space Flight Center and Robert Mallozzi of the University of Alabama in Huntsville have statistically analyzed the observed bursts and report that the weakest, and therefore the most distant, show both a time dilation and a redshift. And the dark lines in the spectrum of GRB 970508 are substantially shifted to the red.

A Cosmic Catastrophe

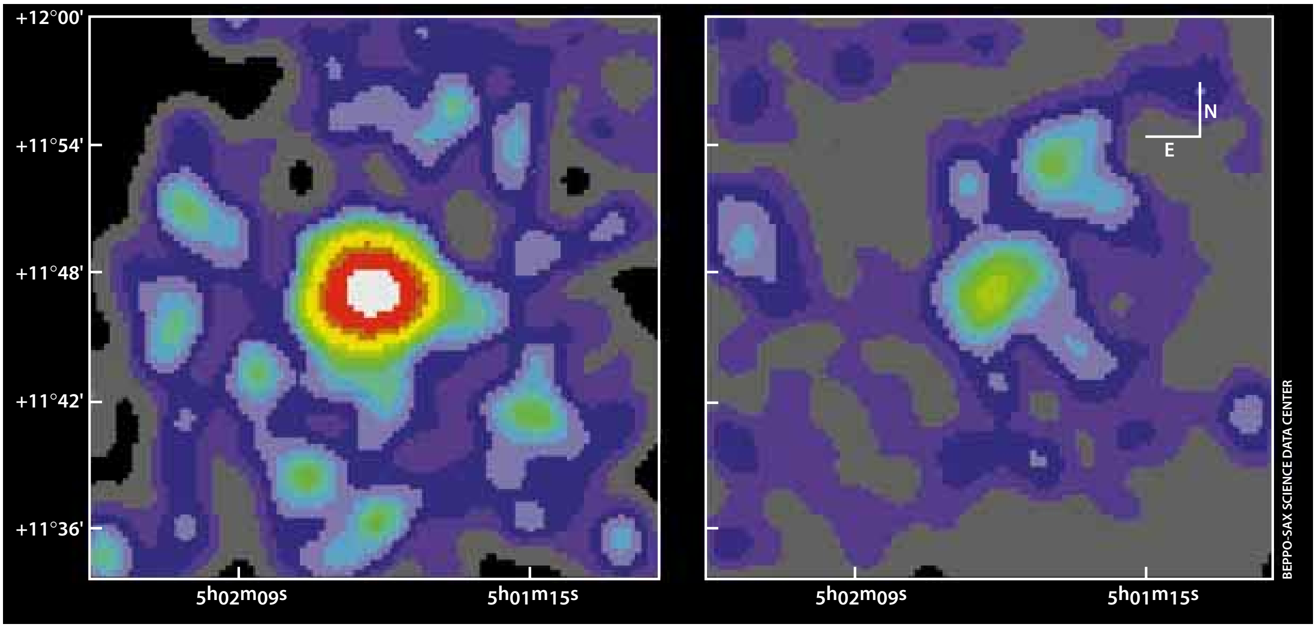
One feature that makes it difficult to explain the bursts is their great variety. A burst may last from about 30 milliseconds to almost 1,000 seconds—and in one case, 1.6 hours. Some bursts show spasms of intense radiation, with no detectable emission in between, whereas others are smooth. Also complicated are the spectra—essentially, the colors of the radiation, invisible though they are. The bulk of a burst's energy is in radiation of between 100,000 and one million electron volts, implying an exceedingly hot source. (The photons of optical light, the primary radiation from the sun, have energies of a few electron volts.) Some bursts evolve smoothly to lower frequencies such as x-rays as time passes. Although this x-ray tail has less energy, it contains many photons.

If originating at cosmological distances, the bursts must have energies of perhaps 10^{52} ergs. (About 1,000 ergs can lift a gram by one centimeter.) This energy must be emitted within seconds or less from a tiny region of space, a few tens of kilometers across. It would seem we are dealing with a fireball.

The first challenge is to conceive of circumstances that would create a sufficiently energetic fireball. Most theorists favor a scenario in which a binary neutron-star system collapses [see "Binary Neutron Stars," by Tsvi Piran; Scientific American, May 1995]. Such a pair gives off gravitational energy in the form of radiation. Consequently, the stars spiral in toward each other and may ultimately merge to form a black hole. Theoretical models estimate that one such event occurs every 10,000 to one million years in a galaxy. There are about 10 billion galaxies in the volume of space that BATSE observes; that yields up to 1,000 bursts a year in the sky, a number that fits the observations.

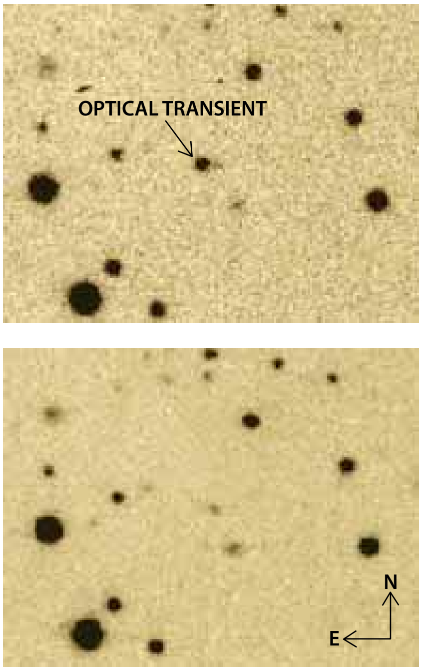
Variations on this scenario involve a neutron star, an ordinary star or a white dwarf colliding with a black hole. The details of such mergers are a focus of intense study. Nevertheless, theorists agree that before two neutron stars,

X-RAY IMAGE taken by Beppo-SAX on February 28, 1997 (*left image*), localized GRB 970228 to within a few arc minutes, allowing ground-based telescopes to search for it. On March 3 the source was much fainter (*right image*).



BEPPO-SAX SCIENCE DATA CENTER

OPTICAL IMAGES of the region of GRB 970228 were taken by the William Herschel Telescope on the Canary Islands, on February 28 (*top*) and March 8 (*bottom*). A point of light in the first image has faded away in the second one, indicating a transient afterglow.



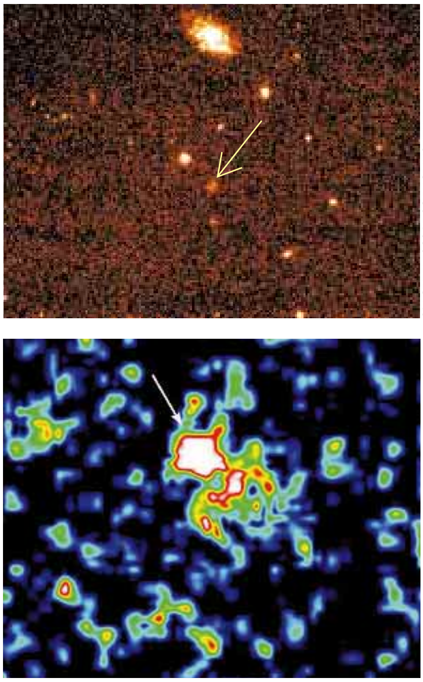
PAUL GROOT, TITUS GALAMA AND JAN VAN PARADIS *University of Amsterdam*

say, collapse into a black hole, their death throes release as much as 10^{53} ergs. This energy emerges in the form of neutrinos and antineutrinos, which must somehow be converted into gamma rays. That requires a chain of events: neutrinos collide with antineutrinos to yield electrons and positrons, which then annihilate one another to yield photons. Unfortunately, this process is very inefficient, and recent simulations by Max Ruffert and Hans-Thomas Janka of the Max Planck Institute in Munich, as well as by other groups, suggest it may not yield enough photons.

Worse, if too many heavy particles such as protons are in the fireball, they reduce the energy of the gamma rays. Such proton pollution is to be expected, because the collision of two neutron stars must yield a potpourri of particles. But then all the energy ends up in the kinetic energy of the protons, leaving none for radiation. As a way out of this dilemma, Peter Mészáros of Pennsylvania State University and Martin J. Rees of the University of Cambridge have suggested that when the expanding fireball—essentially hot protons—hits surrounding gases, it produces a shock wave. Electrons accelerated by the intense electromagnetic fields in this wave then emit gamma rays.

A variation of this scenario involves internal shocks, which occur when different parts of the fireball hit one another at relativistic speeds, also generating gamma rays. Both the shock models imply that gamma-ray bursts should be followed by long afterglows of x-rays and visible light. In particular, Mario Vietri of the Astronomical Observatory of Rome has predicted detectable x-ray afterglows lasting for a month—and also noted that such afterglows do not occur in halo models. GRB 970228 provides the strongest evidence yet for such a tail.

There are other ways of generating the required gamma rays. Nir Shaviv and Arnon Dar of the Israel Institute of Technology in Haifa start with a fireball of unknown origin that is rich in heavy metals. Hot ions of iron or nickel could then interact with radiation from nearby stars to give off gamma rays. Simulations show that the time profiles of the resulting bursts are quite close to observations, but a fireball consisting entirely of heavy metals seems unrealistic.



SPACE TELESCOPE SCIENCE INSTITUTE

SPACE TELESCOPE SCIENCE INSTITUTE

OPTICAL REMNANT of GRB 970228 was pictured by the Hubble Space Telescope. The afterglow (*near center of top image*), when seen in close-up (*bottom*), has a faint, elongated background glow that may correspond to a galaxy in which the burst occurred.

Another popular mechanism invokes immensely powerful magnetic engines, similar to the dynamos that churn in the cores of galaxies. Theorists envision that instead of a fireball, a merger of two stars—of whatever kind—could yield a black hole surrounded by a thick, rotating disk of debris. Such a disk would be very short-lived, but the magnetic fields inside it would be astounding, some 10^{15} times those on Earth. Much as an ordinary dynamo does, the fields would extract rotational energy from the system, channeling it into two jets bursting out along the rotation axis.

The cores of these jets—the regions closest to the axis—would be free of proton pollution. Relativistic electrons inside them can then generate an intense, focused pulse of gamma rays. Although quite a few of the details remain to be worked out, many such scenarios ensure that mergers are the leading contenders for explaining bursts.

Still, gamma-ray bursts have been the subject of more than 3,000 papers—about one publication per recorded burst. Their transience has made them difficult to observe with a variety of instruments, and the resulting paucity of data has allowed for a proliferation of theories.

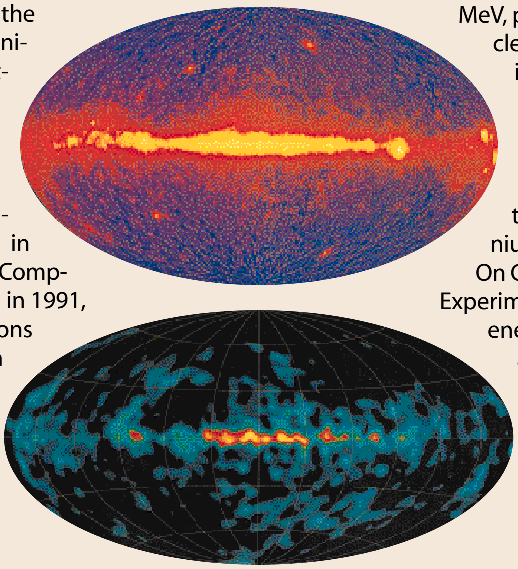
If one of the satellites detects a lensed burst, astronomers would have further confirmation that bursts occur at cosmological distances. Such an event might occur if an intervening galaxy or other massive object serves as a gravitational lens to bend the rays from a burst toward Earth. When optical light from a distant star is focused in this manner, it appears as multiple images of the original star, arranged in arcs around the lens. Gamma rays cannot be pinpointed with such accuracy; instead they are currently detected by instruments that have poor directional resolution.

Moreover, bursts are not steady sources like stars. A lensed gamma-ray burst would therefore show up as two bursts coming from roughly the same direction, having identical spectra and time profiles but different intensities and arrival times. The time difference would come from the rays' traversing curved paths of different lengths through the lens.

To further nail down the origins of the underlying explosion, we need data on other kinds of radiation that might

The Gamma-Ray Sky

Gamma-ray astronomy elucidates the structure and evolution of the universe by means of the photons of greatest energy. Because gamma rays are absorbed by the atmosphere of Earth and, moreover, are hard to detect, their study poses a challenge to technology.



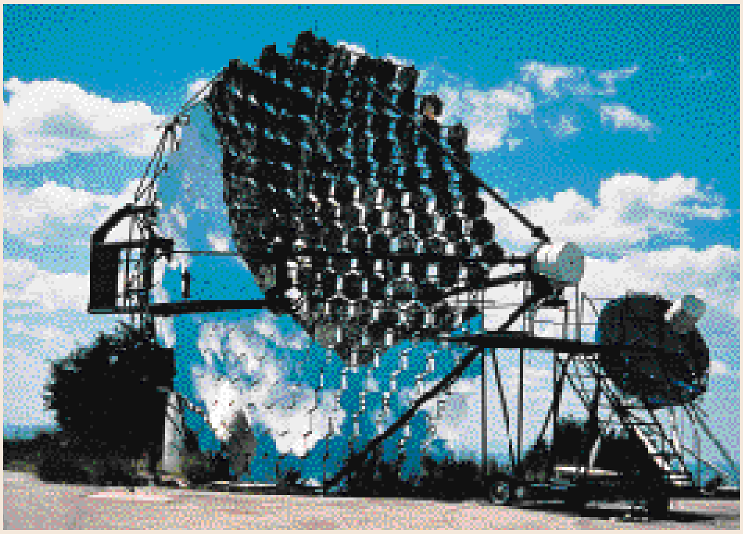
SKY MAP (*top*) in photons of more than 100 million electron volts (MeV) traces the fastest particles in the universe, glowing when they interact with interstellar matter and light.The isolated spots far from the Milky Way (*lateral line*) are blazars. The map in photons of a precise energy, 1.8 MeV (*bottom*), reveals the presence of aluminum 26 and therefore the distribution of past supernovae. It demonstrates that the synthesis of elements and their dispersion from supernovae is an ongoing process throughout the galaxy.

EGRET, COMPTON OBSERVATORY (*top*); R. DIEHL AND U. OBERLACK *Max Planck Institute for Extraterrestrial Physics* AND COMPTEL, COMPTON OBSERVATORY (*bottom*)

Early detectors were flown on balloons. Nowadays instruments based in space survey the sky for these rays. The Compton Gamma Ray Observatory, launched in 1991, uses complex detectors to catch photons in the energy range of 10 kilo electron volts to 30 giga electron volts (GeV). Future instruments, such as the Gamma-ray Large Area Space Telescope (GLAST) planned for 2004, will survey the sky even more sensitively at higher energies of up to 300 GeV.

When a photon's energy becomes large enough, it creates an avalanche of particles on penetrating the atmosphere. These particles then emit optical light that can be detected on the ground by large mirrored collectors such as Whipple in Arizona. Whipple currently detects particles of energy 300 GeV or higher. If it is upgraded, as planned, to VERITAS (Very Energetic Radiation Imaging Telescope Array System), the array will detect particles of energy as low as 100 GeV, closing the gap with the satellite data.

Gamma rays are emitted by the most violent explosions in the universe. As a result, they allow astronomers to study essential processes such as the production of elements in the universe. Heavy elements created within stars are dispersed by supernovae explosions; new stars and planets are then born from the chemically enriched gas, ultimately incorporating the new substances into emerging life.



WHIPPLE OBSERVATORY

WHIPPLE OBSERVATORY on Mount Hopkins in Arizona surveys the sky in gamma rays. An energetic gamma-ray photon penetrating the atmosphere releases a shower of optical photons that the 10-meter reflector detects.

One of the nuclei thus produced and ejected is aluminum 26, which decays in about a million years by emitting a photon of 1.8 million eV (MeV). Two instruments on the Compton observatory map the sky in this line, thereby providing an image of the past supernovae activity in the Milky Way. Tens of thousands of supernovae (occurring at a rate of one per century) contribute to a diffuse glow at 1.8 MeV, providing testimony to the amazing cycle of stellar birth and death as an ongoing process. The INTEGRAL (International Gamma Ray Astrophysics Laboratory) satellite, to be launched early in the next century, will continue the quest for gamma-ray maps of special spectral lines, such as aluminum 26 and titanium 44.

On Compton, EGRET (Energetic Gamma Ray Experiment Telescope) has mapped the sky at energies above 100 MeV, finding a bright and diffuse glow concentrated along the galactic midplane. The radiation is emitted by fast particles—from supernovae explosions—as they slam into the molecular gas between the stars. Apart from tracing such remnants of violent processes, the EGRET image also shows pointlike sources. Some of these, close to the plane of the galaxy, are pulsars, stable cores left in the wake of supernovae.

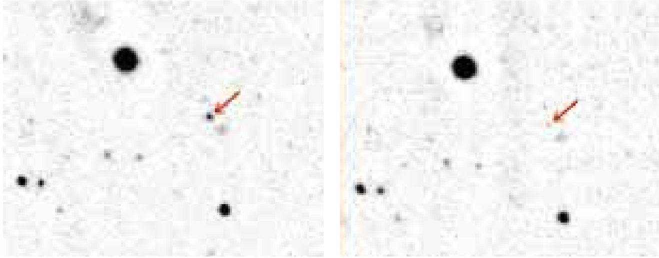
These dense, compact objects often have extremely strong magnetic fields—a trillion times that of Earth—and rotate very rapidly, with periods of milliseconds. The magnetized atmospheres of some such "dead stars" emit strongly beamed gamma rays. But if the rays miss the detectors, astronomers may never notice the star even if it is nearby. Although radio astronomers have found about 1,000 pulsars, gamma-ray astronomers have detected only half a dozen. Even so, these gamma-ray pulsars have taught us a great deal about the behavior of matter under extreme conditions. One example is the process by which electrons emit radiation when in magnetic fields too high to be created on Earth.

Yet another kind of point source of gamma rays is a blazar. Blazars are active galaxies in whose centers lie black holes as massive as a billion suns. The gas and stars the black hole draws in emit a beam of gamma rays. These rays allow us to probe the conditions of matter near a black hole and the ways in which it spirals inward.

And then, of course, there are the gamma-ray bursters, perhaps the most mysterious of all.

—G.J.F. and D.H.H.

accompany a burst. Even better would be to identify the source. Until the observation of GRB 970228 such "counterparts" had proved exceedingly elusive. To find others, we must locate the bursts very precisely.


ASTROPHYSICAL RESEARCH CONSORTIUM

OPTICAL TRANSIENT for a third burst, GRB 971214, provided a New Year's gift to astronomers. The images were taken at the Apache Point Observatory in Sunspot, N.M., on December 15 (*left*) and 16 (*right*).

Since the early 1970s, Kevin Hurley of the University of California at Berkeley and Thomas Cline of the NASA Goddard Space Flight Center have worked to establish "interplanetary networks" of burst instruments. They try to put a gamma-ray detector on any spacecraft available or to send aloft dedicated devices. The motive is to derive a location to within arc minutes, by comparing the times at which a burst arrives at spacecraft separated by large distances.

From year to year, the network varies greatly in efficacy, depending on the number of participating instruments and their separation. At present, there are five components: BATSE, Beppo-SAX and the military satellite DMSP, all near Earth; Ulysses, far above the plane of the solar system; and the spacecraft Wind, orbiting the sun. The data from Beppo-SAX, Ulysses and Wind were used to triangulate GRB 970228. (BATSE was in Earth's shadow at the time.) The process, unfortunately, is slow—eight hours at best.

Watching and Waiting

Time is of the essence if we are to direct diverse detectors at a burst while it is glowing. Scott Barthelmy of the Universities Space Research Association at the NASA Goddard Space Flight Center has developed a system called GCN (Gamma-ray burst Coordinate Network) to transmit within seconds BATSE data on burst locations to ground-based telescopes.

BATSE consists of eight gamma-ray detectors pointing in different directions from eight corners of the Compton satellite; comparing the intensity of a burst at these detectors provides its location to roughly a few degrees but within several seconds. Often GCN can locate the burst even while it is in progress. The location is transmitted over the Internet to several dozen sites worldwide. In five more seconds, robotically controlled telescopes at Lawrence Livermore National Laboratory, among others, slew to the location for a look.

Unfortunately, only the fast-moving, smaller telescopes, which would miss a faint image, can contribute to the effort. The Livermore devices, for instance, could not have seen the afterglow of GRB 970228. Telescopes that are 100 times more sensitive are required. These mid-size telescopes would also need to be robotically controlled so they can slew very fast, and they must be capable of searching reasonably large regions. If they do find a transient afterglow, they will determine its location rather well, allowing much larger telescopes such as Hubble and Keck to look for a counterpart.

The long-lasting, afterglow following GRB 970228 gives new hope for this strategy. The HETE mission, directed by George Ricker of the Massachusetts Institute of Technology, is to be rebuilt and launched in about two years. It will survey the sky with x-ray detectors that can localize bursts to within several arc minutes. Ground-based optical telescopes will receive these locations instantly and start searching for transients.

Of course, we do not know what fraction of bursts exhibit a detectable afterglow. Moreover, even a field as small as arc minutes contains too many faint objects to make a search for counterparts easy.

To further constrain the models, we will need to look at radiation of both higher and lower frequency than that currently observed. The Compton satellite has seen a handful of bursts that emit radiation of up to 10 billion electron volts. Better data in this regime, from the Gamma-ray Large Area Space Telescope (GLAST), a satellite being developed by an international team of scientists, will greatly aid theorists. Photons of even higher energy—of about a trillion electron volts—might be captured by special ground-based gamma-ray telescopes. At the other end of the spectrum, soft x-rays, which have energies of up to roughly one kilo electron volt (keV), can help test models of bursts and obtain better fixes on position. In the range of 0.1 to 10 keV, there is a good chance of discovering absorption or emission lines that would tell volumes about the underlying fireball.

When the Hubble telescope was pointed to the location of GRB 970508, it picked up the fading light from the optical transient. Much to our surprise, however, it saw no galaxy in the immediate vicinity—not even a hint of one [*see left illustration on page 69*]. This absence emphasizes a potential problem noted by Bradley E. Schaefer of Yale University: bursts do not occur in the kind of bright galaxies within which one would expect an abundance of stars. So whereas astrophysicists now have strong evidence of the cosmological distances of bursts, we are still confounded as to their host environments and physical origins.

Just in time for New Year's 1998, nature provided a third afterglow from a gamma-ray burst. Again, Beppo-SAX discovered the initial event, following which Jules P. Halpern of Columbia University and John R. Thorstensen of Dartmouth College used the 2.4-meter telescope on Kitt Peak to find an optical transient. The glow dimmed in a manner similar to that of the previous two transients. As this article goes to press, we wait for Hubble to discern if this burst, GRB 971214, has a bright underlying galaxy or not.

The Authors

GERALD J. FISHMAN and DIETER H. HARTMANN bring complementary skills to the study of gamma-ray bursts. Fishman is an experimenter—the principal investigator for BATSE and a senior astrophysicist at the National Aeronautics and Space Administration Marshall Space Flight Center in Huntsville, Ala. He has received the NASA Medal for Exceptional Scientific Achievement three times and in 1994 was awarded the Bruno Rossi Prize of the American Astronomical Society. Hartmann is a theoretical astrophysicist at Clemson University in South Carolina; he obtained his Ph.D. in 1989 from the University of California, Santa Cruz. Apart from gamma-ray astronomy, his primary interests are the chemical dynamics and evolution of galaxies and stars. This article updates a version that appeared in *Scientific American* in July 1997.

Colossal Galactic Explosions

Enormous outpourings of gas from the centers of nearby galaxies may ultimately help explain both star formation and the intergalactic medium

by Sylvain Veilleux, Gerald Cecil and Jonathan Bland-Hawthorn

U.S. INFORMATION SERVICE

HALE OBSERVATORIES

Millions of galaxies shine in the night sky, most made visible by the combined light of their billions of stars. In a few, however, a pointlike region in the central core dwarfs the brightness of the rest of the galaxy. The details of such galactic dynamos are too small to be resolved even with the Hubble Space Telescope. Fortunately, debris from these colossal explosions—in the form of hot gas glowing at temperatures well in excess of a million degrees—sometimes appears outside the compact core, on scales that can be seen directly from Earth.

The patterns that this superheated material traces through the interstellar gas and dust surrounding the site of the explosion provide important clues to the nature and history of the powerful forces at work inside the galactic nucleus. Astronomers can now determine what kind of engines drive these dynamos and the effects of their tremendous outpourings on the intergalactic medium.

Furthermore, because such cataclysms appear to have been taking place since early in the history of the universe, they have almost certainly affected the environment in which our own Milky Way galaxy evolved. Understanding how such events take place now may illuminate the distribution of chemical elements that has proved crucial to formation of stars like the sun.

Astronomers have proposed two distinctly different mechanisms for galactic dynamos. The first was the brainchild of Martin J. Rees of the University of Cambridge and Roger D. Blandford, now at the California Institute of Technology. During the early 1970s, the two sought to explain the prodigious luminosity—thousands of times that of the Milky Way—and the spectacular "radio jets" (highly focused streams of energetic material) that stretch over millions of light-years from the centers of some hyperactive young galaxies known as quasars. They suggested that an ultramassive black hole—not much larger than the sun but with perhaps a million times its mass—could power a quasar.

A black hole itself produces essentially no light, but the disk of accreted matter spiraling in toward the hole heats up and radiates as its density increases. The inner, hotter part of the disk produces ultraviolet and x-ray photons over a broad range of energies, a small fraction of which are absorbed by the surrounding gas and reemitted as discrete spectral lines of ultraviolet and visible light. In the years since Rees and Blandford proposed their model, astronomers have come to understand that similar black holes may be responsible for the energy output of nearer active galaxies.

As the disk heats up, gas in its vicinity reaches temperatures of millions of degrees and expands outward from the galactic nucleus at high speed. This flow, an enormous cousin to the solar wind that streams away from the sun or other stars, can sweep up other interstellar gases and expel them from the nucleus. The resulting luminous shock waves can span thousands of light-years—comparable to the visible sizes of the galaxies themselves—and can be studied from space or ground-based observatories. Some of these galaxies also produce radio jets: thin streams of rapidly moving gas that emit radio waves as they traverse a magnetic field that may be anchored within the accretion disk.

Black holes are not the only engines that drive violent galactic events. Some galaxies apparently undergo short episodes of rapid star formation in their cores: so-called nuclear starbursts. The myriad new stars produce strong stellar winds and, as the stars age, a rash of supernovae. The fast-moving gas ejected from the supernovae strikes the background interstellar dust and gas and heats it to millions of degrees.

The pressure of this hot gas forms a cavity, like a steam bubble in boiling water. As the bubble expands, cooler gas and dust accumulate in a dense shell at the edge of the bubble, slowing its expansion. The transition from free flow inside the bubble to near stasis at its boundary gives rise to a zone of turbulence that is readily visible from Earth. If the energy injected into the cavity is large enough, the bubble bursts out of the galaxy's gas disk and spews the shell's fragments and hot gas into the galaxy halo or beyond, thousands of light-years away from their origins.

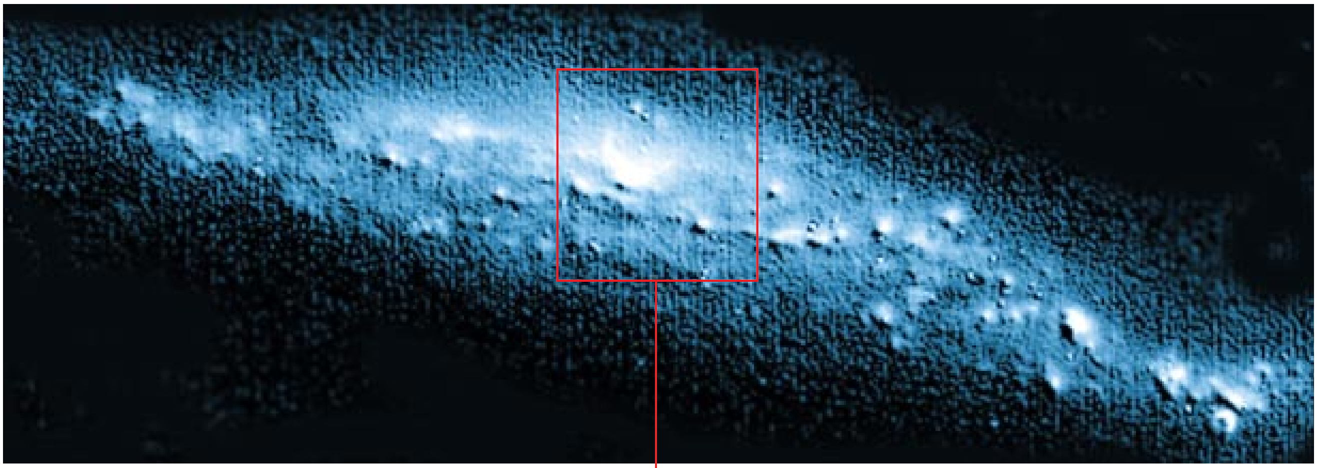
Roberto Terlevich of the Royal Greenwich Observatory and his collaborators have led the most recent research aimed at determining whether starbursts alone can drive the outpourings of hot gas characteristic of active galaxies. In 1985 Terlevich and Jorge Melnick, now at the European Southern Observatory, argued that many such galaxies contain unusual stars they dubbed "warmers"—extremely hot stars with temperatures higher than 100,000 degrees and very powerful stellar winds. Such stars, the two scientists proposed, arise naturally when a starburst occurs in a region enriched in heavy chemical elements from previous supernovae. Terlevich and his colleagues contend that their model explains the spectra and many other properties of certain active galaxies.

Identifying the Engine

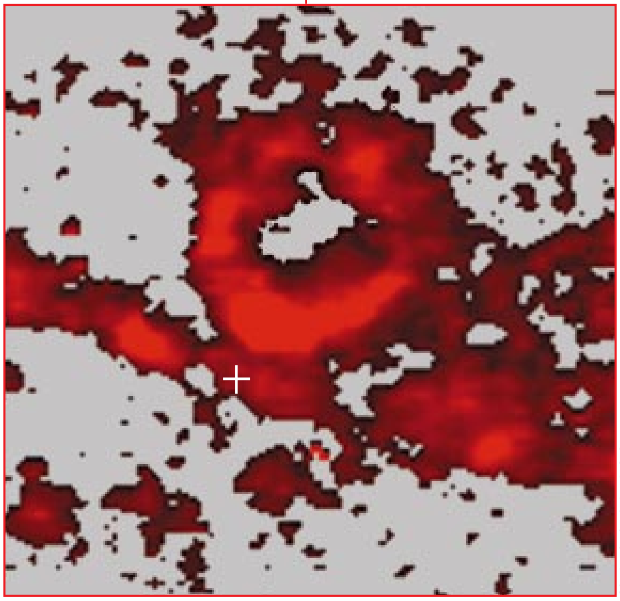
Both the starburst and the black-hole explanations appear plausible, but there are important differences between the two that may reveal which one is at work in a given galaxy. A black hole can convert as much as 10 percent of the infalling matter to energy. Starbursts, in contrast, rely on nuclear fusion, which can liberate only 0.1 percent of the

GALAXY M82 (*a, b*), about 10 million light-years away from Earth, is distinguished by an outpouring of incandescent gas from the area around its core (*c*). Astronomers have deduced that the upheaval is caused by the rapid formation of stars near the galactic nucleus. The resulting heat and radiation cause dust and gas from the galactic disk to rush into intergalactic space. The galaxy's activity may have been triggered by interaction with its neighbor, M81.

COURTESY OF PATRICK SHOPBELL



COURTESY OF GERALD CECIL



COURTESY OF GERALD CECIL

STARBURST, a sudden pulse of star formation, may be responsible for the activity of NGC 3079 (*top*) even though the galaxy has a black hole at its center. A close-up view of the area near the nucleus (*white cross*) reveals the outlines of an enormous bubble that has been blown into the interstellar medium by the heat of the stars forming at the galaxy's center.

reacting mass. As a result, they require at least 100 times as much matter, most of which accumulates as unburned fuel. Over the lifetime of a starburst-powered quasar, the total mass accumulated in the nucleus of the galaxy could reach 100 billion times the mass of the sun, equivalent to the mass of all the stars in the Milky Way galaxy.

The more mass near the nucleus, the more rapidly the orbiting stars must move. Recent ground-based near-infrared observations have revealed the presence of a dark compact object with a mass two million times that of the sun at the center of our own Milky Way. And recent radio-telescope findings have revealed an accretion disk with an inner radius of half a light-year spinning rapidly around a mass 20 million times that of the sun at the center of a nearby spiral galaxy called NGC 4258.

Several research groups are now measuring the distributions of gas and stellar motions across galactic nuclei using the recently upgraded spectrograph on board the Hubble telescope. The discovery that gas in the inner cores of the active galaxies M87 and M84 is moving in a manner consistent with a black-hole accretion disk has demonstrated how such techniques are capable of weighing the dark compact component at the centers of these objects.

Starbursts and black holes also differ in the spectra of the most energetic photons they produce. Near a black hole, the combination of a strong magnetic field and a dense accretion disk creates a soup of very fast particles that collide with one another and with photons to generate x-rays and gamma rays. A starburst, in contrast, produces most of its high-energy radiation from collisions between supernova ejecta and the surrounding galactic gas and dust. This impact heats gas to no more than about a billion degrees and so cannot produce any radiation more energetic than x-rays. The large numbers of gamma rays detected recently from some quasars by the Compton Gamma Ray Observatory imply that black holes are at their centers [see "The Compton Gamma Ray Observatory," by Neil Gehrels, Carl E. Fichtel, Gerald J. Fishman, James D. Kurfess and Volker Schönfelder; Scientific American, December 1993].

A final difference between black holes and starbursts lies in the forces that focus the flow of outrushing gas. The magnetic-field lines attached to the accretion disk around a black hole direct outflowing matter along the rotation axis of the disk in a thin jet. The material expelled by a starburst bubble, in contrast, simply follows the path of least resistance in the surrounding environment. A powerful starburst in a spiral galaxy will spew gas perpendicular to the plane of the galaxy's disk of stars and gas, but the flow will be distributed inside an hourglass-shaped region with a wide opening. The narrow radio jets that extend millions of light-years from the core of some active galaxies clearly suggest the presence of black holes.

All that we know about galaxies—active or otherwise—comes from the radiation they emit. Our observations supply the data that astrophysicists can use to choose among competing theories. The three of us have concentrated on visible light, from which we can determine the temperatures, pressures and concentrations of different atoms in the gas agitated by galactic explosions. We compare the wavelength and relative intensities of emission lines from excited or ionized atoms with those measured in terrestrial laboratories or derived from theoretical calculations.

Thanks to the Doppler shift, which changes the frequency and wavelength of light emitted by moving sources, this analysis also reveals how fast the gas is moving. Approaching gas emits light shifted toward the blue end of the spectrum, and receding gas emits light shifted toward the red end.

Until recently, astronomers unraveled gas behavior by means of two complementary methods: emission-line imaging and long-slit spectroscopy. The first produces images through a filter that selects light of a particular wavelength emitted by an element such as hydrogen. Such images often dramatically reveal the filamentary patterns of explosions, but they cannot tell observers anything about the speed or direction of the gases' motions, because the filter does not discriminate finely enough to measure redshifts or blueshifts. Long-slit spectrometers, which disperse light into its constituent colors, provide detailed information about gas motions but only over a tiny region.

For almost a decade, our group has used an instrument that combines the advantages of these two methods without the main drawbacks. The Hawaii Imaging Fabry-Perot Interferometer (HIFI) yields detailed spectral information over a large field of view. Named after the turn-of-the-century French inventors Charles Fabry and Alfred Perot, such interferometers have found wide-ranging applications in astronomy. At their heart are two glass plates that are kept perfectly parallel while separated by less than a twentieth of a millimeter. The inner surfaces of the plates are highly reflecting, so light passing through the plates is trapped into repeated reflections. Light of all but a specific wavelength—determined by the precise separation—is attenuated by destructive interference as the light waves bounce back and forth between the plates. By adjusting the separation between the plates, we can produce a series of images that are essentially a grid of spectra obtained by the interferometer at every position over the field of view.

The HIFI takes its pictures atop the 4,200-meter dormant volcano Mauna Kea, using the 2.2-meter telescope owned by the University of Hawaii and the 3.6-meter Canada-France-Hawaii instrument. The smooth airflow at the mountaintop produces sharp images. Charge-coupled devices, which are very stable and sensitive to faint light, collect the photons. In a single night, this powerful combination can generate records of up to a million spectra across the full extent of a galaxy.

We have used the HIFI to explore NGC 1068, an active spiral galaxy 46 million light-years away. As the nearest and brightest galaxy of this type visible from the Northern Hemisphere, it has been studied extensively. At radio wavelengths, NGC 1068 looks like a miniature quasar: two jets extend about 900 light-years from the core, with more diffuse emission from regions farther out. Most likely, emission from gaseous plasma moving at relativistic speeds creates the radio jets, and the "radio lobes" arise where the plasma encounters matter from the galactic disk. As might a supersonic aircraft, the leading edge of the northeast jet produces a V-shaped shock front.

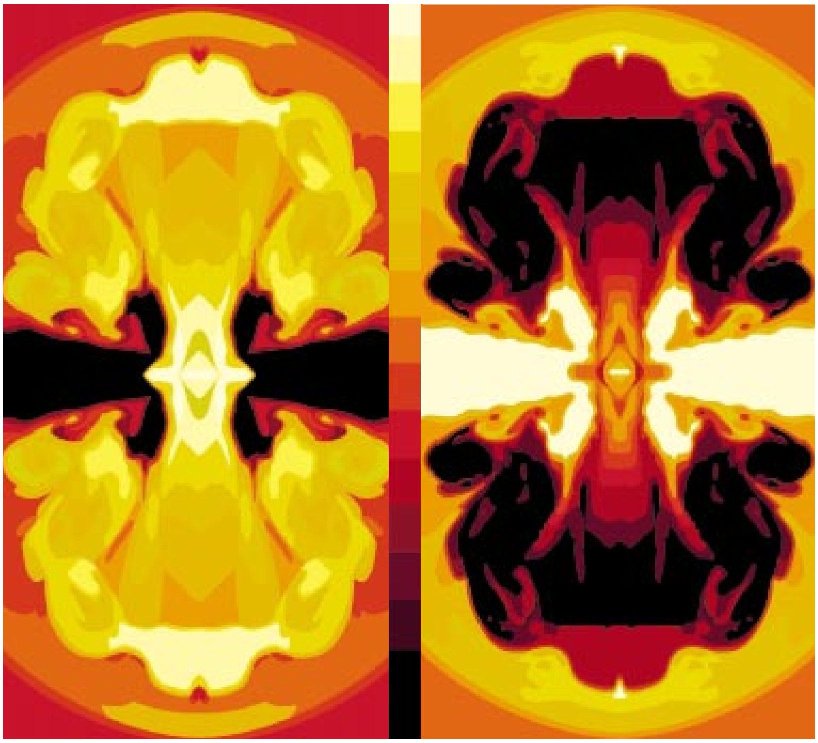
The same regions also emit large amounts of visible and ultraviolet light. We have found, however, that only 10 percent of the light comes from the nucleus. Another 5 percent comes from galaxy-disk gas that has piled up on the expanding edge of the northeast radio lobe. All the rest comes from two fans of high-velocity gas moving outward from the center at speeds of up to 1,500 kilometers per second.

The gas flows outward in two conical regions; it is probably composed of dense filaments of matter that have been swept up by the hot wind from the accretion disk. The axis of the cones of outflowing wind is tilted above the plane of the galaxy but does not point toward the poles.

The effects of the activity within the nucleus reach out several thousand light-years, well beyond the radio lobes. The diffuse interstellar gas exhibits unusually high temperatures and a large fraction of the atoms have lost one or more electrons and become ionized. At the same time, phenomena in the disk appear to influence the nucleus. Infrared images reveal an elongated bar of stars that extends more than 3,000 light-years from the nucleus. The HIFI velocity measurements suggest that the bar distorts the circular orbit of the gas in the disk, funneling material toward the center of the galaxy. This inflow of material may in fact fuel the black hole.

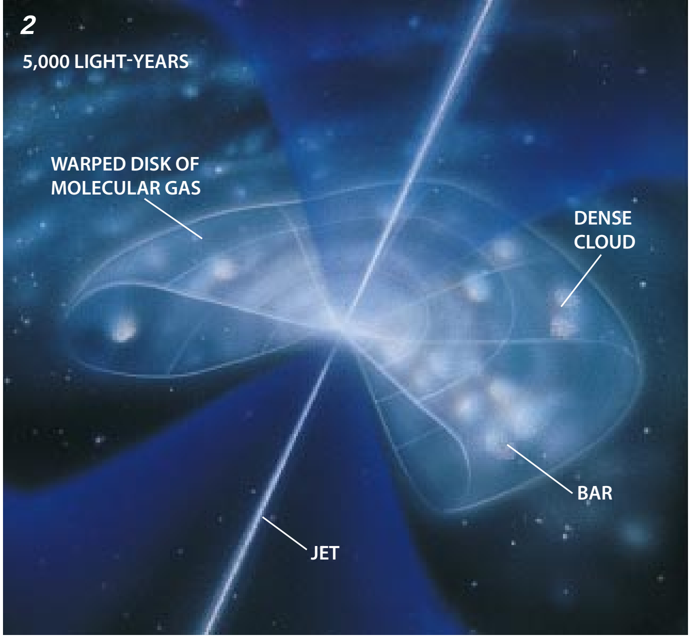
Nearby Active Galaxies

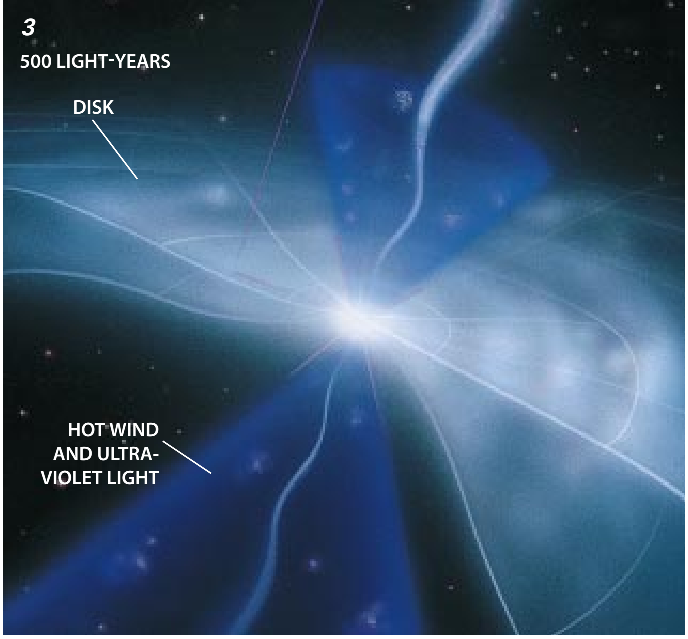
Another tremendous explosion is occurring in the core of one of our nearest neighbor galaxies, M82, just a few million light-years away. In contrast to NGC 1068, this cataclysm appears to be an archetypal starburst-driven event. Images exposed through a filter that passes the red light of forming hydrogen atoms reveal a web of filaments spraying outward along the galactic poles. Our spectral grids of emission from filaments perpendicular to the galactic disk reveal two main masses of gas, one receding and the other approaching.

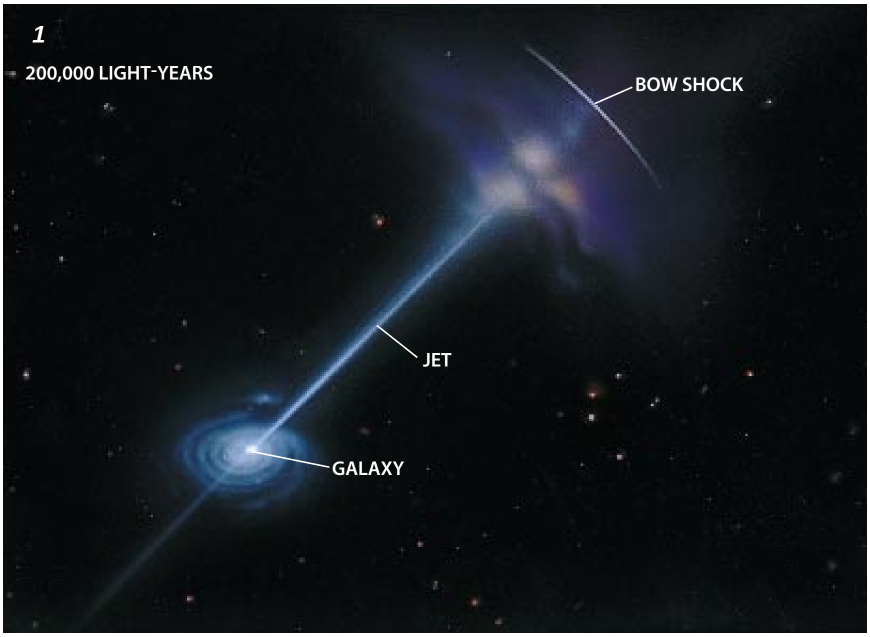


COURTESY OF GERALD CECIL, AFTER ANATOLY A. SUCHKOV ET AL.

OUTPOURING OF GAS rapidly becomes turbulent in this computer simulation of an active starburst-driven galaxy. A temperature map (*left*) shows how the hot gas emanating from the nucleus displaces the cooler galactic gas around it. The resulting shock appears clearly in a map of gas density (*right*).







ALFRED T. KAMAJIAN

COLOSSAL FORCES at work in the center of an active galaxy can make themselves felt half a million light-years or more away as jets of gas moving at relativistic speeds plow into the intergalactic medium and create enormous shock waves (*1*). Closer to the center of the galaxy (*2, 3*), a dense equatorial disk of dust and molecular gas feeds matter to the active nucleus while hot gas and radiation spill out along the poles. The high density of the infalling gas within a few dozen light-years of the center of the galaxy causes a burst of star formation (*4*). Even closer to the center (*5*), the disk, glowing at ultraviolet and x-ray wavelengths, tapers inward to feed what astronomers believe is a black hole containing millions of stellar masses but still so small as to be invisible on this scale.

The difference in velocity between the two increases as the gas moves outward from the core, reaching about 350 kilometers per second at a distance of 3,000 light-years. At a distance of 4,500 light-years from the core, the velocity separation diminishes.

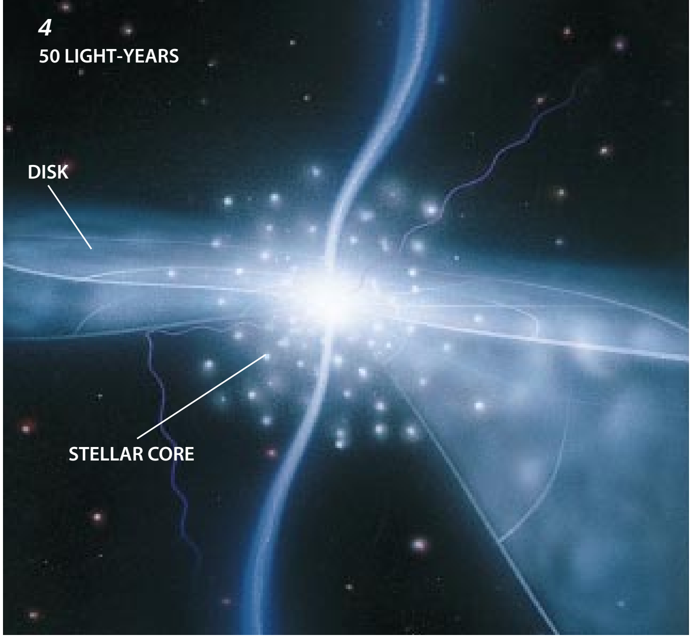
The core of M82 is undergoing an intense burst of star formation, possibly triggered by a recent orbital encounter with its neighbors M81 and NGC 3077. Its infrared luminosity is 30 billion times the total luminosity of the sun, and radio astronomers have identified the remnants of large numbers of supernovae. The filamentary web visible from Earth results from two elongated bubbles oriented roughly perpendicular to the disk of M82 and straddling the nucleus. X-ray observatories in space have detected the hot wind that inflates these bubbles; their foamy appearance probably arises from instabilities in the hot gas as it cools. The upcoming launch of the Advanced X-ray Astrophysics Facility (AXAF), the third of NASA's four planned Great Observatories, should open up exciting new avenues of research in the study of this hot-wind component.

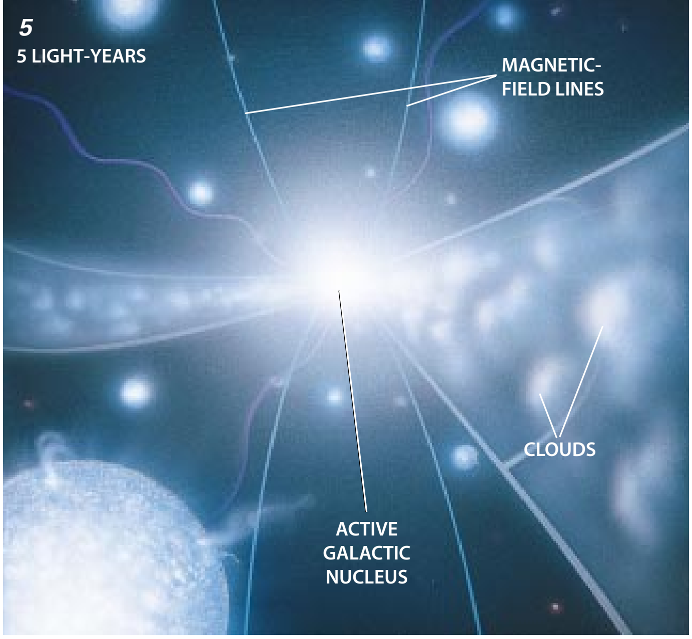
Ambiguous Activity

Unfortunately, the identity of the principal source of energy in active galaxies is not always so obvious. Sometimes a starburst appears to coexist with a black-hole engine. Like M82, many of these galaxies are abnormally bright at infrared wavelengths and rich in molecular gas, the raw material of stars. Radio emission and visual spectra resembling those of a quasar, however, suggest that a black hole may also be present.

Such ambiguity plagues interpretations of the behavior of the nearby galaxy NGC 3079. This spiral galaxy appears almost edge-on from Earth—an excellent vantage point from which to study the gas expelled from the nucleus. Like galaxy M82, NGC 3079 is anomalously bright in the infrared, and it also contains a massive disk of molecular gas spanning 8,000 light-years around its core. At the same time, the core is unusually bright at radio wavelengths, and the linear shape of radio-emitting regions near the core suggests a collimated jet outflow. On a larger scale, the radio-emission pattern is complex and extends more than 6,500 light-years from either side of the galactic disk.

Images made in red hydrogen light show a nearly circular ring 3,600 light-years across just east of the nucleus; velocity





measurements from the HIFI confirm that the ring marks the edge of a bubble as seen from the side. The bubble resembles an egg with its pointed extremity balanced on the nucleus and its long axis aligned with the galactic pole. There is another bubble on the west side of the nucleus, but most of it is hidden behind the dusty galaxy disk.

Our spectral observations imply that the total energy of this violent outflow is probably 10 times that of the explosions in NGC 1068 or M82. The alignment of the bubble along the polar axis of the host galaxy implies that galactic dust and gas, rather than a central black hole, are collimating the outflow. Nevertheless, the evidence is fairly clear that NGC 3079 contains a massive black hole at its core.

Is the nuclear starburst solely responsible for such a gigantic explosion? We have tried to answer this question by analyzing the infrared radiation coming from the starburst area. Most of the radiation from young stars embedded in molecular clouds is absorbed and reemitted in the infrared, so the infrared luminosity of NGC 3079's nucleus may be a good indicator of the rate at which supernovae and stellar winds are injecting energy at the center of the galaxy. When we compare the predictions of the starburst model with our observations, we find that the stellar ejecta appears to have enough energy to inflate the bubble. Although the black hole presumed to exist in the core of NGC 3079 may contribute to the outflow, there is no need to invoke it as an energy source.

How Active Galaxies Form

Although astronomers now understand the basic principles of operation of the engines that drive active galaxies, many details remain unclear. There is a vigorous debate about the nature of the processes that ignite a starburst or form a central black hole. What is the conveyor belt that transports fuel down to the pointlike nucleus? Most likely, gravitational interactions with gas-rich galaxies redistribute gas in the host galaxy, perhaps by forming a stellar bar such as the one in NGC 1068. Computer simulations appear to indicate that the bar, once formed, may be quite stable [see "Colliding Galaxies," by Joshua Barnes, Lars Hernquist and François Schweizer; Scientific American, August 1991]. (Indeed, the bar must be stable, because NGC 1068 currently has no close companion.)

Researchers are also divided on which comes first, nuclear starburst or black hole. Perhaps the starburst is an early phase in the evolution of active galaxies, eventually fading to leave a dense cluster of stellar remnants that rapidly coalesce into a massive black hole.

The anomalous gas flows that we and others have observed are almost certainly only particularly prominent examples of widespread, but more subtle, processes that affect many more galaxies. Luminous infrared galaxies are common, and growing evidence is leading astronomers to believe that many of their cores are also the seats of explosions. These events may profoundly affect the formation of stars throughout the galactic neighborhood. The bubble in NGC 3079, for instance, is partially ruptured at the top and so probably leaks material into the outer galactic halo or even into the vast space between galaxies. Nuclear reactions in the torrent of supernovae unleashed by the starburst enrich this hot wind in heavy chemical elements. As a result, the wind will not only heat its surroundings but also alter the environment's chemical composition.

The full impact of this "cosmic bubble bath" over the history of the universe is difficult to assess accurately because we currently know very little of the state of more distant galaxies. Images of distant galaxies taken by the Hubble will help clarify some of these questions. Indeed, as the light that left those galaxies billions of years ago reaches our instruments, we may be watching an equivalent of our own galactic prehistory unfolding elsewhere in the universe.

The Authors

SYLVAIN VEILLEUX, GERALD CECIL and JONATHAN BLAND-HAWTHORN met while working at observatories in Hawaii and were drawn to collaborate by a shared interest in peculiar galaxies. Veilleux, now an assistant professor of astronomy at the University of Maryland, received his Ph.D. from the University of California, Santa Cruz. Cecil, an associate professor of astronomy and physics at the University of North Carolina at Chapel Hill and project scientist of the SOAR four-meter telescope in Chile, received his doctorate from the University of Hawaii. Bland-Hawthorn received his Ph.D. in astronomy and astrophysics from the University of Sussex and the Royal Greenwich Observatory. He is now a research astronomer at the Anglo-Australian Observatory in Sydney. This article updates a version that appeared in *Scientific American* in February 1996.

The Ghostliest Galaxies

Astronomers have found more than 1,000 low-surface-brightness galaxies over the past decade, significantly altering our views of how galaxies evolve and how mass is distributed in the universe

by Gregory D. Bothun

Astronomers have known for decades that galaxies exist in three basic types: elliptical, spiral and irregular. The ellipticals are spheroidal, with highest light intensity at their centers. Spirals, which include our own Milky Way, have a pronounced bulge at their center, which is much like a mini-elliptical galaxy. Surrounding this bulge is a spiral-patterned disk populated with younger, bluish stars. And irregular galaxies have relatively low mass and, as their name implies, fit none of the other categories.

With only minor refinements, this system of galactic classification has changed little since astronomer Edwin Hubble originated it some 70 years ago. Technological advances, however, have significantly improved astronomers' ability to find objects outside the Milky Way galaxy that are extraordinarily hard to detect. Over the past decade my colleagues and I have used an ingenious method of photographic contrast enhancement invented by astronomer David J. Malin of the Anglo-Australian Observatory, as well as electronic imaging systems based on improved charge-coupled devices (CCDs).


LOW-SURFACE-BRIGHTNESS GALAXY Malin 1 dwarfs a conventional spiral galaxy about the size of the Milky Way, shown for scale at the upper right in this artist's conception.

ALFRED T. KAMAJIAN

Using these techniques, we have discovered that the universe contains, in addition to the other types, galaxies that, because of their extreme diffuseness, went essentially unnoticed until the mid- to late 1980s. These galaxies have the same general shape and even the same approximate number of stars as a conventional spiral galaxy. In comparison, though, the diffuse galaxies tend to be much larger, with far fewer stars per unit volume. In a conventional spiral galaxy, for example, the arms are hotbeds of stellar formation and are ordinarily populated with young stars emitting more bluish light. In the diffuse galaxies, the arms have much more gas and much less of a spiral structure. Apparently these low-surface-brightness galaxies, as they are known, take much longer to convert gas to stars. The result is galaxies that evolve four or five times more slowly; the universe literally is not old enough for these galaxies to have evolved fully.

Our work over the past decade demonstrates that, remarkably, these galaxies may be as numerous as all other galaxies combined. In other words, up to 50 percent of the general galaxy population of the universe has been missed.

Although low-surface-brightness galaxies are not numerous and massive enough to be cosmologists' long-sought dark matter, they may solve a different long-standing cosmological puzzle, concerning the baryonic mass in galaxies. Baryons are subatomic particles that are generally either protons or neutrons. They are the source of stellar—and therefore galactic—luminosity. But the amount of helium in the universe, as measured by spectroscopy, indicates that there should be far more baryons than exist in the known population of galaxies. The missing baryons may be in intergalactic space, or they may be in an unknown or difficult-to-detect population of galaxies—such as low-surface-brightness galaxies. More knowledge about these galaxies may not only settle this issue but may also force us to revise drastically our current conception of how galaxies form and evolve.

Low-surface-brightness galaxies have only recently begun shaking up the world of extragalactic astronomy, although the first temblors were felt 20 years ago. In 1976 astronomer Michael J. Disney, now at the University of Wales in Cardiff, realized that the catalogues of galaxies discovered by optical telescopes were potentially biased. Disney noted that astronomers had catalogued only the most conspicuous galaxies—those relatively detectable because they exhibited high contrast with respect to the background of the night sky. There was no reason to believe these galaxies were representative of the general population, Disney maintained. At that time, however, astronomers had not yet detected any very diffuse galaxies to substantiate Disney's suspicions. Thus, for a decade or so, the astronomical community dismissed his theory as applicable to, at most, an inconsequential population of extragalactic objects.

Ultimately, Disney was vindicated. In 1986 my colleagues and I serendipitously discovered an extremely large, low-surface-brightness disk galaxy that is the most massive (and luminous) disk galaxy yet observed. In extragalactic terms, it is fairly close—a mere 800 million light-years away. If this galaxy were as close as the spiral Andromeda galaxy (2.3 million light-years away), it would subtend an arc of fully 20 degrees in Earth's sky—40 times the apparent width of a full moon.

Why did an object this massive and nearby elude us for so many years? The answers require some background on galactic characteristics and the way astronomers measure them. Spiral galaxies have two main components: a central bulge and a surrounding disk with spiral arms. The disks usually emit light in a specific pattern, in which the intensity falls off exponentially with radial distance from the galaxy's center.

This characteristic provides astronomers with a convenient means of measuring the size of a galaxy. The scale length of a spiral galaxy (the size indicator preferred by astronomers) is a measure of the distance from the center of the galaxy to the point in the disk where the surface brightness falls to the reciprocal of *e*, the base of natural logarithms.

The other key parameter astronomers use to characterize galaxies is the central surface light intensity, which is a measure of bluish light in the center of the galaxy, an indicator of stellar density. The word "surface" in this expression refers to the fact that galaxies, which are three-dimensional, are viewed on the two-dimensional plane of the sky; thus, their brightness is projected onto this two-dimensional "surface."

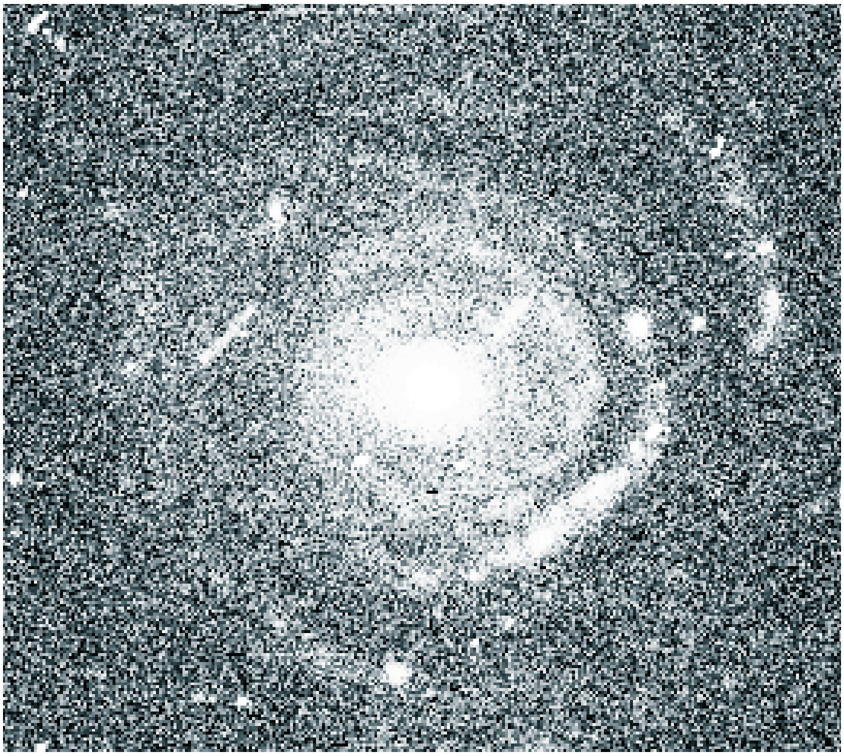
A typical spiral galaxy might have a central surface light intensity (in the blue part of the spectrum) of about 21.5 magnitudes per square arc second. For the purposes of this article, we might define a low-surface-brightness galaxy as one whose central surface light intensity has a value of at least 23 magnitudes per square arc second. (Remember, the higher the magnitude value, the less luminous the object.) To put this value into perspective, it is about equal to the brightness of the background night sky, as measured in the bluish spectrum between 400 and 500 nanometers, on a dark, moonless night at a good astronomical observing site.

Together, by simple integration, the scale length and the central surface light intensity can give us a galaxy's total mass and luminosity. Astronomers' standard catalogues of galaxies generally list them according to diameter or luminosity, as derived from scale length and central surface light intensity. As the discovery of low-surface-brightness galaxies attests, however, the complete range of galactic types is still being determined. Thus, the full range of scale lengths and central surface light intensities is not yet known. The range of these parameters is controlled by the process of galaxy formation, which remains a mystery.

Discovery and Verification

In 1984 astronomer Allan R. Sandage of the Carnegie Institution of Washington released a survey of the Virgo Cluster, which sparked our group's initial quest to locate very diffuse galaxies. In his survey, Sandage had found some very diffuse galaxies that were most likely low-mass dwarf galaxies. Pondering those images lead my colleague Chris D. Impey of the University of Arizona and me to consider whether more diffuse galaxies existed below the detection threshold of the Sandage survey. To test this hypothesis, we enlisted the aid of Malin, who provided us with contrast-enhanced prints of several regions in the Virgo Cluster. These high-contrast prints had many apparent "smudges" on them that were candidates for very diffuse galaxies.

Whereas skeptics suggested these smudges were probably artifacts (dust, water spots and so on) of Malin's photographic contrast-enhancement process, we remained uncertain. In February 1986 our first CCD ran to see if the "smudge" galaxies could be detected and verified. All of Malin's candidates turned out to be detectable in our CCD data. This finding indicated, of course, that these were real galaxies. To understand these galaxies, we had to measure their distances. Yet, because they are so faint, obtaining their optical spectra was nearly impossible. Our only hope was that these diffuse galaxies had sufficient amounts of atomic hydrogen to detect with the



COURTESY OF UNIVERSITY OF OREGON

MALINIZATION TECHNIQUE enables the imaging of low-surface-brightness galaxies. This one is known, appropriately enough, as Malin 2; it was discovered in 1990 and was the second such galaxy to be found. It is about 450 million light-years away and, with a scale length of 15 kiloparsecs, is about five times the size of the Milky Way.

Arecibo radio telescope in Puerto Rico. During the course of these radio observations in October 1986, we made a discovery.

Atomic hydrogen makes up roughly 10 percent of the baryonic mass of many galaxies and usually concentrates in the spiral arms. It was perfectly possible that some of our smudge galaxies appeared so diffuse because they were composed mostly of gas. Thus, emissions from atomic hydrogen in the smudge galaxies would corroborate their existence. One object turned out to be unique, displaying a redshift 25 times greater than that of Virgo. This was the discovery of Malin 1, an absolutely immense and extraordinarily diffuse disk galaxy. Malin 1 has a central surface light intensity only 1 percent as bright as a typical, conventional spiral. This was the first direct verification of the existence of low-surface-brightness galaxies.

Finding More Galaxies

Based on these results, Impey and I initiated three new surveys, hoping to characterize the extent and nature of this apparent population of previously undetected galaxies. The first survey relied heavily on the goodwill of James M. Schombert, at that time a postdoctoral researcher at the California Institute of Technology. Schombert was associated with the new Palomar Sky Survey and had access to the survey's plates, which he let us inspect for diffuse galaxies with sizes larger than one arc minute.

A second survey using the malinization technique was initiated in the Fornax cluster. In this survey, we could detect galaxies with central surface light intensities as low as 27 magnitudes per square arc second—a mere 2 percent brighter than the background night sky. The final survey was initiated with Michael J. Irwin of the Royal Greenwich Observatory in Cambridge, England; it was able to make use of automatic techniques to scan photographic plates.

As a result of these surveys, we detected a total of approximately 1,000 objects that we believe to be low-surface-brightness galaxies. The group includes both very small, gas-poor dwarfs and about a dozen extremely large, gas-rich objects like Malin 1. (A decade after its discovery Malin 1 remains the largest known galaxy.) In general, these galaxies span the same range of physical size, rotational velocity and mass as normal spiral galaxies. But a small percentage of the low-surface-brightness population is relatively gigantic, with scale lengths that exceed 15 kiloparsecs.

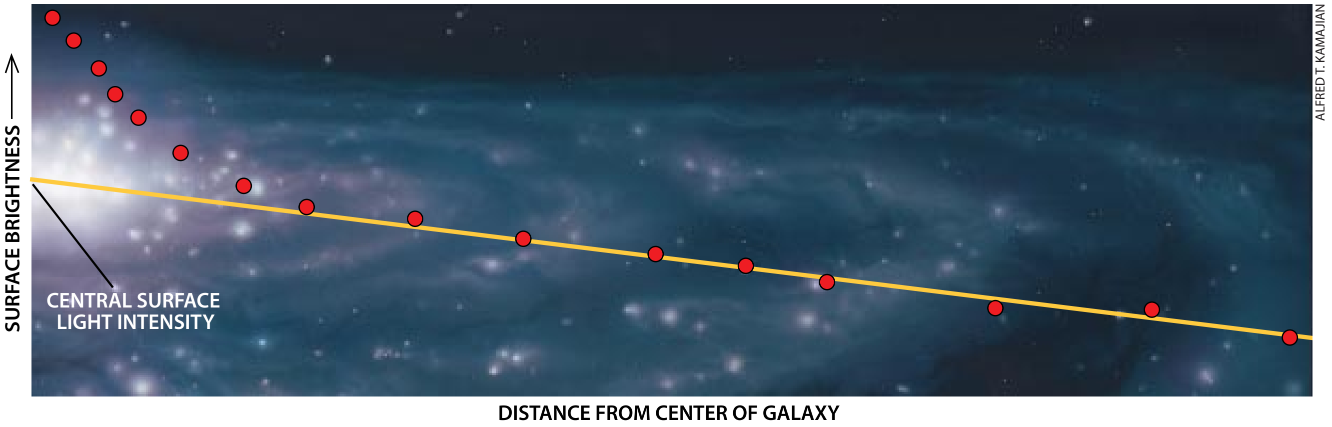
We found that in clusters of galaxies—and perhaps in the universe at large—low-surface-brightness galaxies seem to be much more numerous than conventional ones. Furthermore, if the ratio of mass to luminosity increases with decreasing surface brightness (that is, if there is more mass in less visible galaxies), then these diffuse galaxies harbor a great deal—perhaps most—of the baryonic mass in the universe.

The most startling result of these surveys has come from a recent analysis by Stacy S. McGaugh, now at the Carnegie Institution. McGaugh found that if the space density of galaxies is plotted as a function of their central surface brightness intensity, the plot is flat out to the limits of the data. In other words, there seem to be just as many very diffuse galaxies with a central surface light intensity of 27 magnitudes per square arc second as there are conventional galaxies for which this value is 21—or 23.5 or 22 or 20 and so on. This fact means that up to 50 percent of all galaxies are spirals with a central surface light intensity fainter than 22 magnitudes per square arc second.

Interestingly, low-surface-brightness galaxies are similar in several ways to the enormous number of faint, blue galaxies detected in CCD surveys of very, very distant galaxies. The two galactic types share such attributes as color, luminosity, mean surface brightness and extent of clustering. It may well be that these faint, blue galaxies are low-surface-brightness galaxies in their initial phase of star formation. At closer distances, where the objects are seen as they were in the less distant past, these objects have faded to surface brightness levels that are not intense enough for us to detect. If these faint, blue galaxies are indeed young low-surface-brightness galaxies, then there must be a still larger space density of these low-surface-brightness galaxies than is accepted at present.

This view is supported by studies of the color of low-surface-brightness galaxies, which are generally quite blue. This blueness, typically a sign of star formation, is difficult to understand. It generally indicates a galaxy that has not progressed past an early formative stage, a fact consistent with the low densities of these structures. Thus, it appears that most low-surface-brightness galaxies collapsed quite late and that their first stars formed rather late as well.

Several other findings had intriguing implications for our views about how galaxies evolve. For example, the amounts of neutral hydrogen in low-surface-brightness and convention-



SURFACE BRIGHTNESS of a spiral galaxy declines more or less exponentially with radial distance away from the galaxy's center. Past the central bulge, however, the decline in brightness is almost linear. If this linear region is extended leftward to the vertical axis, it intersects the axis at a value known as the central surface light intensity, an indicator of stellar density.

al galaxies tend to be similar, except that the low-surface-brightness galaxies have much lower densities of the gas. This and other data support the idea that a rotating gas disk must reach a minimum, or threshold, surface gas density before widespread star formation can occur. Furthermore, low-surface-brightness spirals are comparatively deficient in molecular gas.

Taken together, these observations suggest that the density of neutral hydrogen gas on the surface of the diffuse galaxies is insufficient to transform the gas into the giant molecular clouds that, in conventional galaxies, subsequently fragment to form massive stars. It seems that low-surface-brightness spiral galaxies are on a parallel evolutionary track, one in which only small stars form within lower-density clouds of neutral hydrogen gas. Because they lack massive stars, low-surface-brightness galaxies produce the heavier elements (those with atomic numbers greater than 12) at quite low rates. Ordinarily, the more massive a galaxy is, the more heavy elements it tends to contain. The fact that low-surface-brightness galaxies, regardless of their mass, are so deficient in heavy elements suggests that these diffuse galaxies are among the most unevolved objects in the universe and have changed little over the course of billions of years.

Startling Conclusions

Only during the past decade have we come to realize that up to half of all galaxies have been ignored, simply because we could not detect them through the immense noise of the night sky. Now we know that these diffuse galaxies may harbor much baryonic matter. The fact that low-surface-brightness galaxies show properties so different from normal spiral galaxies indicates that many galactic features may exist that we simply cannot detect.

Yet, given the dominance of dark matter in all galaxies, differences in their optical properties may not matter so much. Compelling new evidence now suggests that low-surface-brightness galaxies also reveal differences in the nature of their dark matter, compared with spiral galaxies.

In 1997 our team measured nearly a dozen rotation curves of low-surface-brightness disk galaxies—which differ substantially from those of high-surface-brightness rotating galaxies. In general, a galaxy's rotation speed stems from its total mass at a given radius. If most of a galaxy's mass falls near its center, then the rate at which it rotates will drop as its radius grows, much the same way that the speed of a planet's orbit falls as its distance from its host star rises.

For about 30 years, astronomers have known that most disk galaxies show a constant rotational velocity as their radii extend, indicating that the galaxy's mass grows with its radius. This observation tells us these galaxies must have dark-matter halos, containing roughly 90 percent of their total mass.

Our data have led us to two startling conclusions about low-surface-brightness galaxies. One is that their dark-matter halos extend farther and are less dense than those of spiral galaxies. Second, they contain a much smaller fraction of baryonic matter than spirals galaxies do.

Low-surface-brightness galaxies may well have fundamentally different dark-matter distributions than normal spiral galaxies do. They appear to be physically distinct from normal spirals, even though they share global properties. More important, the data also indicate that these galaxies have less baryonic matter than normal galaxies do. They are close to a hypothetical class of "dark galaxies" in which no baryons collapsed to form stars. Indeed, these galaxies may represent the tip of the iceberg of a large population of dark objects that could account for some of the universe's "missing" mass.

In just over a decade, a whole new population of galaxies has presented a unique window onto the evolution of galaxies and the distribution of matter in the universe. Over the next few years we will search for these galaxies more rigorously, with CCD surveys of wide fields of the sky at the darkest sites. In these new surveys, we should be able to find galaxies with central surface light intensities of 27 magnitudes per square arc second. SA

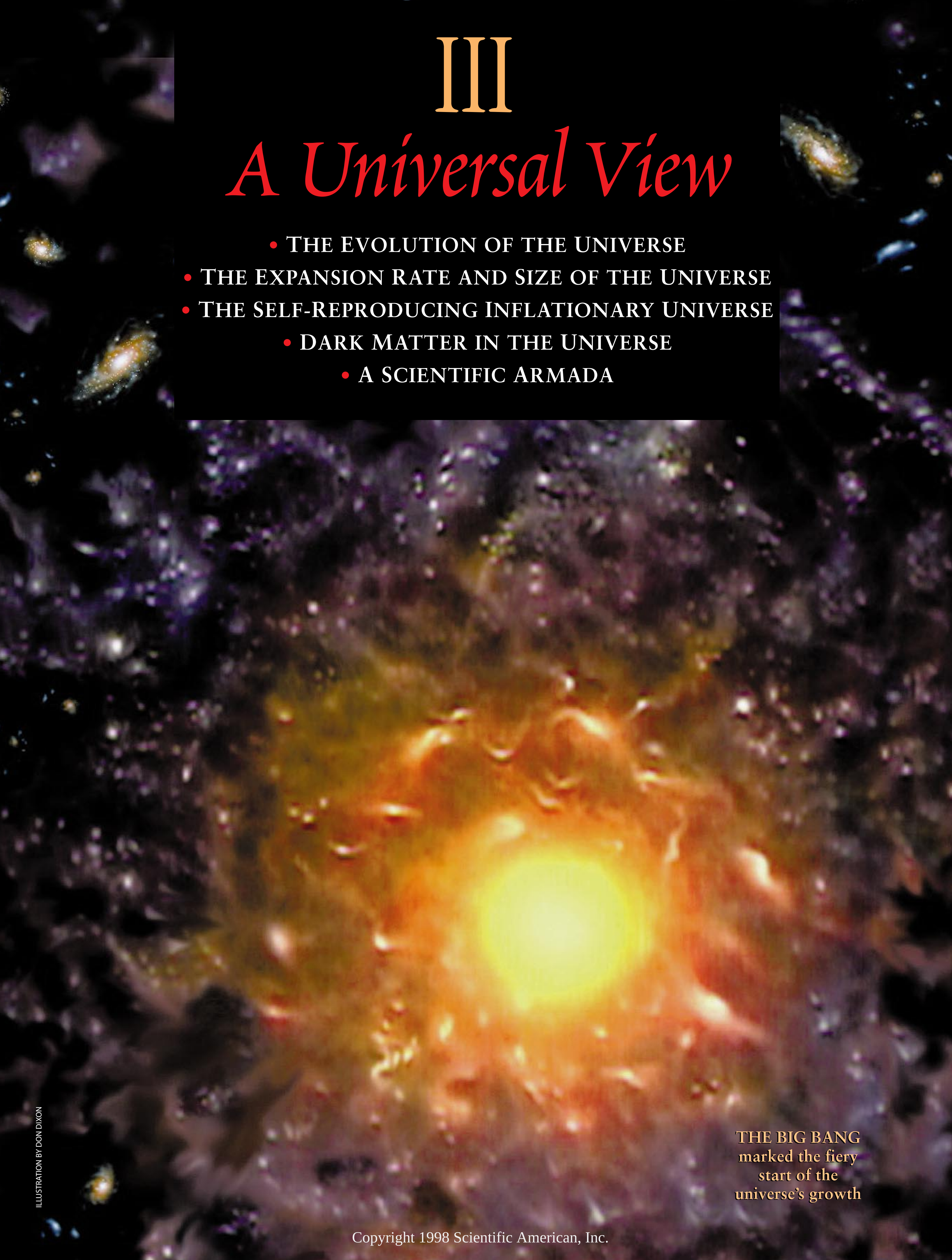
The Author

GREGORY D. BOTHUN is professor of physics at the University of Oregon. After receiving a Ph.D. in astronomy in 1981 from the University of Washington, he went on to hold positions at Harvard University, the California Institute of Technology and the University of Michigan. His research interests center on observational cosmology, especially involving large-scale structure in the universe and the formation and evolution of galaxies. He currently serves as director of Pine Mountain Observatory at the University of Oregon, which is now being transformed into a "digital observatory" accessible to the public over the Internet. This article updates a version that appeared in *Scientific American* in February 1997.

III

A Universal View

- THE EVOLUTION OF THE UNIVERSE
- THE EXPANSION RATE AND SIZE OF THE UNIVERSE
- THE SELF-REPRODUCING INFLATIONARY UNIVERSE
- DARK MATTER IN THE UNIVERSE
- A SCIENTIFIC ARMADA



THE BIG BANG marked the fiery start of the universe's growth

ILLUSTRATION BY DON DIXON

The Evolution of the

ALAN DRESSLER *Carnegie Institution/National Aeronautics and Space Administration*

by P. James E. Peebles, David N. Schramm,
Edwin L. Turner and Richard G. Kron

Universe

Some 12 billion years ago the universe emerged from a hot, dense sea of matter and energy. As the cosmos expanded and cooled, it spawned galaxies, stars, planets and life

At a particular instant roughly 12 billion years ago, all the matter and energy we can observe, concentrated in a region smaller than a dime, began to expand and cool at an incredibly rapid rate. By the time the temperature had dropped to 100 million times that of the sun's core, the forces of nature assumed their present properties, and the elementary particles known as quarks roamed freely in a sea of energy. When the universe had expanded an additional 1,000 times, all the matter we can measure filled a region the size of the solar system.

At that time, the free quarks became confined in neutrons and protons. After the universe had grown by another factor of 1,000, protons and neutrons combined to form atomic nuclei, including most of the helium and deuterium present today. All of this occurred within the first minute of the expansion. Conditions were still too hot, however, for atomic nuclei to capture electrons. Neutral atoms appeared in abundance only after the expansion had continued for 300,000 years and the universe was 1,000 times smaller than it is now. The neutral atoms then began to coalesce into gas clouds, which later evolved into stars. By the time the universe had expanded to one fifth its present size, the stars had formed groups recognizable as young galaxies.

When the universe was half its present size, nuclear reactions in stars had produced most of the heavy elements from which terrestrial planets were made. Our solar system is relatively young: it formed five billion years ago, when the universe was two thirds its present size. Over time the formation of stars has consumed the supply of gas in galaxies, and hence the population of stars is waning. Fifteen billion years from now stars like our sun will be relatively rare, making the universe a far less hospitable place for observers like us.

Our understanding of the genesis and evolution of the universe is one of the great achievements of 20th-century science. This knowledge comes from decades of innovative experiments and theories. Modern telescopes on the ground and in space detect the light from galaxies billions of light-years away, showing us what the universe looked like when it was young. Particle accelerators probe the basic physics of the high-energy environment of the early universe. Satellites detect the cosmic background radiation left over from the early stages of expansion, providing an image of the universe on the largest scales we can observe.

Our best efforts to explain this wealth of data are embodied in a theory known as the standard cosmological model or the big bang cosmology. The major claim of the theory is that in the large-scale average, the universe is expanding in a nearly homogeneous way from a dense early state. At present, there are no fundamental challenges to the big bang theory, although there are certainly unresolved issues within the theory itself. Astronomers are not sure, for example, how the galaxies were formed, but there is no reason to think the process did not occur within the framework of the big bang. Indeed, the predictions of the theory have survived all tests to date.

Yet the big bang model goes only so far, and many fundamental mysteries remain. What was the universe like before it was expanding? (No observation we have made allows us to look back beyond the moment at which the expansion began.) What will happen in the distant future, when the last of the stars exhaust the supply of nuclear fuel? No one knows the answers yet.

Our universe may be viewed in many lights—by mystics, theologians, philosophers or scientists. In science we adopt the plodding route: we accept only what is tested by experiment or observation. Albert Einstein gave us the now well-tested and accepted general theory of relativity, which establishes the relations between mass, energy, space and time. Einstein showed that a homogeneous distribution of matter in space fits nicely with his theory. He assumed without discussion that the universe is static, unchanging in the large-scale average [see "How Cosmology Became a Science," by Stephen G. Brush; Scientific American, August 1992].

GALAXY CLUSTER
is representative of what the universe looked like when it was 60 percent of its present age. The Hubble Space Telescope captured the image by focusing on the cluster as it completed 10 orbits. Several pairs of galaxies appear to be caught in one another's gravitational field. Such interactions are rarely found in nearby clusters and are evidence that the universe is evolving.

In 1922 the Russian theorist Alexander A. Friedmann realized that Einstein's universe is unstable; the slightest perturbation would cause it to expand or contract. At that time, Vesto M. Slipher of Lowell Observatory was collecting the first evidence that galaxies are actually moving apart. Then, in 1929, the eminent astronomer Edwin P. Hubble showed that the rate a galaxy is moving away from us is roughly proportional to its distance from us.

The existence of an expanding universe implies that the cosmos has evolved from a dense concentration of matter into the present broadly spread distribution of galaxies. Fred Hoyle, an English cosmologist, was the first to call this process the big bang. Hoyle intended to disparage the theory, but the name was so catchy it gained popularity. It is somewhat misleading, however, to describe the expansion as some type of explosion of matter away from some particular point in space.

That is not the picture at all: in Einstein's universe the concept of space and the distribution of matter are intimately linked; the observed expansion of the system of galaxies reveals the unfolding of space itself. An essential feature of the theory is that the average density in space declines as the universe expands; the distribution of matter forms no observable edge. In an explosion the fastest particles move out into empty space, but in the big bang cosmology, particles uniformly fill all space. The expansion of the universe has had little influence on the size of galaxies or even clusters of galaxies that are bound by gravity; space is simply opening up between them. In this sense, the expansion is similar to a rising loaf of raisin bread. The dough is analogous to space, and the raisins, to clusters of galaxies. As the dough expands, the raisins move apart. Moreover, the speed with which any two raisins move apart is directly and positively related to the amount of dough separating them.

The evidence for the expansion of the universe has been accumulating for some 60 years. The first important clue is the redshift. A galaxy emits or absorbs some wavelengths of light more strongly than others. If the galaxy is moving away from us, these emission and absorption features are shifted to longer wavelengths—that is, they become redder as the recession velocity increases.

Hubble's Law

Hubble's measurements indicated that the redshift of a distant galaxy is greater than that of one closer to Earth. This relation, now known as Hubble's law, is just what one would expect in a uniformly expanding universe. Hubble's law says the recession velocity of a galaxy is equal to its distance multiplied by a quantity called Hubble's constant. The redshift effect in nearby galaxies is relatively subtle, requiring good instrumentation to detect it. In contrast, the redshift of very distant objects—radio galaxies and quasars—is an awesome phenomenon; some appear to be moving away at greater than 90 percent of the speed of light.

Hubble contributed to another crucial part of the picture. He counted the number of visible galaxies in different directions in the sky and found that they appear to be rather uniformly distributed. The value of Hubble's constant seemed to be the same in all directions, a necessary consequence of uniform expansion. Modern surveys confirm the fundamental tenet that the universe is homogeneous on large scales. Although maps of the distribution of the nearby galaxies display clumpiness, deeper surveys reveal considerable uniformity.

The Milky Way, for instance, resides in a knot of two dozen galaxies; these in turn are part of a complex of galaxies that protrudes from the so-called local supercluster. The hierarchy of clustering has been traced up to dimensions of about 500 million light-years. The fluctuations in the average density of matter diminish as the scale of the structure being investigated increases. In maps that cover distances that reach close to the observable limit, the average density of matter changes by less than a tenth of a percent.

To test Hubble's law, astronomers need to measure distances to galaxies. One method for gauging distance is to observe the apparent brightness of a galaxy. If one galaxy is four times fainter than an otherwise comparable gal-

W. N. COLLEY, J. A. TYSON AND E. L. TURNER

MULTIPLE IMAGES of a distant galaxy, which appear as faint blue ovals, are the result of an effect known as gravitational lensing. The effect occurs when light from a distant body is bent by the gravitational field of an intervening object. In this case, the cluster of red galaxies, concentrated in the center of the picture, produces distorted images of the more distant galaxy, which lies far behind the red galaxies. The photograph was produced using the Hubble Space Telescope.

axy, then it can be estimated to be twice as far away. This expectation has now been tested over the whole of the visible range of distances.

Some critics of the theory have pointed out that a galaxy that appears to be smaller and fainter might not actually be more distant. Fortunately, there is a direct indication that objects whose redshifts are larger really are more distant. The evidence comes from observations of an effect known as gravitational lensing [*see illustration on opposite page*]. An object as massive and compact as a galaxy can act as a crude lens, producing a distorted, magnified image (or even many images) of any background radiation source that lies behind it. Such an object does so by bending the paths of light rays and other electromagnetic radiation. So if a galaxy sits in the line of sight between Earth and some distant object, it will bend the light rays from the object so that they are observable [see "Gravitational Lenses," by Edwin L. Turner; SCIENTIFIC AMERICAN, July 1988]. During the past decade, astronomers have discovered about two dozen gravitational lenses. The object behind the lens is always found to have a higher redshift than the lens itself, confirming the qualitative prediction of Hubble's law.

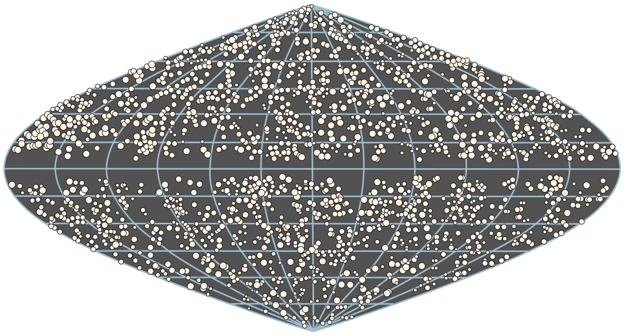
Hubble's law has great significance not only because it describes the expansion of the universe but also because it can be used to calculate the age of the cosmos. To be precise, the time elapsed since the big bang is a function of the present value of Hubble's constant and its rate of change. Astronomers have determined the approximate rate of the expansion, but no one has yet been able to measure the second value precisely.

Still, one can estimate this quantity from knowledge of the universe's average density. One expects that because gravity exerts a force that opposes expansion, galaxies would tend to move apart more slowly now than they did in the past. The rate of change in expansion is thus related to the gravitational pull of the universe set by its average density. If the density is that of just the visible material in and around galaxies, the age of the universe probably lies between 10 and 15 billion years. (The range allows for the uncertainty in the rate of expansion.)

Yet many researchers believe the density is greater than this minimum value. So-called dark matter would make up the difference. A strongly defended argument holds that the universe is just dense enough that in the remote future the expansion will slow almost to zero. Under this assumption, the age of the universe decreases to the range of seven to 13 billion years.

To improve these estimates, many astronomers are involved in intensive research to measure both the distances to galaxies and the density of the universe. Estimates of the expansion time provide an important test for the big bang model of the universe. If the theory is correct, everything in the visible universe should be younger than the expansion time computed from Hubble's law.

These two timescales do appear to be in at least rough concordance. For example, the oldest stars in the disk of the Milky Way galaxy are about nine billion years old—an estimate derived from the rate of cooling of white dwarf stars. The stars in the halo of the Milky Way are somewhat older, about 12 billion years—a value derived from the rate of nuclear fuel consumption in the cores of these stars. The ages of the oldest known chemical elements are also approximately 12 billion years—a number that comes from radioactive dating techniques. Workers in laboratories have derived these age estimates from atomic and nuclear physics. It is noteworthy that their results agree, at least approximately, with the age that astronomers have derived by measuring cosmic expansion.



JOHNNY JOHNSON, AFTER PAT McCARTHY *Carnegie Institution;* NIVR, NASA AND SERC

HOMOGENEOUS DISTRIBUTION of galaxies is apparent in a map that includes objects from 300 million to 1,000 million light-years away. The only inhomogeneity, a gap near the center line, occurs because part of the sky is obscured by the Milky Way. Michael Strauss, now at Princeton University, created the map using data from the Infrared Astronomical Satellite.

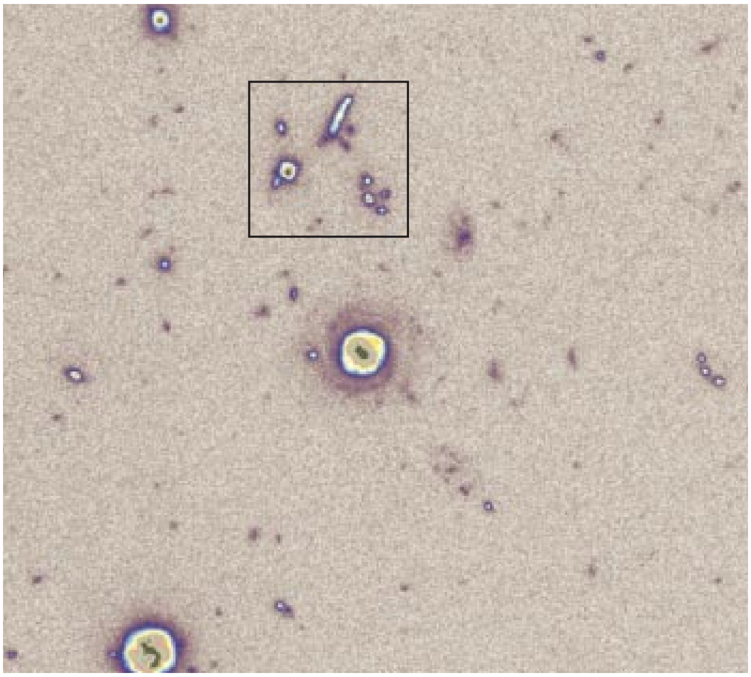
Another theory, the steady-state theory, also succeeds in accounting for the expansion and homogeneity of the universe. In 1946 three physicists in England—Hoyle, Hermann Bondi and Thomas Gold—proposed such a cosmology. In their theory the universe is forever expanding, and matter is created spontaneously to fill the voids. As this material accumulates, they suggested, it forms new stars to replace the old. This steady-state hypothesis predicts that ensembles of galaxies close to us should look statistically the same as those far away. The big bang cosmology makes a different prediction: if galaxies were all formed long ago, distant galaxies should look younger than those nearby because light from them requires a longer time to reach us. Such galaxies should contain more short-lived stars and more gas out of which future generations of stars will form.

Testing the Steady-State Hypothesis

The test is simple conceptually, but it took decades for astronomers to develop detectors sensitive enough to study distant galaxies in detail. When astronomers examine nearby galaxies that are powerful emitters of radio wavelengths, they see, at optical wavelengths, relatively round systems of stars. Distant radio galaxies, on the other hand, appear to have elongated and sometimes irregular structures. Moreover, in most distant radio galaxies, unlike the ones nearby, the distribution of light tends to be aligned with the pattern of the radio emission.

Likewise, when astronomers study the population of massive, dense clusters of galaxies, they find differences between those that are close and those far away. Distant clusters contain bluish galaxies that show evidence of ongoing star formation. Similar clusters that are nearby contain reddish galaxies in which active star formation ceased long ago. Observations made with the Hubble Space Telescope confirm that at least some of the enhanced star formation in these younger clusters may be the result of collisions between their member galaxies, a process that is much rarer in the present epoch.

So if galaxies are all moving away from one another and are evolving from earlier forms, it seems logical that they were once crowded together in some dense sea of matter and energy. Indeed, in 1927, before much was known about distant galaxies, a Belgian cosmologist and priest, Georges



DISTANT GALAXIES are visible in this blowup of a Hubble Deep Field image. The configuration in the box is 10.6 billion light-years away and thus appears as it did when the universe was only 12 percent of its present age. Some of the other galaxies shown here are closer to Earth, so this one image contains many galaxies at widely different distances, stacked up along the line of sight. Pictures such as this one provide important information about how galaxies evolve from being loose, irregular forms in the past into more regular shapes in the present epoch. (Astronomers often look at negative images like this one, in which the background is light and the stars are dark, because weak features are easier to see.)

SPACE TELESCOPE SCIENCE INSTITUTE AND NASA

Lemaître, proposed that the expansion of the universe might be traced to an exceedingly dense state he called the primeval "super-atom." It might even be possible, he thought, to detect remnant radiation from the primeval atom. But what would this radiation signature look like?

When the universe was very young and hot, radiation could not travel very far without being absorbed and emitted by some particle. This continuous exchange of energy maintained a state of thermal equilibrium; any particular region was unlikely to be much hotter or cooler than the average. When matter and energy settle to such a state, the result is a so-called thermal spectrum, where the intensity of radiation at each wavelength is a definite function of the temperature. Hence, radiation originating in the hot big bang is recognizable by its spectrum.

In fact, this thermal cosmic background radiation has been detected. While working on the development of radar in the 1940s, Robert H. Dicke, then at the Massachusetts Institute of Technology, invented the microwave radiometer—a device capable of detecting low levels of radiation. In the 1960s Bell Laboratories used a radiometer in a telescope that would track the early communications satellites Echo-1 and Telstar. The engineer who built this instrument found that it was detecting unexpected radiation. Arno A. Penzias and Robert W. Wilson identified the signal as the cosmic background radiation. It is interesting that Penzias and Wilson were led to this idea by the news that Dicke had suggested that one ought to use a radiometer to search for the cosmic background.

Astronomers have studied this radiation in great detail using the Cosmic Background Explorer (COBE) satellite and a number of rocket-launched, balloon-borne and ground-based experiments. The cosmic background radiation has two distinctive properties. First, it is nearly the same in all directions. (As the COBE team, led by John Mather of the National Aeronautics and Space Administration Goddard Space Flight Center, showed in 1992, the variation is just one part per 100,000.) The interpretation is that the radiation uniformly fills space, as predicted in the big bang cosmology. Second, the spectrum is very close to that of an object in thermal equilibrium at 2.726 kelvins above absolute zero. To be sure, the cosmic background radiation was produced when the universe was far hotter than 2.726 kelvins, yet researchers anticipated correctly that the apparent temperature of the radiation would be low. In the 1930s Richard C. Tolman of the California Institute of Technology showed that the temperature of the cosmic background would diminish because of the universe's expansion.

The cosmic background radiation provides direct evidence that the universe did expand from a dense, hot state, for this is the condition needed to produce the radiation. In the dense, hot early universe thermonuclear reactions produced elements heavier than hydrogen, including deuterium, helium and lithium. It is striking that the computed mix of the light elements agrees with the observed abundances. That is, all evidence indicates that the light elements were produced in the hot young universe, whereas the heavier elements appeared later, as products of the thermonuclear reactions that power stars.

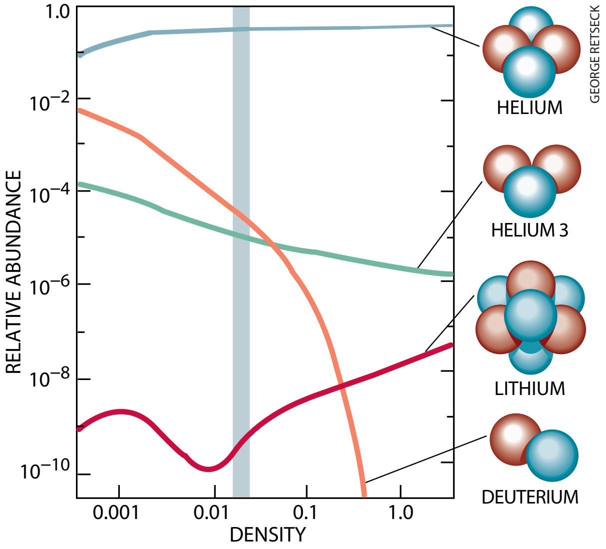
The theory for the origin of the light elements emerged from the burst of research that followed the end of World War II. George Gamow and graduate student Ralph A. Alpher of George Washington University and Robert Herman of the Johns Hopkins University Applied Physics Laboratory and others used nuclear physics data from the war effort to predict what kind of nuclear processes might have occurred in the early universe and what elements might have been produced. Alpher and Herman also realized that a remnant of the original expansion would still be detectable in the existing universe.

Despite the fact that significant details of this pioneering work were in error, it forged a link between nuclear physics and cosmology. The workers demonstrated that the early universe could be viewed as a type of thermonuclear reactor. As a result, physicists have now precisely calculated the abundances of light elements produced in the big bang and how those quantities have changed because of subsequent events in the interstellar medium and nuclear processes in stars.

Putting the Puzzle Together

Our grasp of the conditions that prevailed in the early universe does not translate into a full understanding of how galaxies formed. Nevertheless, we do have quite a few pieces of the puzzle. Gravity causes the growth of density fluctuations in the distribution of matter, because it more strongly slows the expansion of denser regions, making them grow still denser. This process is observed in the growth of nearby clusters of galaxies, and the galaxies themselves were probably assembled by the same process on a smaller scale.

The growth of structure in the early universe was prevented by radiation pressure, but that changed when the universe had expanded to about 0.1 percent of its present size. At that point, the temperature was about 3,000 kelvins, cool enough to allow the ions and electrons to combine to form neutral hydrogen and helium. The neutral matter was able to slip through the radiation and to form gas clouds that could collapse into star clusters. Observations show that by the time the universe was one fifth its present size, matter had gathered into gas clouds large enough to be called young galaxies.



GEORGE RETSECK

DENSITY OF NEUTRONS AND PROTONS in the universe determined the abundances of certain elements. For a higher-density universe, the computed helium abundance is little different, and the computed abundance of deuterium is considerably lower. The shaded region is consistent with the observations, ranging from an abundance of 24 percent for helium to one part in 10^{10} for the lithium isotope. This quantitative agreement of theory and observation is a prime success of the big bang cosmology.

A pressing challenge now is to reconcile the apparent uniformity of the early universe with the lumpy distribution of galaxies in the present universe. Astronomers know that the density of the early universe did not vary by much, because they observe only slight irregularities in the cosmic background radiation. So far it has been easy to develop theories that are consistent with the available measurements, but more critical tests are in progress. In particular, different theories for galaxy formation predict quite different fluctuations in the cosmic background radiation on angular scales less than about one degree. Measurements of such tiny fluctuations have not yet been done, but they might be accomplished in the generation of experiments now under way. It will be exciting to learn whether any of the theories of galaxy formation now under consideration survive these tests.

The present-day universe has provided ample opportunity for the development of life as we know it—there are some 100 billion billion stars similar to the sun in the part of the universe we can observe. The big bang cosmology implies, however, that life is possible only for a bounded span of time: the universe was too hot in the distant past, and it has limited resources for the future. Most galaxies are still producing new stars, but many others have already exhausted their supply of gas. Thirty billion years from now, galaxies will be much darker and filled with dead or dying stars, so there will be far fewer planets capable of supporting life as it now exists.

The universe may expand forever, in which case all the galaxies and stars will eventually grow dark and cold. The alternative to this big chill is a big crunch. If the mass of the universe is large enough, gravity will eventually reverse the expansion, and all matter and energy will be reunited. During the next decade, as researchers improve techniques for measuring the mass of the universe, we may learn whether the present expansion is headed toward a big chill or a big crunch.

In the near future, we expect new experiments to provide a better understanding of the big bang. New measurements of the expansion rate and the ages of stars are beginning to confirm that the stars are indeed younger than the expanding universe. New telescopes such as the twin 10-meter Keck telescopes in Hawaii and the 2.5-meter Hubble Space Telescope, other new telescopes at the South Pole and new satellites looking at background radiation as well as new physics experiments searching for "dark matter" may allow us to see how the mass of the universe affects the curvature of space-time, which in turn influences our observations of distant galaxies.

We will also continue to study issues that the big bang cosmology does not address. We do not know why there was a big bang or what may have existed before. We do not know whether our universe has siblings—other expanding regions well removed from what we can observe. We do not understand why the fundamental constants of nature have the values they do. Advances in particle physics suggest some interesting ways these questions might be answered; the challenge is to find experimental tests of the ideas.

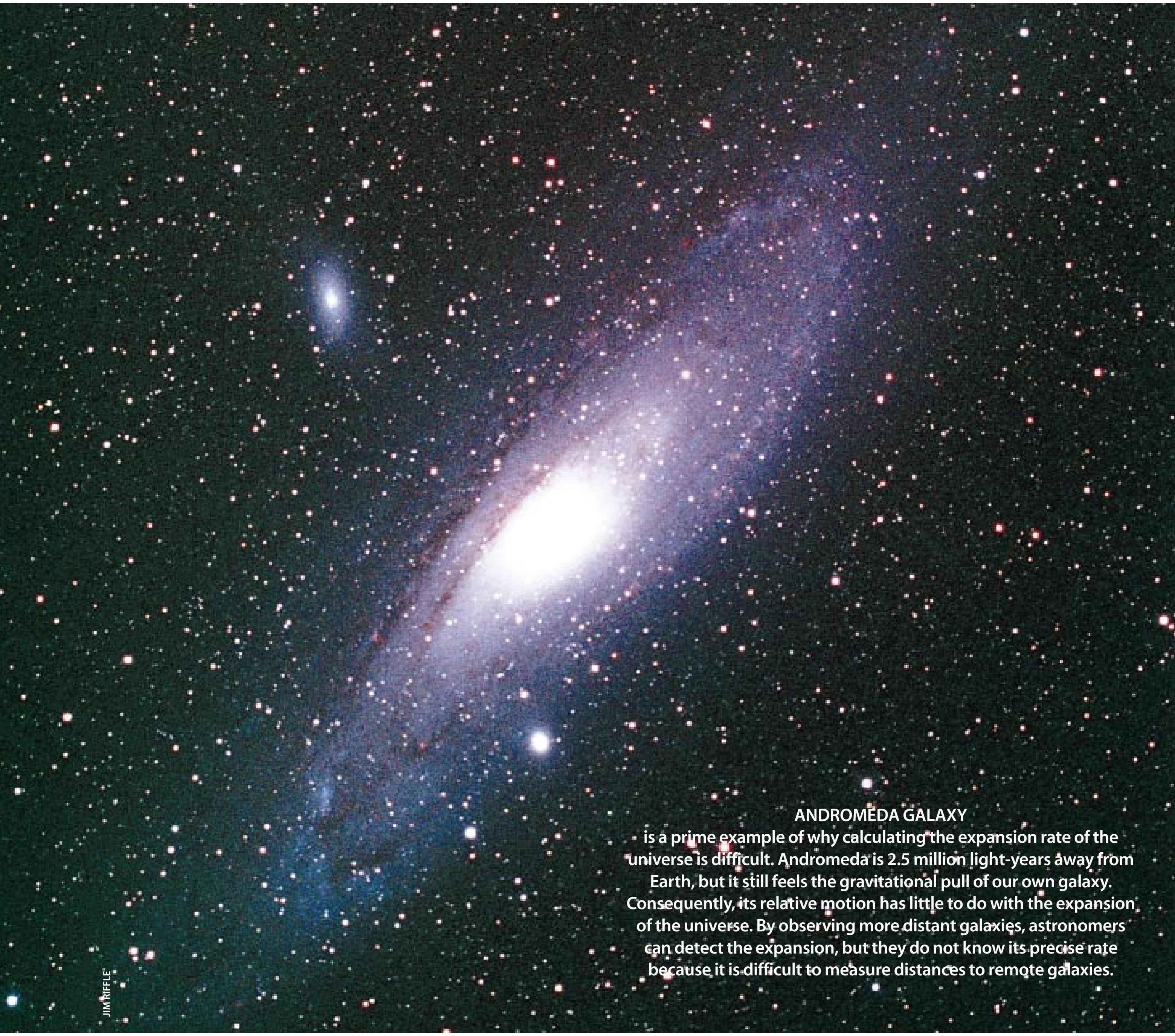
In following the debate on such matters of cosmology, one should bear in mind that all physical theories are approximations of reality that can fail if pushed too far. Physical science advances by incorporating earlier theories that are experimentally supported into larger, more encompassing frameworks. The big bang theory is supported by a wealth of evidence: it explains the cosmic background radiation, the abundances of light elements and the Hubble expansion. Thus, any new cosmology surely will include the big bang picture. Whatever developments the coming decades may bring, cosmology has moved from a branch of philosophy to a physical science where hypotheses meet the test of observation and experiment. SA

The Authors

P. JAMES E. PEEBLES, DAVID N. SCHRAMM, EDWIN L. TURNER and RICHARD G. KRON have individually earned top honors for their work on the evolution of the universe. Peebles is professor of physics at Princeton University, where in 1958 he began an illustrious career in gravitational physics. Most of his free time is spent with his three grandchildren. Turner is chair of astrophysical sciences at Princeton and director of the 3.5-meter ARC telescope in New Mexico. He has a personal, cultural and religious interest in Japan. Since 1978 Kron has served on the faculty of the department of astronomy and astrophysics at the University of Chicago, and he is also a member of the experimental astrophysics group at Fermi National Accelerator Laboratory. He enjoys observing distant galaxies almost as much as directing Yerkes Observatory near Lake Geneva, Wis. Schramm, who was Louis Block Distinguished Service Professor in the Physical Sciences and vice president for research at the University of Chicago, died in a tragic airplane accident while this special issue was being prepared for publication. This article updates a version that appeared in *Scientific American* in October 1994.

The Expansion Rate and Size of the Universe

by Wendy L. Freedman



ANDROMEDA GALAXY is a prime example of why calculating the expansion rate of the universe is difficult. Andromeda is 2.5 million light-years away from Earth, but it still feels the gravitational pull of our own galaxy. Consequently, its relative motion has little to do with the expansion of the universe. By observing more distant galaxies, astronomers can detect the expansion, but they do not know its precise rate because it is difficult to measure distances to remote galaxies.

JIM RIFFLE

Our Milky Way and all other galaxies are moving away from one another as a result of the big bang, the fiery birth of the universe. As we near the end of the millennium, it is interesting to reflect that during the 20th century, cosmologists discovered this expansion, detected the microwave background radiation from the original explosion, deduced the origin of chemical elements in the universe and mapped the large-scale structure and motion of galaxies. Despite these advances, elementary questions remain. When did the colossal expansion begin? Will the universe expand forever, or will gravity eventually halt its expansion and cause it to collapse back on itself?

For decades, cosmologists have attempted to answer such questions by measuring the universe's size-scale and expansion-rate. To accomplish this task, astronomers must determine both how fast galaxies are moving and how far away they are. Techniques for measuring the velocities of galaxies are well established, but estimating the distances to galaxies has proved far more difficult. During the past decade, several independent groups of astronomers have developed better methods for measuring the distances to galaxies, leading to completely new estimates of the expansion rate. Recently the superb resolution of the Hubble Space Telescope has extended and strengthened the calibration of the extragalactic distance scale, leading to new estimates of the expansion rate.

The age, evolution and fate of the universe depend on just how fast it is expanding. By measuring the size of the universe using a variety of new techniques, astronomers have recently improved estimates of the expansion rate

At present, several lines of evidence point toward a high expansion rate, implying that the universe is relatively young, perhaps only 10 billion years old. The evidence also suggests that the expansion of the universe may continue indefinitely. Still, many astronomers and cosmologists do not yet consider the evidence definitive. We actively debate the merits of our techniques.

An accurate measurement of the expansion rate is essential not only for determining the age of the universe and its fate but also for constraining theories of cosmology and models of galaxy formation. Furthermore, the expansion rate is important for estimating fundamental quantities, from the density of the lightest elements (such as hydrogen and helium) to the amount of nonluminous matter in galaxies, as well as clusters of galaxies. Because we need accurate distance measurements to calculate the luminosity, mass and size of astronomical objects, the issue of the cosmological distance scale, or the expansion rate, affects the entire field of extragalactic astronomy.

Astronomers began measuring the expansion rate of the universe some 70 years ago. In 1929 the eminent astronomer Edwin P. Hubble of the Carnegie Institution's observatories made the remarkable observation that the velocity of a galaxy's recession is proportional to its distance. His observations provided the first evidence that the entire universe is expanding.

The Hubble Constant

Hubble was the first to determine the expansion rate. Later this quantity became known as the Hubble constant: the recession velocity of the galaxy divided by its distance. A very rough estimate of the Hubble constant is 100 kilometers per second per megaparsec. (Astronomers commonly represent distances in terms of megaparsecs, where one megaparsec is the distance light travels in 3.26 million years.) Thus, a typical galaxy at a distance of 50 megaparsecs moves away at about 5,000 kilometers (3,000 miles) per second. A galaxy at 500 megaparsecs therefore moves at about 50,000 kilometers per second, or more than 100 million miles per hour!

For seven decades, astronomers have hotly debated the precise value of the expansion rate. Hubble originally obtained a value of 500 kilometers per second per megaparsec (km/s/Mpc). After Hubble's death in 1953, his protégé Allan R. Sandage, also at Carnegie, continued to map the expansion of the universe. As Sandage and others made more accurate and extensive observations, they revised Hubble's original value downward into the range of 50 to 100 km/s/Mpc, thereby indicating a universe far older and larger than suggested by the earliest measurements.

During the past two decades, new estimates of the Hubble constant have continued to fall within this same range, but preferentially toward the two extremes. Notably, Sandage and his longtime collaborator Gustav A. Tammann of the University of Basel have argued for a value of 50 km/s/Mpc, whereas the late Gérard de Vaucouleurs of the University of Texas advocated a value of 100 km/s/Mpc. The controversy has created an unsatisfactory situation in which scientists have been free to choose any value of the Hubble constant between the two extremes.

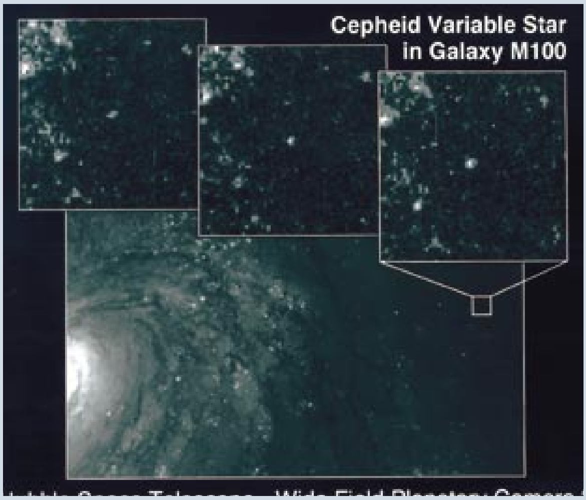
In principle, determining the Hubble constant is simple, requiring only a measurement of velocity and distance. Measuring a galaxy's velocity is straightforward: Astronomers disperse light from a galaxy and record its spectrum. A galaxy's spectrum has discrete spectral lines, which occur at characteristic wavelengths caused by emission or absorption of elements in the gas and stars making up the galaxy. For a galaxy receding from Earth, these spectral lines shift to longer wavelengths by an amount proportional to the velocity—an effect known as redshift.

If the measurement of the Hubble constant is so simple in principle, then why has it remained one of the outstanding problems in cosmology for almost 70 years? In practice, measuring the Hubble constant is extraordinarily difficult, primarily for two reasons. First, although we can measure their velocities accurately, galaxies interact gravitationally with their neighbors. In so doing, their velocities become perturbed, inducing "peculiar" motions that are superimposed onto the general expansion of the universe. Second, establishing an accurate distance scale has turned out to be

Why Cepheid Variables Pulsate

Several times more massive than the sun, a Cepheid variable is a relatively young star whose luminosity changes in a periodic way: a Cepheid brightens and then dims more slowly over a period of a few days to months. It pulsates because the force of gravity acting on the atmosphere of the star is not quite balanced by the pressure of the hot gases from the interior of the star.

The imbalance occurs because of changes in the atmosphere of a Cepheid. An important ingredient in the atmosphere is singly ionized helium (that is, helium atoms that have lost a single electron). As radiation flows out of the interior of a Cepheid, singly ionized helium in the atmosphere absorbs and scatters radiation, and it may become doubly ionized (that is, each helium atom releases a second electron). Consequently, the atmosphere becomes more opaque, making it difficult for radiation to escape from the atmosphere. This interaction between radiation and matter generates a pressure that forcefully pushes out the atmosphere of the star. As a result, the Cepheid variable increases in size and in brightness.



CEPHEID VARIABLE in the galaxy M100 is shown here at three different times in its light cycle (*top*). As seen from left to right, the star brightens.

WENDY L. FREEDMAN *Carnegie Institution's Observatories* AND NASA

Yet as the atmosphere expands, it also cools, and at lower temperatures the helium returns to its singly ionized state. Hence, the atmosphere allows radiation to pass through more freely, and the pressure on the atmosphere decreases. Eventually, the atmosphere collapses back to its initial size, and the Cepheid returns to its original brightness. The cycle then repeats.

—*W.L.F.*

much more difficult than anticipated. Consequently, an accurate measure of the Hubble constant requires us not only to establish an accurate extragalactic distance scale but also to do this already difficult task at distances great enough that peculiar motions of galaxies are small compared with the overall expansion, or Hubble flow. To determine the distance to a galaxy, astronomers must choose from a variety of complicated methods. Each has its advantages, but none is perfect.

Measuring Distances to Galaxies

Astronomers can most accurately measure distances to nearby galaxies by monitoring a type of star commonly known as a Cepheid variable. Over time, the star changes in brightness in a periodic and distinctive way. During the first part of the cycle, its luminosity increases very rapidly, whereas during the remainder of the cycle, the luminosity of the Cepheid decreases slowly. On average, Cepheid variables are about 10,000 times brighter than the sun.

Remarkably, the distance to a Cepheid can be calculated from its period (the length of its cycle) and its average apparent brightness (its luminosity as observed from Earth). In 1908 Henrietta S. Leavitt of Harvard College Observatory discovered that the period of a Cepheid correlates closely with its brightness. She found that the longer the period, the brighter the star. This relation arises from the fact that a Cepheid's brightness is proportional to its surface area. Large, bright Cepheids pulsate over a long period just as, for example, large bells resonate at a low frequency (or longer period).

By observing a Cepheid's variations in luminosity over time, astronomers can obtain its period and average apparent luminosity, thereby calculating its absolute luminosity (that is, the apparent brightness the star would have if it were a standard distance of 10 parsecs away). Furthermore, they know that the apparent luminosity decreases as the distance it travels increases—because the apparent luminosity falls off in proportion to the square of the distance to an object. Therefore, we can compute the distance to the Cepheid from the ratio of its absolute brightness to its apparent brightness.

During the 1920s, Hubble used Cepheid variables to establish that other galaxies existed far beyond the Milky Way. By measuring apparent brightnesses and periods of faint, starlike images that he discovered on photographs of objects such as the Andromeda Nebula (also known as M31), the Triangulum Nebula (M33) and NGC 6882, he could show that these objects were located more than several hundred thousand light-years from the sun, well outside the Milky Way. From the 1930s to the 1960s, Hubble, Sandage and others struggled to find Cepheids in nearby galaxies. They succeeded in measuring the distances to about a dozen galaxies. About half these galaxies are useful for the derivation of the Hubble constant.

One of the difficulties with the Cepheid method is that dust between stars diminishes apparent luminosity. Dust particles absorb, scatter and redden light from all types of stars. Another complication is that it is hard to establish how Cepheids of different chemical element abundances differ in brightness. The effects of both dust and element abundances are most severe for blue and ultraviolet light. Astronomers must either observe Cepheids at infrared wavelengths, where the effects are less significant, or observe them at many different optical wavelengths so that they can assess the effects and correct for them.

During the 1980s, my collaborator (and husband) Barry F. Madore of the California Institute of Technology and I remeasured the distances to the nearest galaxies using charge-coupled devices (CCDs) and the large reflecting telescopes at many sites, including Mauna Kea in Hawaii, Las Campanas in Chile and Mount Palomar in California. As a result, we determined the distances to nearby galaxies with much greater accuracy than has been done before.

These new CCD observations proved critical to correct for the effects of dust and to improve previous photographic photometry. In some cases, we revised distances to nearby galaxies downward by a factor of two. Were it feasible, we would use Cepheids directly to measure distances associated with the universe's expansion. Unfortunately, so far we cannot detect Cepheids in galaxies sufficiently far away so that we know they are part of a "pure" Hubble expansion of the universe.

Nevertheless, astronomers have developed several other methods for measuring relative distances between galaxies on vast scales, well beyond Cepheid range. Because we must use the Cepheid distance scale to calibrate these techniques, they are considered secondary distance indicators.

During the past decade, astronomers have made great strides developing techniques to measure such relative distances. These methods include observing and measuring a special category of supernovae: catastrophic explosions signaling the death of certain low-mass stars. Sandage and his collaborators are now determining the Hubble constant by studying such supernovae based on the calibration of Cepheids. Other secondary distance-determining methods include measuring the brightnesses and rotations of velocities of entire spiral galaxies, the fluctuations (or graininess) in the light of elliptical galaxies, and the analysis and measurement of the expansion properties of another category of younger, more massive supernovae. The key to measuring the Hubble constant using these techniques is to determine the distance to selected galaxies using Cepheids; their distances can, in turn, be used to calibrate the relative extragalactic distance scale by applying secondary methods.

Yet scientists have not reached a consensus about which, if any, secondary indicators are reliable. As the saying goes, "the devil is in the details." Astronomers disagree on how to apply these methods, whether they should be adjusted for various effects that might bias the results, and what the true uncertainties are. Differences in the choice of secondary methods lie at the root of most current debates about the Hubble constant.

COURTESY OF HARVARD UNIVERSITY ARCHIVES

HENRIETTA S. LEAVITT of Harvard College Observatory found, in 1908, a correlation between the period of a Cepheid variable and its absolute brightness. This correlation allows astronomers to measure distances to the nearest galaxies.

Establishing a Distance Scale

One technique for measuring great distances, the Tully-Fisher relation, relies on a correlation between a galaxy's brightness and its rotation rate. High-luminosity galaxies typically have more mass than low-luminosity galaxies, and so bright galaxies rotate slower than dim galaxies. Several groups have tested the Tully-Fisher method and shown that the relation does not appear to depend on environment; it remains the same in the dense and outer parts of rich clusters and for relatively isolated galaxies. The Tully-Fisher relation can be used to estimate distances as far away as 300 million light-years. A disadvantage is that astronomers lack a detailed theoretical understanding of the Tully-Fisher relation.

Another distance indicator that has great potential is a particular kind of supernova known as type Ia. Type Ia supernovae, astronomers believe, occur in double-star systems in which one of the stars is a very dense object known as a white dwarf. When a companion star transfers its mass to a white dwarf, it triggers an explosion. Because supernovae release tremendous amounts of radiation, astronomers should be able to see supernovae as far away as five billion light-years—that is, a distance spanning a radius of half the visible universe.

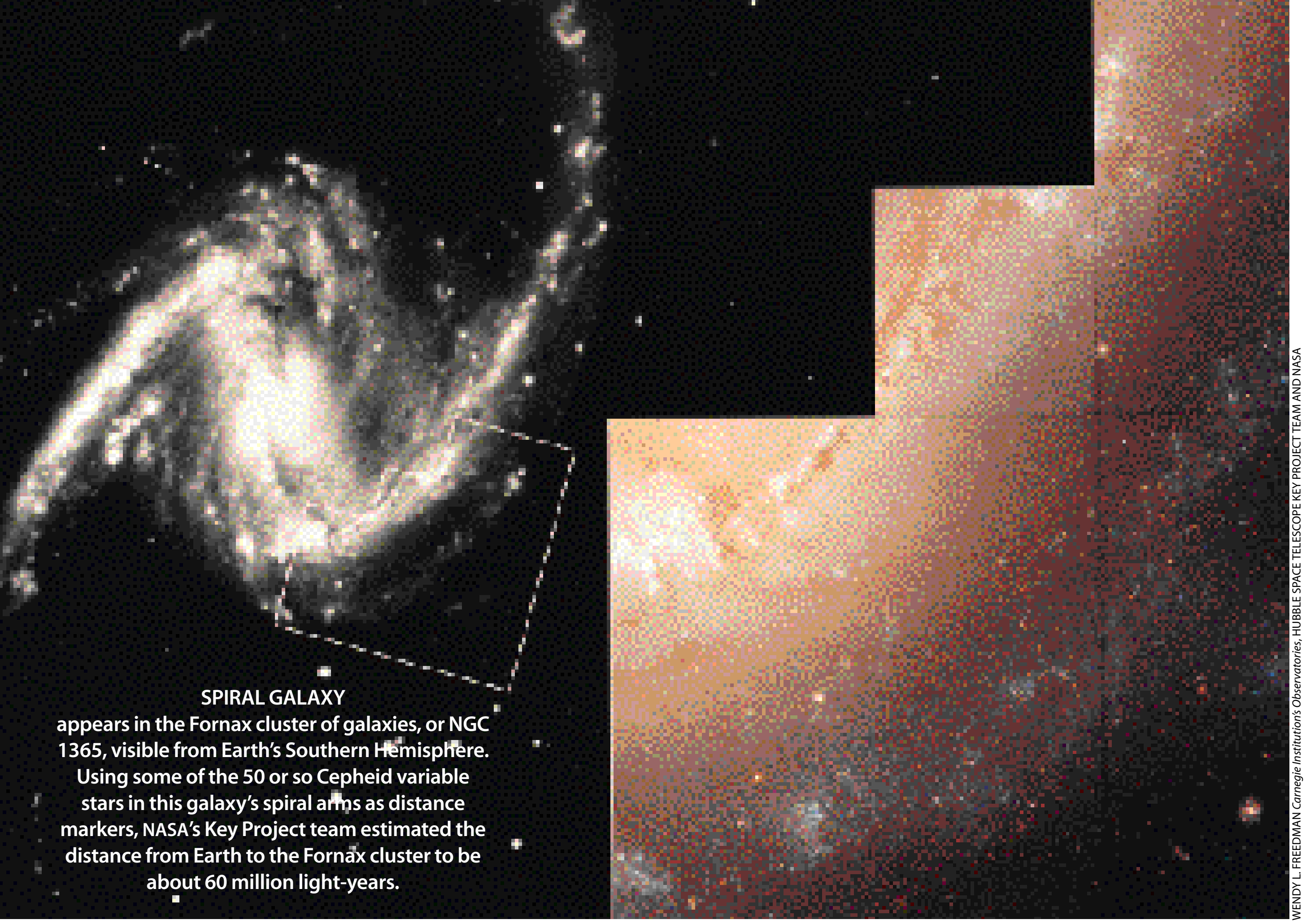
Type Ia supernovae make good distance indicators because, at the peak of their brightness, they all produce roughly the same amount of light. Using this information, astronomers can infer their distance.

If supernovae are also observed in galaxies for which Cepheid distances can be measured, then the brightnesses of supernovae can be used to infer distances. In practice, however, the brightnesses of supernovae are not all the same; there is a range of brightnesses that must be taken into account. A difficulty is that supernovae are very rare events, so the chance of seeing one nearby is very small. Unfortunately, a current limitation of this method is that about half of all supernovae observed in galaxies close enough to have Cepheid distances were observed decades ago, and these measurements are of low quality.

An interesting method, developed by John L. Tonry of the Massachusetts Institute of Technology and his colleagues, exploits the fact that nearby galaxies appear grainy, whereas remote galaxies are more uniform in their surface-brightness distribution. The graininess decreases with distance because the task of resolving individual stars becomes increasingly difficult. Hence, the distance to a galaxy can be gauged by how much the apparent brightness of the galaxy fluctuates over its surface. This method cannot currently extend as far as the Tully-Fisher relation or supernovae, but it and other methods offer an important, independent way to test and compare relative distances. These comparisons yield excellent agreement, representing one of the most important advancements in recent years.

For decades, astronomers have recognized that the solution to the impasse on the extragalactic distance scale would require observations made at very high spatial resolution. The Hubble telescope can now resolve Cepheids at distances 10 times farther (and therefore in a volume 1,000 times larger) than we can do from the ground. A primary motivation for building an orbiting optical telescope was to enable the discovery of Cepheids in remoter galaxies and to measure accurately the Hubble constant.

More than a decade ago several colleagues and I were awarded time on the Hubble telescope to undertake this project. This program involves 26 astronomers, led by me, Jeremy R. Mould of Mount Stromlo and Siding Springs Observatory, and Robert C. Kennicutt of Steward Obser-



SPIRAL GALAXY appears in the Fornax cluster of galaxies, or NGC 1365, visible from Earth's Southern Hemisphere. Using some of the 50 or so Cepheid variable stars in this galaxy's spiral arms as distance markers, NASA's Key Project team estimated the distance from Earth to the Fornax cluster to be about 60 million light-years.

WENDY L. FREEDMAN *Carnegie Institution's Observatories*, HUBBLE SPACE TELESCOPE KEY PROJECT TEAM AND NASA

vatory. Our effort involves measuring Cepheid distances to about 20 galaxies, enough to calibrate a wide range of secondary distance methods. We aim to compare and contrast results from many techniques and to assess the true uncertainties in the measurement of the Hubble constant.

Though still incomplete, new Cepheid distances to a dozen galaxies have been measured as part of this project. Preliminary results yield a value of the Hubble constant of about 70 km/s/Mpc with an uncertainty of about 15 percent. This value is based on a number of methods, including the Tully-Fisher relation, type Ia supernovae, type II supernovae, surface-brightness fluctuations, and Cepheid measurements to galaxies in the nearby Virgo and Fornax clusters.

Sandage and his collaborators have reported a value of 59 km/s/Mpc, based on type Ia supernovae. Other groups (including our own) have found a value in the middle 60 range, based on the same type Ia supernovae. Nevertheless, these current disagreements are much smaller than the earlier discrepancies of a factor of two, which have existed until now. This progress is encouraging.

Two other methods for determining the Hubble constant spark considerable interest because they do not involve the Cepheid distance scale and can be used to measure distances on vast cosmological scales. The first of these alternative methods relies on an effect called gravitational lensing: if light from some distant source travels near a galaxy on its way to Earth, the light can be deflected, as a result of gravity, according to Einstein's general theory of relativity. The light may take many different paths around the galaxy, some shorter, some longer, and consequently arrives at Earth at different times. If the brightness of the source varies in some distinctive way, the signal will be seen first in the light that takes the shortest path and will be observed again, some time later, in the light that traverses the longest path. The difference in the arrival times reveals the difference in length between the two light paths. By applying a theoretical model of the mass distribution of the galaxy, astronomers can calculate a value for the Hubble constant.

The second method uses a phenomenon known as the Sunyaev-Zel'dovich (SZ) effect. When photons from the microwave background travel through galaxy clusters, they can gain energy as they scatter off the hot plasma (x-ray) electrons found in the clusters. The net result of the scattering is a decrease in the microwave background toward the position of the cluster. By comparing the microwave and x-ray distributions, a distance to the cluster can be inferred. To determine the distance, however, astronomers must also know the average density of the electrons, as well as their distribution and temperature, and have an accurate measure of the decrement in the temperature of the microwave background. By calculating the distance to the cluster and measuring its recessional velocity, astronomers can then obtain the Hubble constant.

The SZ method and the gravitational-lensing technique are promising. Yet, to date, few objects are available with the required characteristics. Hence, these methods have not yet been tested rigorously. Fortunately, impressive progress is being made in both these areas with large, new surveys. Current applications of these methods result in values of the Hubble constant in the range of 40 to 80 km/s/Mpc.

The debate continues regarding the best method for determining distances to remote galaxies. Consequently, astronomers hold many conflicting opinions about what the best current estimate is for the Hubble constant.

How Old Is the Universe?

The value of the Hubble constant has many implications for the age, evolution and fate of the universe. A low value for the Hubble constant implies an old age for the universe, whereas a high value suggests a young age. For example, a value of 100 km/s/Mpc indicates the universe is about 6.5 to 8.5 billion years old (depending on the amount of matter in the universe and the corresponding deceleration caused by that matter). A value of 50 km/s/Mpc suggests, however, an age of 13 to 16.5 billion years.

And what of the ultimate fate of the universe? If the average density of matter in the universe is low, as current observations indicate, the standard cosmological model predicts that the universe will expand forever.

Nevertheless, theory and observations suggest that the universe contains more mass than what can be attributed to luminous matter. A very active area of cosmological research is the search for this additional "dark" matter in the universe. To answer the question about the fate of the universe unambiguously, cosmologists require not only a knowledge of the Hubble constant and the average mass density of the universe but also an independent measure of the age of the universe. These three quantities are needed to specify uniquely the geometry and the evolution of the universe.

> **The value of the Hubble constant has many implications for the fate of the universe.**

If the Hubble constant turns out to be high, it would have profound implications for our understanding of the evolution of galaxies and the universe. A Hubble constant of 70 km/s/Mpc yields an age estimate of nine to 12 billion years (allowing for uncertainty in the value of the average density of the universe). A high-density universe corresponds to an age of about nine billion years. A low-density universe corresponds to an age of about 12 billion years for this same value of the Hubble constant.

These estimates are all shorter than what theoretical models suggest for the age of old stellar systems known as globular clusters. Globular clusters are believed to be among the first objects to form in our galaxy, and their age is estimated to be between 13 and 17 billion years. Obviously, the ages of the globular clusters cannot be older than the age of the universe itself.

Age estimates for globular clusters are often cited as a reason to prefer a low value for the Hubble constant and therefore an older age of the universe. Some astronomers argue, however, that the theoretical models of globular clusters on which these estimates depend may not be complete and may be based on inaccurate assumptions. For instance, the models rely on knowing precise ratios of certain elements in globular clusters, particularly oxygen and iron. Moreover, accurate ages require accurate measures of luminosities of globular cluster stars, which in turn require accurate measurements of the distances to the globular clusters.

Recent measurements from the Hipparcos satellite suggest that the distances to globular clusters might have to be increased slightly. The resulting effect of this change, if confirmed, would be to lower the globular cluster ages, perhaps to 11 or 12 billion years. Given the current uncertainties in the measurements of both the Hubble constant and the models and distances for globular clusters, these new results may indicate that no serious discrepancy exists between the age of the universe, based on expansion, and the age of globular clusters.

In any case, these subtle inconsistencies highlight the importance of accurate distance measurements, not only for studying galaxies and determining the Hubble constant but also for understanding globular clusters and their ages.

A high value for the Hubble constant raises another potentially serious problem: it disagrees with standard theories of how galaxies are formed and distributed in space. For example, the theories predict how much time is required for large-scale clustering, which has been observed in the distribution of galaxies, to occur. If the Hubble constant is large (that is, the universe is young), the models cannot reproduce the observed distribution of galaxies.

Scientists are excited about results in the next decade. The recently installed NICMOS infrared camera on the Hubble telescope will allow us to refine the Cepheid distances measured so far. Large, ground-based telescope surveys will increase the number of galaxies for which we can measure relative distances beyond the reach of Cepheids.

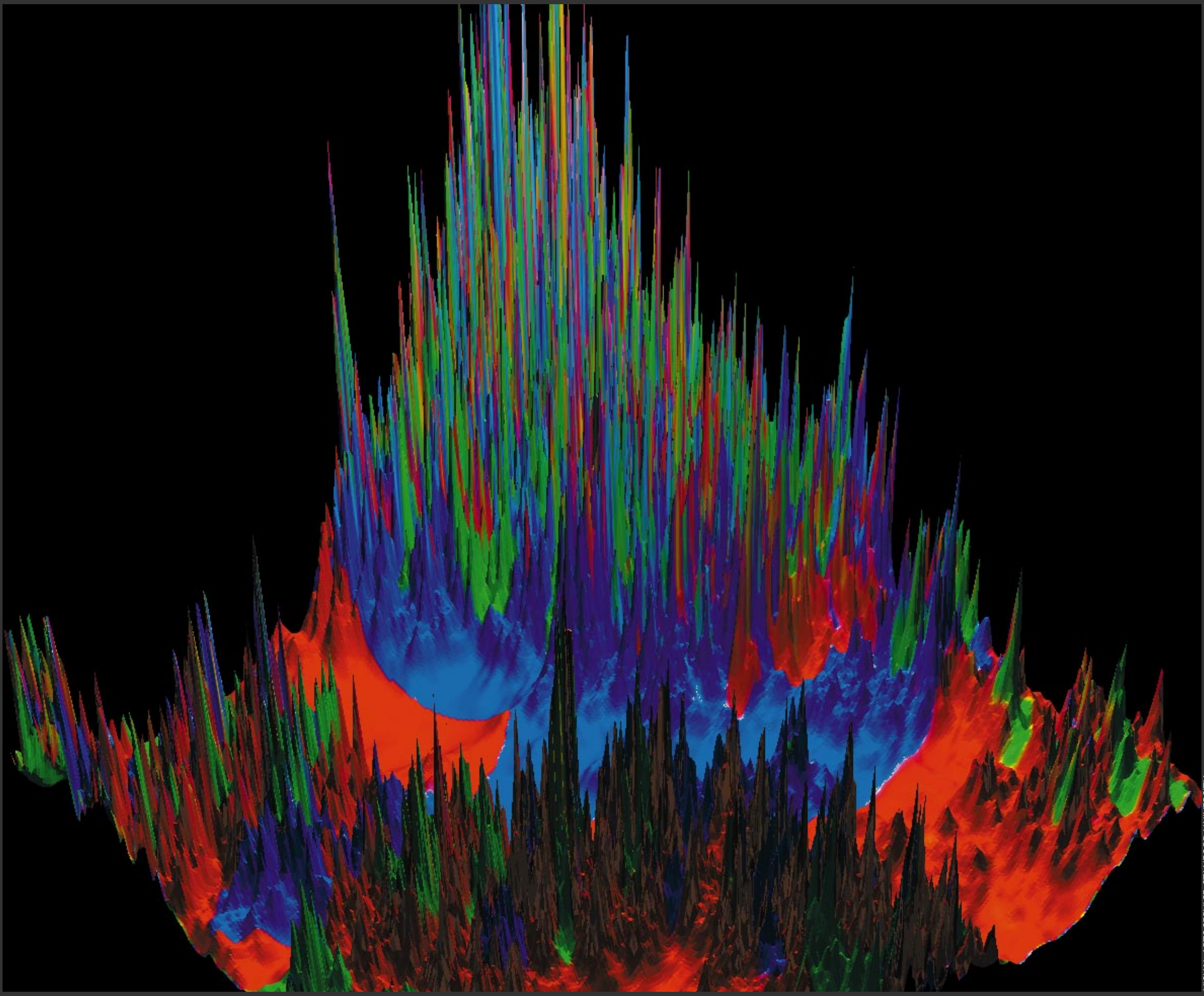
Promising space missions loom on the horizon, such as the National Aeronautics and Space Administration's Microwave Anisotropy Probe (MAP) and the European Space Agency's Planck Surveyor. These two experiments will permit detailed mapping of small fluctuations in the cosmic microwave background. If current cosmological theories prove correct, these measurements will robustly determine the density of matter in the universe and independently constrain the Hubble constant.

Although the history of science suggests that ours is not the last generation to wrestle with these questions, the next decade promises much excitement. There are many reasons to be optimistic that the current disagreement over values of the cosmological parameters governing the evolution of the universe will soon be resolved.

The Author

WENDY L. FREEDMAN is a staff member at the Carnegie Institution's observatories in Pasadena, Calif. Born in Toronto, she received a Ph.D. in astronomy and astrophysics from the University of Toronto in 1984 and, in 1987, became the first woman to join Carnegie's scientific staff. In 1994 she received the Marc Aaronson Prize for her contributions to the study of extragalactic distance and stellar populations of galaxies. A co-leader of the Hubble Space Telescope Key Project to measure the Hubble constant, she is also a member of the National Research Council's Committee on Astronomy and Astrophysics, the executive board of the Center for Particle Astrophysics and the National Aeronautics and Space Administration's scientific oversight committee planning the Next Generation Space Telescope. Beyond astronomy, her main interest is her family: husband Barry, daughter Rachael and son Daniel. This article updates a version that appeared in *Scientific American* in November 1992.

The Self-Reproducing Inflationary Universe



VISUAL ARTS SERVICE, STANFORD UNIVERSITY

SELF-REPRODUCING UNIVERSE
in a computer simulation consists of exponentially large domains, each of which has different laws of physics (*represented by colors*). Sharp peaks are new "big bangs"; their heights correspond to the energy density of the universe there. At the top of the peaks, the colors rapidly fluctuate, indicating that the laws of physics there are not yet settled. They become fixed only in the valleys, one of which corresponds to the kind of universe we live in now.

by Andrei Linde

If my colleagues and I are right, we may soon be saying good-bye to the idea that our universe was a single fireball created in the big bang. We are exploring a new theory based on a 15-year-old notion that the universe went through a stage of inflation. During that time, the theory holds, the cosmos became exponentially large within an infinitesimal fraction of a second. At the end of this period, the universe continued its evolution according to the big bang model. As workers refined this inflationary scenario, they uncovered some surprising consequences. One of them constitutes a fundamental change in how the cosmos is seen. Recent versions of inflationary theory assert that instead of being an expanding ball of fire the universe is a huge, growing fractal. It consists of many inflating balls that produce new balls, which in turn produce more balls, ad infinitum.

Recent versions of the inflationary scenario describe the universe as a self-generating fractal that sprouts other inflationary universes

Cosmologists did not arbitrarily invent this rather peculiar vision of the universe. Several workers, first in Russia and later in the U.S., proposed the inflationary hypothesis that is the basis of its foundation. We did so to solve some of the complications left by the old big bang idea. In its standard form, the big bang theory maintains that the universe was born about 15 billion years ago from a cosmological singularity—a state in which the temperature and density are infinitely high. Of course, one cannot really speak in physical terms about these quantities as being infinite. One usually assumes that the current laws of physics did not apply then. They took hold only after the density of the universe dropped below the so-called Planck density, which equals about 10^{94} grams per cubic centimeter.

As the universe expanded, it gradually cooled. Remnants of the primordial cosmic fire still surround us in the form of the microwave background radiation. This radiation indicates that the temperature of the universe has dropped to 2.7 kelvins. The 1965 discovery of this background radiation by Arno A. Penzias and Robert W. Wilson of Bell Laboratories proved to be the crucial evidence in establishing the big bang theory as the preeminent theory of cosmology. The big bang theory also explained the abundances of hydrogen, helium and other elements in the universe.

As investigators developed the theory, they uncovered complicated problems. For example, the standard big bang theory, coupled with the modern theory of elementary particles, predicts the existence of many superheavy particles carrying magnetic charge—that is, objects that have only one magnetic pole. These magnetic monopoles would have a typical mass 10^{16} times that of the proton, or about 0.00001 milligram. According to the standard big bang theory, monopoles should have emerged very early in the evolution of the universe and should now be as abundant as protons. In that case, the mean density of matter in the universe would be about 15 orders of magnitude greater than its present value, which is about 10^{-29} gram per cubic centimeter.

Questioning Standard Theory

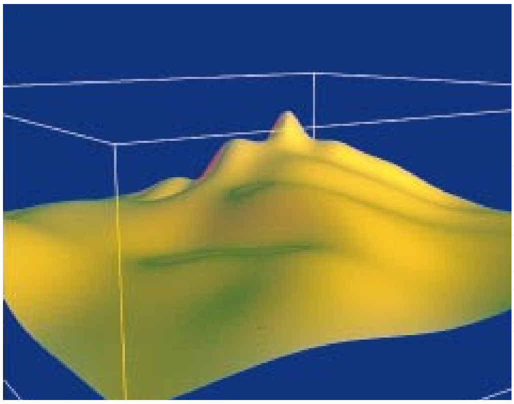
This and other puzzles forced physicists to look more attentively at the basic assumptions underlying the standard cosmological theory. And we found many to be highly suspicious. I will review six of the most difficult. The first, and main, problem is the very existence of the big bang. One may wonder, What came before? If space-time did not exist then, how could everything appear from nothing? What arose first: the universe or the laws determining its evolution? Explaining this initial singularity—where and when it all began—still remains the most intractable problem of modern cosmology.

A second trouble spot is the flatness of space. General relativity suggests that space may be very curved, with a typical radius on the order of the Planck length, or 10^{-33} centimeter. We see, however, that our universe is just about flat on a scale of 10^{28} centimeters, the radius of the observable part of the universe. This result of our observation differs from theoretical expectations by more than 60 orders of magnitude.

A similar discrepancy between theory and observations concerns the size of the universe, a third problem. Cosmological examinations show that our part of the universe contains at least 10^{88} elementary particles. But why is the universe so big? If one takes a universe of a typical initial size given by the Planck length and a typical initial density equal to the Planck density, then, using the standard big bang theory, one can calculate how many elementary particles such a universe might encompass. The answer is rather unexpected: the entire universe should only be large enough to accommodate just one elementary particle—or at most 10 of them. It would be unable to house even a single reader of *Scientific American*, who consists of about 10^{29} elementary particles. Obviously, something is wrong with this theory.

The fourth problem deals with the timing of the expansion. In its standard form, the big bang theory assumes that all parts of the universe began expanding simultaneously. But how could all the different parts of the universe synchronize the beginning of their expansion? Who gave the command?

Fifth, there is the question about the distribution of matter in the universe. On the very large scale, matter has spread out with remarkable uniformity. Across more than 10 billion light-years, its distribution departs from perfect homogeneity by less than one part in 10,000. For a long time, nobody had any idea why the universe was so homogeneous. But those who do not have ideas sometimes have principles. One of the cornerstones of the standard cosmology was the "cosmological principle," which asserts that the universe must be homogeneous. This assumption, however, does not help much, because the universe incorporates important deviations from homogeneity, namely, stars, galaxies and other agglomerations of matter. Hence, we must explain why the universe is so uniform on large scales and at the same time suggest some mechanism that produces galaxies.


VISUAL ARTS SERVICE, STANFORD UNIVERSITY

EVOLUTION OF A SCALAR FIELD leads to many inflationary domains. In most parts of the universe, the scalar field decreases (*represented as depressions and valleys*). In other places, quantum fluctuations cause the scalar field to grow.

Finally, there is what I call the uniqueness problem. Albert Einstein captured its essence when he said, "What really interests me is whether God had any choice in the creation of the world." Indeed, slight changes in the physical constants of nature could have made the universe unfold in a completely different manner. For example, many popular theories of elementary particles assume that space-time originally had considerably more than four dimensions (three spatial and one temporal). In order to square theoretical calculations with the physical world in which we live, these models state that the extra dimensions have been "compactified," or shrunk to a small size and tucked away. But one may wonder why compactification stopped with four dimensions, not two or five.

Moreover, the manner in which the other dimensions become rolled up is significant, for it determines the values of the constants of nature and the masses of particles. In some theories, compactification can occur in billions of different ways. A few years ago it would have seemed rather meaningless to ask why space-time has four dimensions, why the gravitational constant is so small or why the proton is almost 2,000 times heavier than the electron. Now developments in elementary particle physics make answering these questions crucial to understanding the construction of our world.

All these problems (and others I have not mentioned) are extremely perplexing. That is why it is encouraging that many of these puzzles can be resolved in the context of the theory of the self-reproducing, inflationary universe.

The basic features of the inflationary scenario are rooted in the physics of elementary particles. So I would like to take you on a brief excursion into this realm—in particular, to the unified theory of weak and electromagnetic interactions. Both these forces exert themselves through particles. Photons mediate the electromagnetic force; the *W* and *Z* particles are responsible for the weak force. But whereas photons are massless, the *W* and *Z* particles are extremely heavy. To unify the weak and electromagnetic interactions despite the obvious differences between photons and the *W* and *Z* particles, physicists introduced what are called scalar fields.

Although scalar fields are not the stuff of everyday life, a familiar analogue exists. That is the electrostatic potential—the voltage in a circuit is an example. Electrical fields appear only if this potential is uneven, as it is between the poles of a battery or if the potential changes in time. If the entire universe had the same electrostatic potential—say, 110 volts—then nobody would notice it; the potential would seem to be just another vacuum state. Similarly, a constant scalar field looks like a vacuum: we do not see it even if we are surrounded by it.

These scalar fields fill the universe and mark their presence by affecting properties of elementary particles. If a scalar field interacts with the *W* and *Z* particles, they become heavy. Particles that do not interact with the scalar field, such as photons, remain light.

To describe elementary particle physics, therefore, physicists begin with a theory in which all particles initially are light and in which no fundamental difference between weak and electromagnetic interactions exists. This difference arises only later, when the universe expands and becomes filled by various scalar fields. The process by which the fundamental forces separate is called symmetry breaking. The particular value of the scalar field that appears in the universe is determined by the position of the minimum of its potential energy.

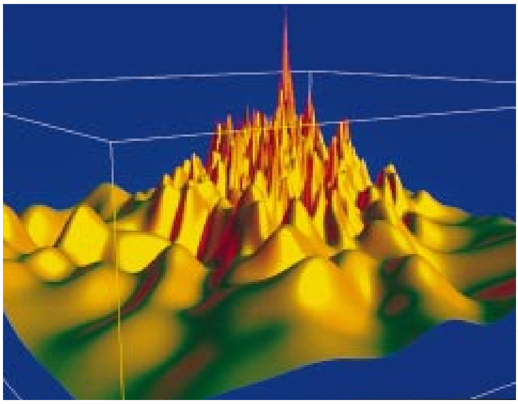
Scalar Fields

Scalar fields play a crucial role in cosmology as well as in particle physics. They provide the mechanism that generates the rapid inflation of the universe. Indeed, according to general relativity, the universe expands at a rate (approximately) proportional to the square root of its density. If the universe were filled by ordinary matter, then the density would rapidly decrease as the universe expanded. Thus, the expansion of the universe would rapidly slow down as density decreased. But because of the equivalence of mass and energy established by Einstein, the potential energy of the scalar field also contributes to the expansion. In certain cases, this energy decreases much more slowly than does the density of ordinary matter.

The persistence of this energy may lead to a stage of extremely rapid expansion, or inflation, of the universe. This possibility emerges even if one considers the very simplest version of the theory of a scalar field. In this version the potential energy reaches a minimum at the point where the scalar field vanishes. In this case, the larger the scalar field, the greater the potential energy. According to Einstein's theory of gravity, the energy of the scalar field must have caused the universe to expand very rapidly. The expansion slowed down when the scalar field reached the minimum of its potential energy.

One way to imagine the situation is to picture a ball rolling down the side of a large bowl. The bottom of the bowl represents the energy minimum. The position of the ball corresponds to the value of the scalar field. Of course, the equations describing the motion of the scalar field in an expanding universe are somewhat more complicated than the equations for the ball in an empty bowl. They contain an extra term corresponding to friction, or viscosity. This friction is akin to having molasses in the bowl. The viscosity of this liquid depends on the energy of the field: the higher the ball in the bowl is, the thicker the liquid will be. Therefore, if the field initially was very large, the energy dropped extremely slowly.

The sluggishness of the energy drop in the scalar field has a crucial implication in the expansion rate. The decline was so gradual that the potential energy of the scalar field remained almost constant as the universe expanded. This behavior contrasts sharply with that of ordinary matter, whose density rapidly decreases in an expanding universe. Thanks to the large energy of the scalar field, the universe continued to expand at a speed much greater than that predicted by preinflation cosmological theories. The size of the universe in this regime grew exponentially.



VISUAL ARTS SERVICE, STANFORD UNIVERSITY

UNIVERSE EXPANDS RAPIDLY in places—represented in the above model as peaks—where quantum fluctuations cause the scalar field to grow. Such expansion creates inflationary regions. In this model, we would exist in a valley, where space is no longer inflating.

This stage of self-sustained, exponentially rapid inflation did not last long. Its duration could have been as short as 10^{-35} second. Once the energy of the field declined, the viscosity nearly disappeared, and inflation ended. Like the ball as it reaches the bottom of the bowl, the scalar field began to oscillate near the minimum of its potential energy. As the scalar field oscillated, it lost energy, giving it up in the form of elementary particles. These particles interacted with one another and eventually settled down to some equilibrium temperature. From this time on, the standard big bang theory can describe the evolution of the universe.

The main difference between inflationary theory and the old cosmology becomes clear when one calculates the size of the universe at the end of inflation. Even if the universe at the beginning of inflation was as small as 10^{-33} centimeter, after 10^{-35} second of inflation this domain acquires an unbelievable size. According to some inflationary models, this size in centimeters can equal $10^{10^{12}}$—that is, a 1 followed by a trillion zeros. These numbers depend on the models used, but in most versions, this size is many orders of magnitude greater than the size of the observable universe, or 10^{28} centimeters.

This tremendous spurt immediately solves most of the problems of the old cosmological theory. Our universe appears smooth and uniform because all inhomogeneities were stretched $10^{10^{12}}$ times. The density of primordial monopoles and other undesirable "defects" becomes exponentially diluted. (Recently we have found that monopoles may inflate themselves and thus effectively push themselves out of the observable universe.) The universe has become so large that we can now see just a tiny fraction of it. That is why, just like a small area on a surface of a huge inflated balloon, our part looks flat. That is why we do not need to insist that all parts of the universe began expanding simultaneously. One domain of a smallest possible size of 10^{-33} centimeter is more than enough to produce everything we see now.

An Inflationary Universe

Inflationary theory did not always look so conceptually simple. Attempts to obtain the stage of exponential expansion of the universe have a long history. Unfortunately, because of political barriers, this history is only partially known to American readers.

The first realistic version of the inflationary theory came in 1979 from Alexei A. Starobinsky of the L. D. Landau Institute of Theoretical Physics in Moscow. The Starobinsky model created a sensation among Russian astrophysicists, and for two years it remained the main topic of discussion at all conferences on cosmology in the Soviet Union. His model, however, was rather complicated (it was based on the theory of anomalies in quantum gravity) and did not say much about how inflation could actually start.

In 1981 Alan H. Guth of the Massachusetts Institute of Technology suggested that the hot universe at some intermediate stage could expand exponentially. His model derived from a theory that interpreted the development of the early universe as a series of phase transitions. This theory was proposed in 1972 by David A. Kirzhnits and me at the P. N. Lebedev Physics Institute in Moscow. According to this idea, as the universe expanded and cooled, it condensed into different forms. Water vapor undergoes such phase transitions. As it becomes cooler, the vapor condenses into water, which, if cooling continues, becomes ice.

Guth's idea called for inflation to occur when the universe was in an unstable, supercooled state. Supercooling is common during phase transitions; for example, water under the right circumstances remains liquid below zero degrees Celsius. Of course, supercooled water eventually freezes. That event would correspond to the end of the inflationary period. The idea to use supercooling for solving many problems of the big bang theory was very attractive. Unfortunately, as Guth himself pointed out, the postinflation universe of his scenario becomes extremely inhomogeneous. After investigating his model for a year, he finally renounced it in a paper he co-authored with Erick J. Weinberg of Columbia University.

In 1982 I introduced the so-called new inflationary universe scenario, which Andreas Albrecht and Paul J. Steinhardt of the University of Pennsylvania also later discovered [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt; SCIENTIFIC AMERICAN, May 1984]. This scenario shrugged off the main problems of Guth's model. But it was still rather complicated and not very realistic.

Only a year later did I realize that inflation is a naturally emerging feature in many theories of elementary particles, including the simplest model of the scalar field discussed earlier. There is no need for quantum gravity effects, phase transitions, supercooling or even the standard assumption that the universe originally was hot. One just considers all possible kinds and values of scalar fields in the early universe and then checks to see if any of them leads to inflation. Those places where inflation does not occur remain small. Those domains where inflation takes place become exponentially large and dominate the total volume of the universe. Because the scalar fields can take arbitrary values in the early universe, I called this scenario chaotic inflation.

In many ways, chaotic inflation is so simple that it is hard to understand why the idea was not discovered sooner. I think the reason was purely psychological. The glorious successes of the big bang theory hypnotized cosmologists. We assumed that the entire universe was created at the same moment, that initially it was hot and that the scalar field from the beginning resided close to the minimum of its potential energy. Once we began relaxing these assumptions, we immediately found that inflation is not an exotic phenomenon invoked by theorists for solving their problems. It is a general regime that occurs in a wide class of theories of elementary particles.

That a rapid stretching of the universe can simultaneously resolve many difficult cosmological problems may seem too good to be true. Indeed, if all inhomogeneities were stretched away, how did galaxies form? The answer is that while removing previously existing inhomogeneities, inflation at the same time made new ones.

These inhomogeneities arise from quantum effects. According to quantum mechanics, empty space is not entirely empty. The vacuum is filled with small quantum fluctuations. These fluctuations can be regarded as waves, or undulations in physical fields. The waves have all possible wavelengths and move in all directions. We cannot detect these waves, because they live only briefly and are microscopic.

In the inflationary universe the vacuum structure becomes even more complicated. Inflation rapidly stretches the waves. Once their wavelengths become sufficiently large, the undulations begin to “feel” the curvature of the universe. At this moment, they stop moving because of the viscosity of the scalar field (recall that the equations describing the field contain a friction term).

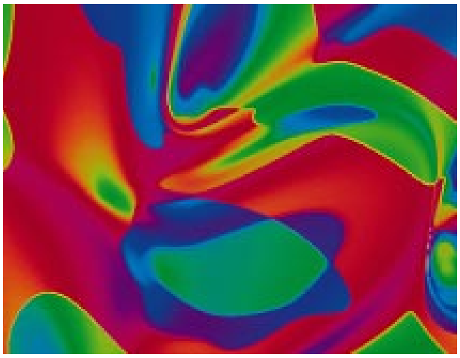
The first fluctuations to freeze are those that have large wavelengths. As the universe continues to expand, new fluctuations become stretched and freeze on top of other frozen waves. At this stage one cannot call these waves quantum fluctuations anymore. Most of them have extremely large wavelengths. Because these waves do not move and do not disappear, they enhance the value of the scalar field in some areas and depress it in others, thus creating inhomogeneities. These disturbances in the scalar field cause the density perturbations in the universe that are crucial for the subsequent formation of galaxies.

Testing Inflationary Theory

In addition to explaining many features of our world, inflationary theory makes several important and testable predictions. First, density perturbations produced during inflation affect the distribution of matter in the universe. They may also accompany gravitational waves. Both density perturbations and gravitational waves make their imprint on the microwave background radiation. They render the temperature of this radiation slightly different in various places in the sky. This nonuniformity was found in 1992 by the Cosmic Background Explorer (COBE) satellite, a finding later confirmed by several other experiments.

Although the COBE results agree with the predictions of inflation, it would be premature to claim that COBE has confirmed inflationary theory. But it is certainly true that the results obtained by the satellite at their current level of precision could have definitively disproved most inflationary models, and it did not happen. At present, no other theory can simultaneously explain why the universe is so homogeneous and still predict the “ripples in space” discovered by COBE.

Inflation also predicts that the universe should be nearly flat.



VISUAL ARTS SERVICE, STANFORD UNIVERSITY

KANDINSKY UNIVERSE, named after the Russian abstractionist painter, is depicted here as a swirling pattern that represents an energy distribution in the theory of axions, a kind of scalar field.

Flatness of the universe can be experimentally verified because the density of a flat universe is related in a simple way to the speed of its expansion. So far observational data are consistent with this prediction. A few years ago it seemed that if someone were to show that the universe is open rather than flat, then inflationary theory would fall apart. Recently, however, several models of an open inflationary universe have been found. The only consistent description of a large homogeneous open universe that we currently know is based on inflationary theory. Thus, even if the universe is open, inflation is still the best theory to describe it. One may argue that the only way to disprove the theory of inflation is to propose a better theory.

One should remember that inflationary models are based on the theory of elementary particles, and this theory is not completely established. Some versions (most notably, superstring theory) do not automatically lead to inflation. Pulling inflation out of the superstring model may require radically new ideas. We should certainly continue the search for alternative cosmological theories. Many cosmologists, however, believe inflation, or something very similar to it, is absolutely essential for constructing a consistent cosmological theory. The inflationary theory itself changes as particle physics theory rapidly evolves. The list of new models includes extended inflation, natural inflation, hybrid inflation and many others. Each model has unique features that can be tested through observation or experiment. Most, however, are based on the idea of chaotic inflation.

Here we come to the most interesting part of our story, to the theory of an eternally existing, self-reproducing inflationary universe. This theory is rather general, but it looks especially promising and leads to the most dramatic consequences in the context of the chaotic inflation scenario.

As I already mentioned, one can visualize quantum fluctuations of the scalar field in an inflationary universe as waves. They first moved in all possible directions and then froze on top of one another. Each frozen wave slightly increased the scalar field in some parts of the universe and decreased it in others.

Now consider those places of the universe where these newly frozen waves persistently increased the scalar field. Such regions are extremely rare, but still they do exist. And they can be extremely important. Those rare domains of the universe where the field jumps high enough begin exponentially expanding with ever increasing speed. The higher the scalar field jumps, the faster the universe expands. Very soon those rare domains will acquire a much greater volume than other domains.

From this theory it follows that if the universe contains at least one inflationary domain of a sufficiently large size, it begins unceasingly producing new inflationary domains. Inflation in each particular point may end quickly, but many other places will continue to expand. The total volume of all these domains will grow without end. In essence, one inflationary universe sprouts other inflationary bubbles, which in turn produce other inflationary bubbles.

This process, which I have called eternal inflation, keeps going as a chain reaction, producing a fractallike pattern of universes. In this scenario the universe as a whole is immortal. Each particular part of the universe may stem from a singularity somewhere in the past, and it may end up in a singularity somewhere in the future. There is, however, no end for the evolution of the entire universe.

The situation with the very beginning is less certain. There is a chance that all parts of the universe were created simultaneously in an initial big bang singularity. The necessity of this assumption, however, is no longer obvious.

Furthermore, the total number of inflationary bubbles on our "cosmic tree" grows exponentially in time. Therefore, most bubbles (including our own part of the universe) grow indefinitely far away from the trunk of this tree. Although this scenario makes the existence of the initial big bang almost irrelevant, for all practical purposes, one can consider the moment of formation of each inflationary bubble as a new "big bang." From this perspective, inflation is not a part of the big bang theory, as we thought 15 years ago. On the contrary, the big bang is a part of the inflationary model.

In thinking about the process of self-reproduction of the universe, one cannot avoid drawing analogies, however superficial they may be. One may wonder, Is not this process similar to what happens with all of us? Some time ago we were born. Eventually we will die, and the entire world of our thoughts, feelings and memories will disappear. But there were those who lived before us, there will be those who will live after, and humanity as a whole, if it is clever enough, may live for a long time.

Inflationary theory suggests that a similar process may occur with the universe. One can draw some optimism from knowing that even if our civilization dies, there will be other places in the universe where life will emerge again and again, in all its possible forms.

A New Cosmology

Could matters become even more curious? The answer is yes. Until now, we have considered the simplest inflationary model with only one scalar field, which has only one minimum of its potential energy. Meanwhile realistic models of elementary particles propound many kinds of scalar fields. For example, in the unified theories of weak, strong and electromagnetic interactions, at least two other scalar fields exist. The potential energy of these scalar fields may have several different minima. This condition means that the same theory may have different "vacuum states," corresponding to different types of symmetry breaking between fundamental interactions and, as a result, to different laws of low-energy physics. (Interactions of particles at extremely large energies do not depend on symmetry breaking.)

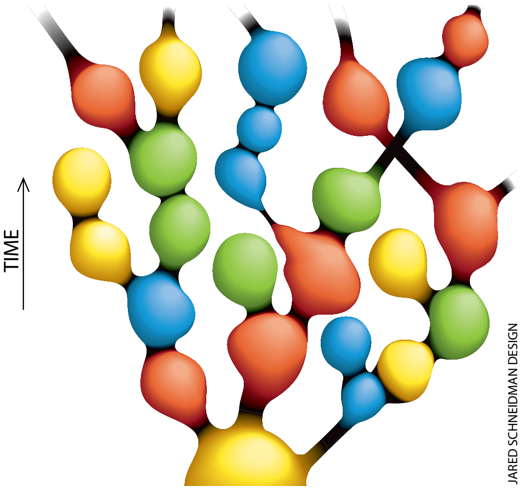
Such complexities in the scalar field mean that after inflation the universe may become divided into exponentially large domains that have different laws of low-energy physics. Note that this division occurs even if the entire universe originally began in the same state, corresponding to one particular minimum of potential energy. Indeed, large quantum fluctuations can cause scalar fields to jump out of their minima. That is, they jiggle some of the balls out of their bowls and into other ones. Each bowl corresponds to alternative laws of particle interactions. In some inflationary models, quantum fluctuations are so strong that even the number of dimensions of space and time can change.

If this model is correct, then physics alone cannot provide a complete explanation for all properties of our allotment of the universe. The same physical theory may yield large parts of the universe that have diverse properties. According to this scenario, we find ourselves inside a four-dimensional domain with our kind of physical laws, not because domains with different dimensionality and with alternative properties are impossible or improbable but simply because our kind of life cannot exist in other domains.

Does this mean that understanding all the properties of our region of the universe will require, besides a knowledge of physics, a deep investigation of our own nature, perhaps even including the nature of our consciousness? This conclusion would certainly be one of the most unexpected that one could draw from the recent developments in inflationary cosmology.

The evolution of inflationary theory has given rise to a completely new cosmological paradigm, which differs considerably from the old big bang theory and even from the first versions of the inflationary scenario. In it the universe appears to be both chaotic and homogeneous, expanding and stationary. Our cosmic home grows, fluctuates and eternally reproduces itself in all possible forms, as if adjusting itself for all possible types of life.

Some parts of the new theory, we hope, will stay with us for years to come. Many others will have to be considerably modified to fit with new observational data and with the ever changing theory of elementary particles. It seems, however, that the past 15 years of development of cosmology have irreversibly changed our understanding of the structure and fate of our universe and of our own place in it. SA



JARED SCHNEIDMAN DESIGN

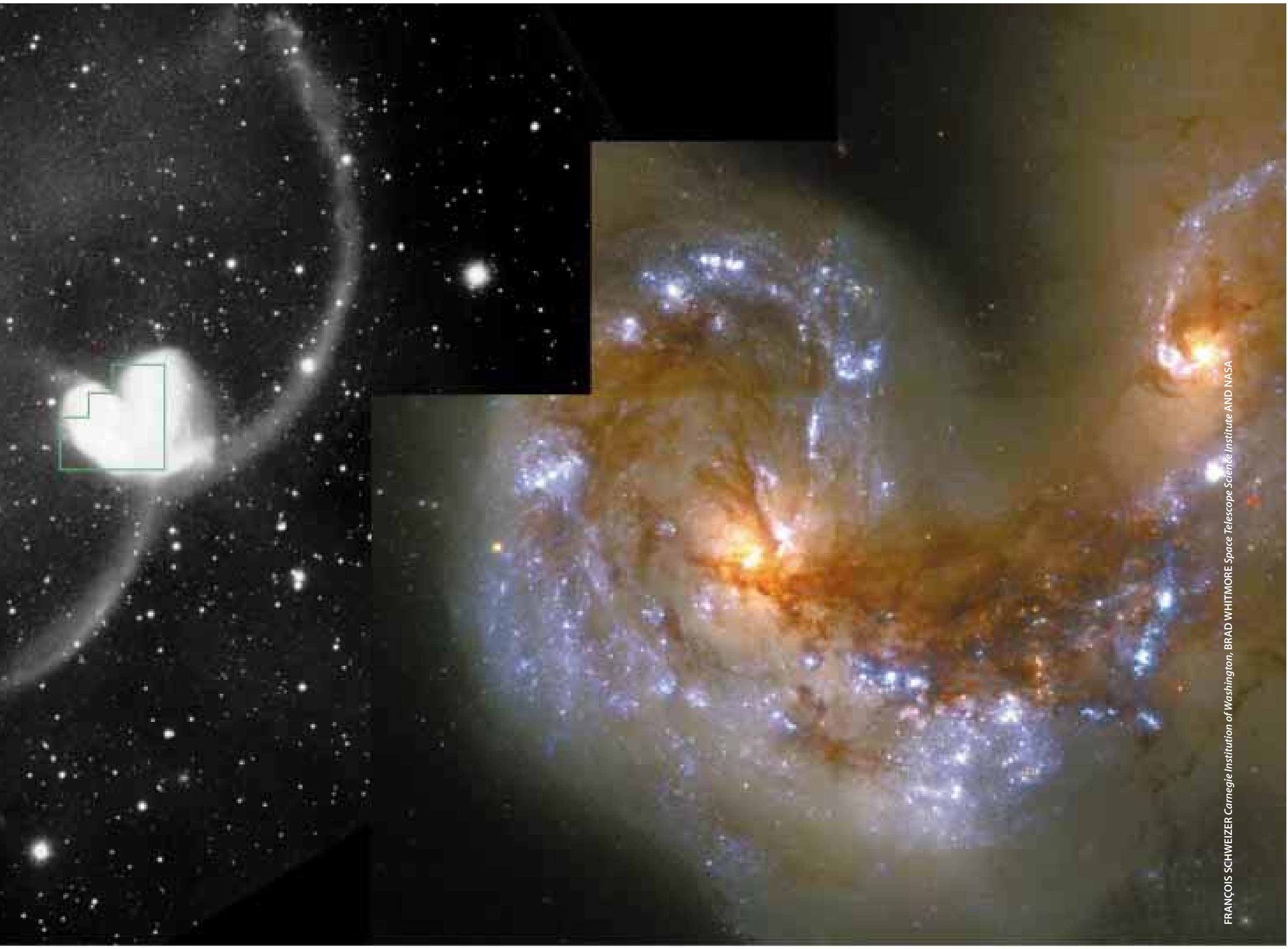
SELF-REPRODUCING COSMOS appears as an extended branching of inflationary bubbles. Changes in color represent "mutations" in the laws of physics from parent universes. The properties of space in each bubble do not depend on the time when the bubble formed. In this sense, the universe as a whole may be stationary, even though the interior of each bubble can be described by the big bang theory.

The Author

ANDREI LINDE is one of the originators of inflationary theory. After graduating from Moscow University, he received his Ph.D. at the P. N. Lebedev Physics Institute in Moscow, where he began probing the connections between particle physics and cosmology. He became a professor of physics at Stanford University in 1990. He lives in California with his wife, Renata Kallosh (also a professor of physics at Stanford), and his sons, Dmitri and Alex. A detailed description of inflationary theory is given in his book *Particle Physics and Inflationary Cosmology* (Harwood Academic Publishers, 1990). This article updates a version that appeared in *Scientific American* in November 1994.

Dark Matter in the Universe

by Vera Rubin



FRANÇOIS SCHWEIZER *Carnegie Institution of Washington*, BRAD WHITMORE *Space Telescope Science Institute* AND NASA

As much as 90 percent of the matter in the universe is invisible. Detecting this dark matter will help astronomers better comprehend the universe's destiny

Imagine, for a moment, that one night you awaken abruptly from a dream. Coming to consciousness, blinking your eyes against the blackness, you find that, inexplicably, you are standing alone in a vast, pitch-black cavern. Befuddled by this predicament, you wonder: Where am I? What is this space? What are its dimensions?

Groping in the darkness, you stumble upon a book of damp matches. You strike one; it quickly flares, then fizzles out. Again, you try; again, a flash and fizzle. But in that moment, you realize that you can glimpse a bit of your surroundings. The next match strike lets you sense faint walls far away. Another flare reveals a strange shadow, suggesting the presence of a big object. Yet another suggests you are moving—or, instead, the room is moving relative to you. With each momentary flare, a bit more is learned.

In some sense, this situation recalls our puzzling predicament on Earth. Today, as we have done for centuries, we gaze into the night sky from our planetary platform and wonder where we are in this cavernous cosmos. Flecks of light provide some clues about great objects in space. And what we do discern about their motions and apparent shadows tells us that there is much more that we cannot yet see.

From every photon we collect from the universe's farthest reaches, we struggle to extract information. Astronomy is the study of light that reaches Earth from the heavens. Our task is not only to collect as much light as possible—from ground- and space-based telescopes—but also to use what we can see in the heavens to understand better what we cannot see and yet know must be there.

Based on 50 years of accumulated observations of the motions of galaxies and the expansion of the universe, most astronomers believe that as much as 90 percent of the stuff constituting the universe may be objects or particles that cannot be seen. In other words, most of the universe's matter does not radiate—it provides no glow that we can detect in the electromagnetic spectrum. First posited some 60 years ago by astronomer Fritz Zwicky, this so-called missing matter was believed to reside within clusters of galaxies. Nowadays we prefer to call the missing mass "dark matter," for it is the light, not the matter, that is missing.

Astronomers and physicists offer a variety of explanations for this dark matter. On the one hand, it could merely be ordinary material, such as ultrafaint stars, large or small black holes, cold gas, or dust scattered around the universe—all of which emit or reflect too little radiation for our instruments to detect. It could even be a category of dark objects called MACHOs (MAssive Compact Halo Objects) that lurk invisibly in the halos surrounding galaxies and galactic clusters. On the other hand, dark matter could consist of exotic, unfamiliar particles that we have not figured out how to observe. Physicists theorize about the existence of these particles, although experiments have not yet confirmed their presence. A third possibility is that our understanding of gravity needs a major revision—but most physicists do not consider that option seriously.

GALAXIES COLLIDE IN THIS MERGER OF THE ANTENNAE GALAXIES, known as NGC 4038/39, setting off stellar fireworks and yielding more than 1,000 bright new star clusters. The galaxies garnered their name from their long, luminous tails—formed by gravitational tidal forces of their encounter—which resemble insect antennae. Visible in the Southern Hemisphere, the galaxies are 63 million light-years from Earth. This figure (*opposite page*) blends two images: the wide angle (*left*) comes from a ground-based telescope in Chile, whereas the detailed view (*right*) comes from the Hubble Space Telescope.

In some sense, our ignorance about dark matter's properties has become inextricably tangled up with other outstanding issues in cosmology—such as how much mass the universe contains, how galaxies formed and whether or not the universe will expand forever. So important is this dark matter to our understanding of the size, shape and ultimate fate of the universe that the search for it will very likely dominate astronomy for the next few decades.

Observing the Invisible

Understanding something you cannot see is difficult—but not impossible. Not surprisingly, astronomers currently study dark matter by its effects on the bright matter that we do observe. For instance, when we watch a nearby star wobbling predictably, we infer from calculations that a "dark planet" orbits around it. Applying similar principles to spiral galaxies, we infer dark matter's presence because it accounts for the otherwise inexplicable motions of stars within those galaxies.

When we observe the orbits of stars and clouds of gas as they circle the centers of spiral galaxies, we find that they move too quickly. These unexpectedly high velocities signal the gravitational tug exerted by something more than that galaxy's visible matter. From detailed velocity measurements, we conclude that large amounts of invisible matter exert the gravitational force that is holding these stars and gas clouds in high-speed orbits. We deduce that dark matter is spread out around the galaxy, reaching beyond the visible galactic edge and bulging above and below the otherwise flattened, luminous galactic disk. As a rough approximation, try to envision a typical spiral galaxy, such as our Milky Way, as a relatively flat, glowing disk embedded in a spherical halo of invisible material—almost like an extremely diffuse cloud.

Looking at a single galaxy, astronomers see within the galaxy's radius (a distance of about 50,000 light-years) only about one tenth of the total gravitating mass needed to account for how fast individual stars are rotating around the galactic hub.

In trying to discover the amount and distribution of dark matter in a cluster of galaxies, x-ray astronomers have found that galaxies within clusters float immersed in highly diffuse clouds of 100-million-degree gas—gas that is rich in energy yet difficult to detect. Observers have learned to use the x-ray-emitting gas's temperature and extent in much the same way that optical astronomers use the velocities of stars in a single galaxy. In both cases, the data provide clues to the nature and location of the unseen matter.

In a cluster of galaxies, the extent of the x-ray-emitting region and temperature of the gas enable us to estimate the amount of gravitating mass within the cluster's radius, which measures almost 100 million light-years. In a typical case, when we add together the luminous matter and the x-ray-emitting hot gas, we are able to sense roughly 20 to 30 percent of the cluster's total gravitating mass. The remainder, which is dark matter, remains undetected by present instruments.

Subtler ways to detect invisible matter have recently emerged. One clever method involves spotting rings or arcs around clusters of galaxies. These "Einstein rings" arise from an effect known as gravitational lensing, which occurs when gravity from a massive object bends light passing by. For instance,

LARGE MAGELLANIC CLOUD, one of the Milky Way's nearest satellite galaxies, is located 180,000 light-years from Earth. Like other small satellite galaxies, the cloud will ultimately merge with the Milky Way, thus becoming one of the galaxy's building blocks. As we view the cloud from Earth, dark objects in the Milky Way's halo gravitationally lens some stars in the cloud, thus providing information about the presence of dark matter in our galaxy's halo.

©ANGLO-AUSTRALIAN OBSERVATORY; PHOTOGRAPHY BY DAVID F. MALIN

when a cluster of galaxies blocks our view of another galaxy behind it, the cluster's gravity warps the more distant galaxy's light, creating rings or arcs, depending on the geometry involved. Interestingly, the nearer cluster acts as nature's telescope, bending light into our detectors—light that would otherwise have traveled elsewhere in the universe. Someday we may exploit these natural telescopes to view the universe's most distant objects.

Using computer models, we can calculate the mass of the intervening cluster, estimating the amount of invisible matter that must be present to produce the observed geometric deflection. Such calculations confirm that clusters contain far more mass than the luminous matter suggests.

Even compact dark objects in our own galaxy can gravitationally lens light. When a foreground object eclipses a background star, the light from the background star is distorted into a tiny ring, whose brightness far exceeds the star's usual brightness. Consequently, we observe an increase, then a decrease, in the background star's brightness. Careful analysis of the light's variations can tease out the mass of the dark foreground lensing object.

Where Is Dark Matter?

Several teams search nightly for nearby lensing events, caused by invisible MACHOs in our own Milky Way's halo. The search for them covers millions of stars in the Magellanic Clouds and the Andromeda galaxy. Ultimately, the search will limit the amount of dark matter present in our galaxy's halo.

Given the strong evidence that spiral and elliptical galaxies lie embedded in large dark-matter halos, astronomers now wonder about the location, amount and distribution of the invisible material.

To answer those questions, researchers compare and contrast observations from specific nearby galaxies. For instance, we learn from the motions of the Magellanic Clouds, two satellite galaxies gloriously visible in the Southern Hemisphere, that they orbit within the Milky Way galaxy's halo and that the halo continues beyond the clouds, spanning a distance of almost 300,000 light-years. In fact, motions of our galaxy's most distant satellite objects suggest that its halo may extend twice as far—to 600,000 light-years.

Because our nearest neighboring spiral galaxy, Andromeda, lies a mere two million light-years away, we now realize that our galaxy's halo may indeed span a significant fraction of the distance to Andromeda and its halo. We have also determined that clusters of galaxies lie embedded in even larger systems of dark matter. At the farthest distances for which we can deduce the masses of galaxies, dark matter appears to dwarf luminous matter by a factor of at least 10, possibly as much as 100.

Overall, we believe dark matter associates loosely with bright matter, because the two often appear together. Yet, admittedly, this conclusion may stem from biased observations, because bright matter typically enables us to find dark matter.

By meticulously studying the shapes and motions of galaxies over decades, astronomers have realized that individual galaxies are actively evolving, largely because of the mutual gravitational pull of galactic neighbors. Within individual galaxies, stars remain enormously far apart relative to their diameters, thus little affecting one another gravitationally. For example, the separation between the sun and its nearest neighbor, Proxima Centauri, is so great that 30 million suns could fit between the two. In contrast, galaxies lie close together, relative to their diameters—nearly all have neighbors within a few diameters. So galaxies do alter one another gravitationally, with dark matter's added gravity a major contributor to these interactions.

As we watch many galaxies—some growing, shrinking, transforming or colliding—we realize that these galactic motions would be inexplicable without taking dark matter into account. Right in our own galactic neighborhood, for instance, such interactions are under way. The Magellanic Clouds, our second nearest neighboring galaxies, pass through our galaxy's plane every billion years. As they do, they mark their paths with tidal tails of gas and, possibly, stars. Indeed, on every passage, they lose energy and spiral inward. In less than 10 billion years, they will fragment and merge into the Milky Way.

Recently astronomers identified a still nearer neighboring galaxy, the Sagittarius dwarf, which lies on the far side of the Milky Way, close to its outer edge. (Viewed from Earth, this dwarf galaxy appears in the constellation Sagittarius.) As it turns out, gravity from our galaxy is pulling apart this dwarf galaxy, which will cease to exist as a separate entity after several orbits. Our galaxy itself may be made up of dozens of such previous acquisitions.

Similarly, the nearby galaxy M31 and the Milky Way are now hurtling toward each other at the brisk clip of 130 kilometers (81 miles) per second. As eager spectators, we must watch this encounter for a few decades to know if M31 will strike our galaxy or merely slide by. If they do collide, we will lose: the Milky Way will merge into the more massive M31. Computer models predict that in about four billion years the galactic pair will become one spheroidal galaxy. Of course, by then our sun will have burned out—so others in the universe will have to enjoy the pyrotechnics.

In many ways, our galaxy, like all large galaxies, behaves as no gentle neighbor. It gobbles up nearby companions and grinds them into building blocks for its own growth. Just as Earth's continents slide beneath our feet, so, too, does our galaxy evolve around us. By studying the spinning, twisting and turning motions and structures of many galaxies as they hurtle through space, astronomers can figure out the gravitational forces required to sustain their motions—and the amount of invisible matter they must contain.

How much dark matter does the universe contain? The destiny of the universe hinges on one still unknown parameter: the total mass of the universe. If we live in a high-density, or "closed," universe, then mutual gravitational attraction will ultimately halt the universe's expansion, causing it to contract—culminating in a big crunch, followed perhaps by reexpansion. If, on the other hand, we live in a low-density, or "open," universe, then the universe will expand forever.

Observations thus far suggest that the universe—or, at least, the region we can observe—is open, forever expanding. When we add up all the luminous matter we can detect, plus all the dark matter that we infer from observations, the total still comes to only a fraction—perhaps 20 percent—of the density needed to stop the universe from expanding forever.

I would be content to end the story there, except that cosmologists often dream of, and model, a universe with "critical" density—meaning one that is finely balanced between high and low density. In such a universe, the density is just right. There is enough matter to slow the universe's continuous expansion, so that it eventually coasts nearly to a halt. Yet this model does not describe the universe we actually measure. As an observer, I recognize that more matter may someday be detected, but this does not present sufficient reason for me to adopt a cosmological model that observations do not yet require.

Another complicating factor to take into account is that totally dark systems may exist—that is, there may be agglomerations of dark matter into which luminous matter has never penetrated. At present, we simply do not know if such totally dark systems exist because we have no observational data either to confirm or to deny their presence.

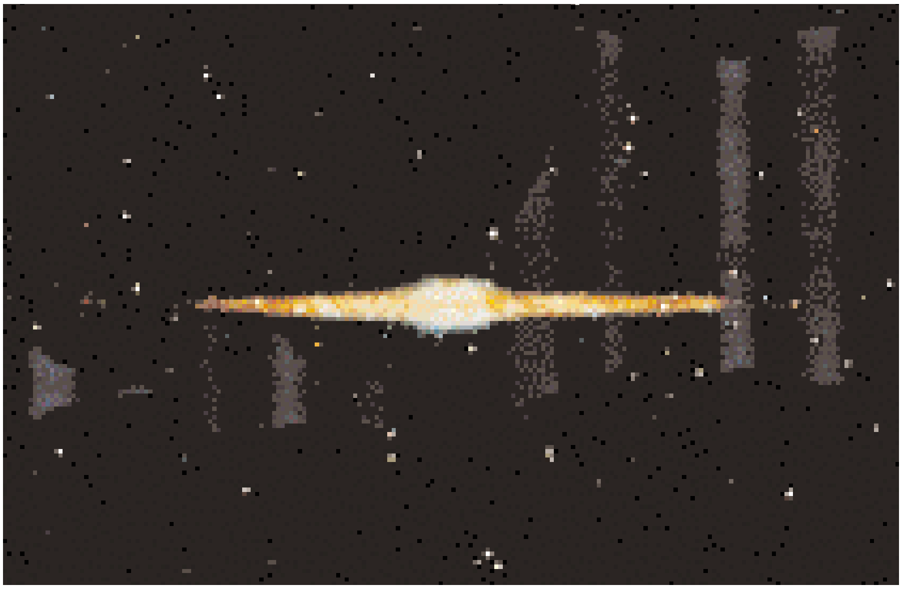
What Is Dark Matter?

Whatever dark matter turns out to be, we know for certain that the universe contains large amounts of it. For every gram of glowing material we can detect, there may be tens of grams of dark matter out there. Currently the astronomical jury is still out as to exactly what constitutes dark matter. In fact, one could say we are still at an early stage of exploration. Many candidates exist to account for the invisible mass, some relatively ordinary, others rather exotic.

Nevertheless, there is a framework in which we must work. Nucleosynthesis, which seeks to explain the origin of elements after the big bang, sets a limit to the number of baryons—particles of ordinary, run-of-the-mill matter—that can exist in the universe. This limit arises out of the Standard Model of the early universe, which has one free parameter—the ratio of the number of baryons to the number of photons.

From the temperature of the cosmic microwave background—which has been measured—the number of photons is now known. Therefore, to determine the number of baryons, we must observe stars and galaxies to learn the cosmic abundance of light nuclei, the only elements formed immediately after the big bang.

Without exceeding the limits of nucleosynthesis, we can construct an acceptable model of a low-density, open universe. In that model, we take approximately equal amounts of baryons and exotic matter (nonbaryonic particles), but in quantities that add up to only 20 percent of the matter needed to close the universe. This model universe matches all our actual observations. On the other hand, a slightly different model of an open universe in which all matter is baryonic would also satis-



NASA GODDARD SPACE FLIGHT CENTER; COURTESY OF COBE SCIENCE WORKING GROUP

INSIDER'S VIEW OF OUR GALAXY is shown here by its near-infrared radiation. Data come from the Diffuse Infrared Background Experiment (DIRBE), part of the National Aeronautics and Space Administration's space-based Cosmic Background Explorer (COBE). Viewing the Milky Way (its central bulge and surrounding disk) from within the disk, DIRBE captures the sky's changing brightness, observed at wavelengths 1.2, 2.3 and 3.5 microns (*blue, green and red, respectively*). The sun sits roughly 28,000 light-years from our galaxy's center. Such data help astronomers fathom the Milky Way's large-scale structure and evolution.

SPIRAL GALAXY NGC 2997, located in the southern Antlia cloud, probably resembles our own galaxy. Like all spiral galaxies, NGC 2997 is embedded in an extended dark diffuse halo, whose composition remains unknown.

©ANGLO-AUSTRALIAN OBSERVATORY; PHOTOGRAPHY BY DAVID F. MALIN

fy observations. Unfortunately, this alternative model contains too many baryons, violating the limits of nucleosynthesis. Thus, any acceptable low-density universe has mysterious properties: most of the universe's baryons would remain invisible, their nature unknown, and in most models much of the universe's matter is exotic.

Exotic Particles

Theorists have posited a virtual smorgasbord of objects to account for dark matter, although many of them have fallen prey to observational constraints. As leading possible candidates for baryonic dark matter, there are black holes (large and small), brown dwarfs (stars too cold and faint to radiate), sun-size MACHOs, cold gas, dark galaxies and dark clusters, to name only a few.

The range of particles that could constitute nonbaryonic dark matter is limited only slightly by theorists' imaginations. The particles include photinos, neutrinos, gravitinos, axions and magnetic monopoles, among many others. Of these, researchers have detected only neutrinos—and whether neutrinos have any mass remains unknown. Experiments are under way to detect other exotic particles. If they exist, and if one has a mass in the correct range, then that particle might pervade the universe and constitute dark matter. But these are very large "ifs."

To a great extent, the details of the evolution of galaxies and clusters depend on properties of dark matter. Without knowing those properties, it is difficult to explain how galaxies evolved into the structures observed today. As knowledge of the early universe deepens, I remain optimistic that we will soon know much more about both galaxy formation and dark matter.

What we fail to see with our eyes, or detectors, we can occasionally see with our minds, aided by computer graphics. Computers now play a key role in the search for dark matter. Historically, astronomers have focused on observations; now the field has evolved into an experimental science. Today's astronomical experimenters sit neither at lab benches nor at telescopes but at computer terminals. They scrutinize cosmic simulations in which tens of thousands of points, representing stars, gas and dark matter, interact gravitationally over a galaxy's lifetime. A cosmologist can tweak a simulation by adjusting the parameters of dark matter and then watch what happens as virtual galaxies evolve in isolation or in a more realistic, crowded universe.

Computer models can thus predict galactic behavior. For instance, when two galaxies suffer a close encounter, violently merging or passing briefly in the night, they sometimes spin off long tidal tails. Yet from the models, we now know these tails appear only when the dark matter of each galaxy's halo is three to 10 times greater than its luminous matter. Heavier halos produce stubbier tails. This realization through modeling has helped observational astronomers to interpret what they see and to understand more about the dark matter they cannot see. For the first time in the history of cosmology, computer simulations actually guide observations.

New tools, no less than new ways of thinking, give us insight into the structure of the heavens. Less than 400 years ago Galileo put a small lens at one end of a cardboard tube and a big brain at the other end. In so doing, he learned that the faint stripe across the sky, called the Milky Way, in fact comprised billions of single stars and stellar clusters. Suddenly, a human being understood what a galaxy is. Perhaps in the coming century, another—as yet unborn—big brain will put her eye to a clever new instrument and definitively answer, What is dark matter?

The Author

VERA RUBIN is a staff member at the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, where she has been since 1965. That same year, she became the first woman permitted to observe at Palomar Observatory. The author of more than 200 papers on the structure of the Milky Way, motions within galaxies and large-scale motions in the universe, she received Carnegie Mellon University's Dickson Prize for Science in 1994 and the Royal Astronomical Society's Gold Medal in 1996. President Bill Clinton awarded her the National Medal of Science in 1993 and appointed her to the President's Committee on the National Medal of Science in 1995.

A Scientific Armada

Space-deployed sensors promise a revolution in science's understanding of the cosmos

by Tim Beardsley, *staff writer*

HUBBLE SPACE TELESCOPE separates from the space shuttle *Discovery* over the Indian Ocean in February 1997, after receiving new instruments. Stunning Hubble images have inspired widespread interest in astrophysical phenomena.

NASA

A decade from now humanity's understanding of the solar system, no less the universe beyond, will have grown vastly more focused and detailed. During the next 10 years, roughly 50 scientific expeditions will blast off from Earth—a veritable armada of missions to visit planets, comets and asteroids, as well as to make sensitive observations of deep space from above Earth's occluding atmosphere. Researchers will very likely resolve some longstanding questions at the same time that they grapple with as yet undreamed of conundrums.

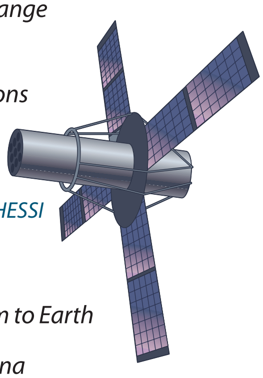
As many as nine spacecraft will intensively survey Mars during the next 10 years, including the Mars Global Surveyor now in orbit. If plans under consideration get the go-ahead, samples from the red planet will return to Earth for analysis sometime after 2005. Missions to Pluto and, perhaps, Mercury and Venus (though not currently scheduled) may advance onto the National Aeronautics and Space Administration's timetable.

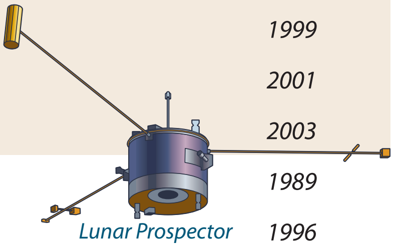
Material samples should also arrive at our home planet from remote interplanetary space, from an asteroid and from the tail of a comet. The moon will once again become a familiar destination for robotic spacecraft, including Japanese and European projects. These missions will map the moon's composition and attempt to settle the question—raised by recent radar observations—of whether Earth's orbital companion might harbor water ice near its south pole. In 2005 the Solar Probe mission might even hurtle toward the sun. Farther afield, by 2004, the giant Cassini spacecraft will arrive at Saturn and dispatch its accompanying Huygens probe to investigate the ringed planet's giant moon, Titan. A variety of scheduled missions are also planned to observe from afar the sun's violent outbursts, including Germany and China's Space Solar Tele-

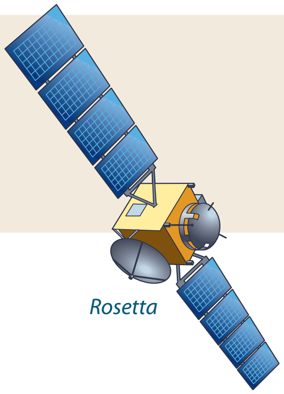
Continued on page 116

KEY SPACE EXPLORATIONS OF THE NEXT DECADE

	NAME OF MISSION (SPONSOR)	MAIN PURPOSE OF MISSION	LAUNCH DATE
The Sun	*ACE, Advanced Composition Explorer (NASA)*	*Monitor solar atomic particles and the interplanetary environment*	*1997*
	Coronas F (Russia)	*Observe the sun's spectrum during a solar maximum*	*1998*
	TRACE, Transition Region and Coronal Explorer (NASA)	*Photograph the sun's coronal plasmas in the ultraviolet range*	*1998*
	HESSI, High Energy Solar Spectroscopic Imager (NASA)	*Study solar flares through x-rays, gamma rays and neutrons*	*2000*
	Photon (Russia)	*Analyze gamma rays from the sun*	*2000*
	SST, Space Solar Telescope (China and Germany)	*Study the sun's magnetic field*	*2001*
	Genesis (NASA)	*Gather atomic nuclei from the solar wind and return them to Earth*	*2001*
	Solar Probe (NASA)	*Measure particles, fields, x-rays and light in the sun's corona*	*2003*
	Solar B (Japan)	*Study the sun's magnetic field around violent events*	*2004*
The Moon	*Lunar Prospector (NASA)*	*Study the moon's magnetic field and search for evidence of water at its poles*	*1998*
	Lunar A (Japan)	*Analyze the moon's subsurface soil*	*1999*
	Euromoon 2000 (ESA)	*Explore the moon's south pole (under study)*	*2001*
	Selene (Japan)	*Map the moon, studying fields and particles*	*2003*
The Planets	*Galileo (NASA)*	*Explore Jupiter and its moons*	*1989*
	Mars Global Surveyor (NASA)	*Map Mars and relay data from other missions*	*1996*
	Cassini (NASA)	*Explore the Saturn system; Huygens (ESA) will descend to Titan*	*1997*
	Planet B (Japan)	*Study interactions between the solar wind and Mars's atmosphere*	*1998*
	Mars Surveyor '98 (NASA)	*Explore a site near Mars's south pole (two-part mission)*	*1998 and 1999*
	Deep Space II (NASA)	*Analyze Martian subsurface soil*	*1999*
	Mars Surveyor 2001 (NASA)	*Land a rover that will travel many kilometers on Mars (two-part mission)*	*2001*
	Mars Surveyor 2003 (NASA)	*Collect Martian soil samples (two-part mission, under study)*	*2003*
	Mars Express (ESA)	*Analyze Martian soil, using an orbiter and two landers*	*2003*
	Pluto/Kuiper Express (NASA)	*Explore the solar system's only unvisited planet and the Kuiper belt (under study)*	*After 2003*
	Mars Sample Return (NASA)	*Return Martian rock and soil samples to Earth (under study)*	*After 2005*
Comets	*Stardust (NASA)*	*Encounter Comet Wild 2, collect particles from its tail and return the sample to Earth*	*1999*
	CONTOUR, Comet Nucleus Tour (NASA)	*Produce spectral maps of three comet nuclei*	*2002*
	Rosetta (ESA and France)	*Land a probe on Comet Wirtanen's nucleus*	*2003*
Asteroid Belt	*NEAR, Near Earth Asteroid Rendezvous (NASA)*	*Measure the composition, magnetic field and mass distribution of the asteroid Eros*	*1996*
	MUSES C (Japan)	*Return a sample of material from an asteroid*	*2002*

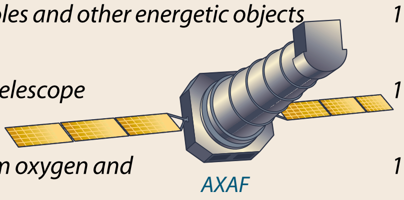

HESSI

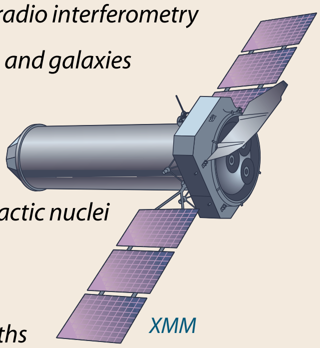

Lunar Prospector


Rosetta

Deep Space

NAME OF MISSION (SPONSOR)	MAIN PURPOSE OF MISSION	LAUNCH DATE
RXTE, Rossi X-ray Timing Explorer (NASA)	*Watch x-ray sources change over time*	*1995*
Beppo-SAX (Italy)	*Observe x-ray sources over a wide energy range*	*1996*
HALCA (Japan)	*Study galactic nuclei and quasars via radio interferometry*	*1997*
FUSE, Far Ultraviolet Spectroscopic Explorer (NASA)	*Detect deuterium in interstellar space*	*1998*
AXAF, Advanced X-ray Astrophysics Facility (NASA)	*Procure x-ray images and spectra of black holes and other energetic objects*	*1998*
WIRE, Wide-Field Infrared Explorer (NASA)	*Observe galaxy formation with a cryogenic telescope*	*1998*
Odin (Sweden)	*Detect millimeter-wavelength emissions from oxygen and water in interstellar gas*	*1998*
SWAS, Submillimeter Wave Astronomy Satellite (NASA)	*Search for oxygen, water and carbon in interstellar clouds*	*1999*
ABRIXAS, A Broad-band Imaging X-ray All-sky Survey (Germany)	*Make a hard x-ray, all-sky survey*	*1999*
Spectrum X-gamma (Russia)	*Measure x-ray emissions from pulsars, black holes, supernova remnants and active galactic nuclei*	*1999*
HETE II, High Energy Transient Explorer (NASA)	*Study gamma-ray bursters with x-ray and gamma-ray detectors*	*1999*
XMM, High-Throughput X-ray Spectroscopy Mission (ESA)	*Observe spectra of cosmic x-ray sources*	*1999*
Astro-E (Japan)	*Make high-resolution x-ray observations*	*2000*
MAP, Microwave Anisotropy Probe (NASA)	*Study the universe's origin and evolution through the cosmic microwave background*	*2000*
Radioastron (Russia)	*Observe quasars and high-energy phenomena via radio interferometry*	*2000*
SIRTF, Space Infrared Telescope Facility (NASA)	*Make high-resolution infrared observations of stars and galaxies*	*2001*
Corot (France)	*Search for evidence of planets around distant stars*	*2001*
INTEGRAL, International Gamma Ray Astrophysics Lab (ESA)	*Obtain spectra of neutron stars, black holes, gamma-ray bursters, x-ray pulsars and active galactic nuclei*	*2001*
GALEX, Galaxy Evolution Explorer (NASA)	*Observe stars, galaxies and heavy elements at ultraviolet wavelengths (under study)*	*2001*
Spectrum UV (Russia)	*Study astrophysical objects at ultraviolet wavelengths*	*2001*
SIM, Space Interferometry Mission (NASA)	*Image stars that may host Earth-like planets (under study)*	*2004*
HTXS, Constellation X-ray Mission (NASA)	*Perform high-resolution x-ray spectroscopy (under study)*	*After 2005*
OWL, Orbiting Wide-angle Light collectors (NASA)	*Study cosmic-ray effects on Earth's atmosphere (under study)*	*After 2005*
FIRST, Far Infrared Submillimeter Telescope, and Planck (ESA)	*Discern the fine structure of the cosmic microwave background (combined mission)*	*2006*
Next Generation Space Telescope (NASA)	*View space, near and far, at infrared wavelengths (under study)*	*2007*
TPF, Terrestrial Planet Finder (NASA)	*Find planets and protoplanets orbiting nearby stars (under study)*	*2009*


AXAF


XMM

ILLUSTRATIONS BY JARED SCHNEIDMAN DESIGN

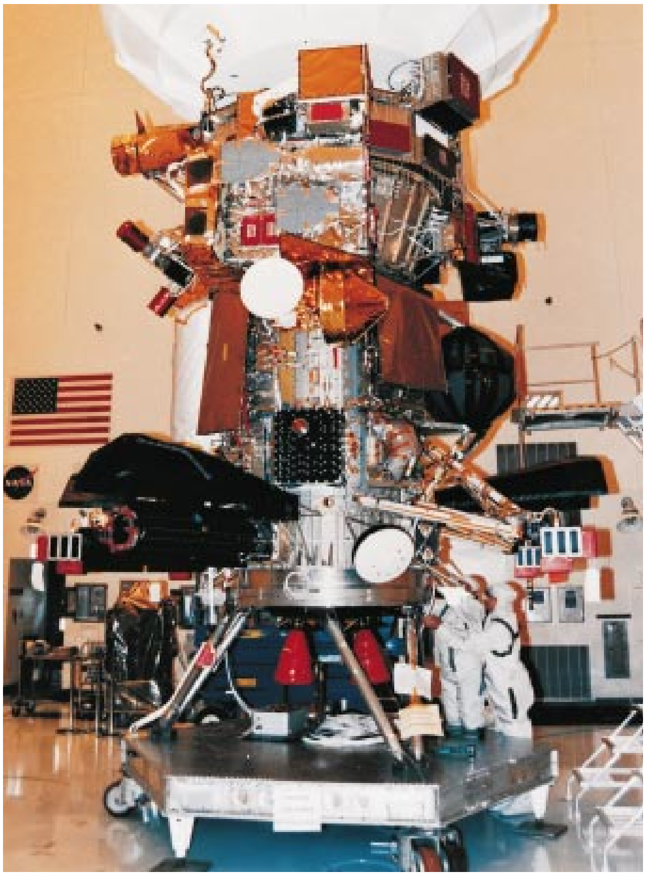
Continued from page 112

scope, Russia's Coronas F and Photon, and Japan's Solar B.

Looking beyond the solar system, an equally impressive fleet of sensitive detectors will soon be launched to make images and analyze radiation and particles emanating from deep space. Although several such observatories have lifted off in recent years, the turn of the century will witness many more launches. These new instruments will vastly outperform their predecessors, because of advances in sensor and computer technology. At least 20 nations will participate in this grand exploration. The great majority of the missions will feature international collaboration at some level—and even a little competition as well.

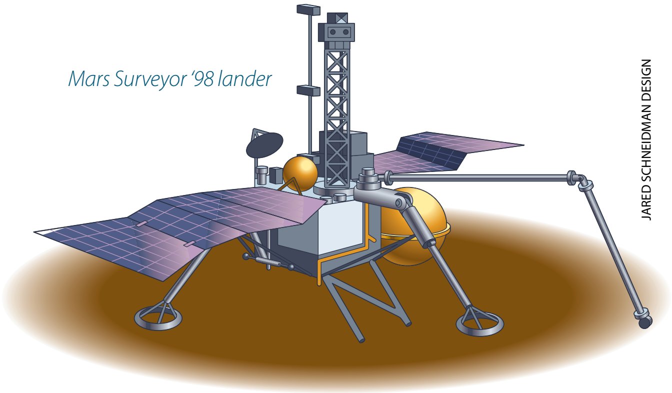
The U.S., Russia, Japan and the European Space Agency (ESA) stand out as major players, but they are not the only ones. Various smaller explorations are being planned. India and Sweden have observation programs, for example, and France, Germany and Italy run substantial projects aside from their membership in the ESA. Many countries not flying spacecraft will contribute instruments or lend the use of tracking facilities. International collaboration seems likely to increase, especially with expensive endeavors such as the projected Mars Sample Return Mission. Although recently NASA has put an emphasis on low-cost, bare-bones space science missions, a number of costly enterprises lie on its drawing board.

By 2008, gamma-ray bursters may seem less mysterious than they do today, thanks to a squadron of satellites designed to identify and observe these brief but cataclysmic events. Quasars and active galactic nuclei of all kinds—not to mention our own galaxy's center—will come under intense scrutiny at x-ray and gamma-ray wavelengths. Among the most significant of these planned high-energy observatories are INTEGRAL, an ESA project to be launched in 2001 to study gamma-ray sources; the High-Throughput X-ray Spectroscopy Mission, due to be lofted in 1999; and the long-delayed Advanced X-ray Astrophysics Facility (AXAF), scheduled for liftoff in 1998. Radioastronomers, too, will observe such high-energy regions with unprecedented resolution by way of very long baseline interferometry, which combines satellite measurements with those taken with antennae on Earth. HALCA, a Japanese radioastronomy satellite using this technique, is already in operation, and Russia plans to launch a larger one, Radioastron, in 2000. Other specialized sensors, NASA's Microwave Anisotropy Probe (MAP) and the ESA's FIRST/Planck combined mission, will eavesdrop on the cosmic microwave background radiation, a survey that could reveal much about conditions in the universe's first moments.



NASA

CASSINI is examined by engineers at the Jet Propulsion Laboratory in Pasadena, Calif. The spacecraft was launched in October 1997 toward Saturn, where it should arrive in 2004.



Mars Surveyor '98 lander

JARED SCHNEIDMAN DESIGN

An important push will come in infrared astronomy, best suited for studying the formation of galaxies, stars and planets. One eagerly anticipated event will arrive with the launch in 2001 of the Space Infrared Telescope Facility (SIRTF). This spacecraft, and some other infrared and submillimeter-wavelength observatories, will extend the capabilities of the ESA's Infrared Space Observatory, which was expected to run out of cryogenic coolant in early 1998.

Although most space-based observatories orbit Earth, infrared observatories benefit from distance. SIRTF will orbit the sun 48 million kilometers (30 million miles) from Earth, whereas some other infrared and submillimeter-wavelength observatories will hover around a gravitationally stable point two million kilometers distant from Earth in the direction away from the sun. The Next Generation Space Telescope, tentatively planned for launch around 2007, will carry a high-resolution infrared observatory—a worthy successor to the Hubble Space Telescope, which will wind down sometime after 2005.

Radioastronomers will not be the only ones using interferometry. The technically daunting challenge of space-based optical interferometry serves as the aim of the Space Interferometry Mission, now under study. Two separate optical telescopes, separated by a 10-meter boom, would combine forces to achieve unprecedented resolution. An even more ambitious mission known as Terrestrial Planet Finder, still in the early planning stage, would employ infrared interferometry to search for Earth-size planets around distant stars—a key component of NASA's theme of studying origins.

All plans and dates for space science missions remain subject to changes. Nevertheless, the size and scope of this scientific armada reveal that human beings have a compelling drive to understand how our universe came to be. SA