SPECIAL EDITION

SCIENTIFIC AMERICAN

WWW.SCIAM.COM

THE SECRET LIVES OF STARS

The Fury of Magnetars

When Stars Collide

Our Fickle Sun

Stormy Dynamos

Ultradense Neutron Stars

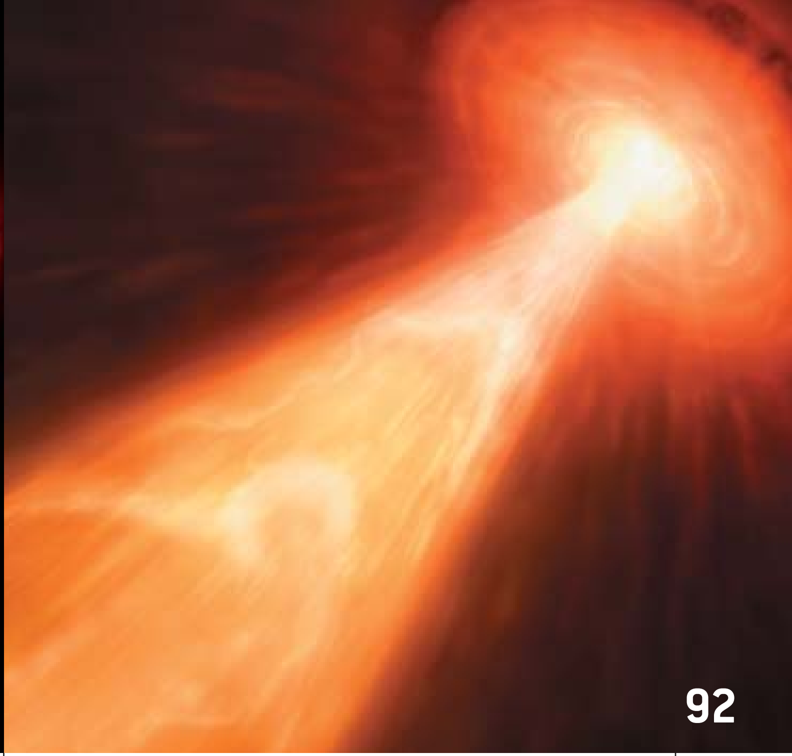
Cannibalistic White Dwarfs

Bright Births of Black Holes

SCIENTIFIC AMERICAN Volume 14, Number 4

The Secret Lives of Stars

contents 2004



Cover illustration by Don Dixon

Scientific American Special (ISSN 1048-0943), Volume 14, Number 4, 2004, published by Scientific American, Inc., 415 Madison Avenue, New York, NY 10017-1111.

Canadian BN No. 127387652RT; QST No. Q1015332537. To purchase additional quantities: U.S., $10.95 each; elsewhere, $13.95 each. Send payment to Scientific American, Dept. STARS, 415 Madison Avenue, New York, NY 10017-1111. Inquiries: fax 212-355-0408 or telephone 212-451-8890. Printed in U.S.A.

letter from the editor

Exposé on the Stars

ENCOUNTER with a tempestuous starlet tears apart our sun.

MARK A. GARLICK

FOR PURE THEATRICS and spectacle, Hollywood celebrities have nothing on the denizens of the heavens. Stars are born, live and die in fiery and fascinating ways—ways that we have only recently been able to study in greater detail, like so many swarming paparazzi, using the long-range lenses created by improved techniques and new, sharper observatories.

We've observed some eye-popping behavior from these brazen orbs. As with many glamorous beings, stars often pair off, but the two may blow apart—literally—one star ripping the shine from its former partner in a highly public display. It seems that only about half the stars live as couples—no better than the average state of human relationships. Or consider the wondrous strange doings of the magnetar, a name itself worthy of a character in the blockbuster X-Men films. These intensely magnetized stars emit huge bursts of magnetic energy that can alter the very nature of the quantum vacuum. Like rebels without a cause, they are furiously active for only a short time: after 10,000 years they wink out. We've also been able to see how, after a long era of flamboyant x-ray emission, x-ray binaries settle down to become some of the most steady, unchanging entities in the cosmos. Not all of these celestial objects make it as luminaries, however. We may feel some sympathy for brown dwarfs, failed stars that glow so dully nobody could even find one until 1995.

In this special edition from *Scientific American,* we invite you to forget about everyday life to spend some time with the stars. In the pages that follow, you'll find the latest gossip on the glitterati, written by the astronomer shutterbugs themselves. Although the stars have revealed more than ever, they've been careful not to tell all, lest we grow bored with their antics. As authors Chryssa Kouveliotou, Robert C. Duncan and Christopher Thompson so aptly put it in "Magnetars": "What other phenomena, so rare and fleeting that we have not recognized them, lurk out there?" We can hardly wait to find out.

Mariette DiChristina
Executive Editor
Scientific American
editors@sciam.com

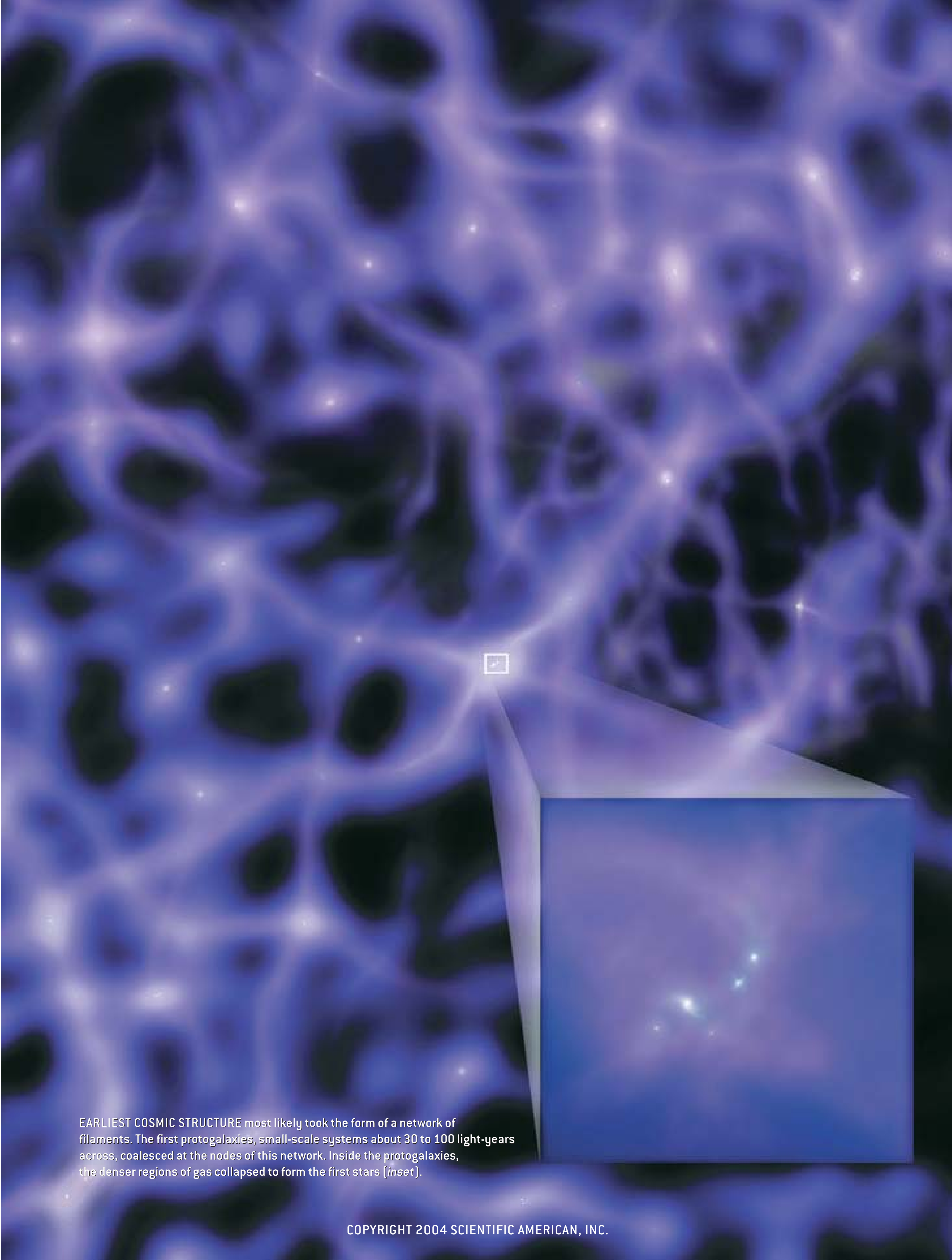
THE FIRST STARS IN THE UNIVERSE

Exceptionally massive and bright, the earliest stars changed the course of cosmic history

BY RICHARD B. LARSON AND VOLKER BROMM
ILLUSTRATIONS BY DON DIXON

We live in a universe that is full of bright objects. On a clear night one can see thousands of stars with the naked eye. These stars occupy merely a small nearby part of the Milky Way galaxy; telescopes reveal a much vaster realm that shines with the light from billions of galaxies. According to our current understanding of cosmology, however, the universe was featureless and dark for a long stretch of its early history. The first stars did not appear until perhaps 100 million years after the big bang, and nearly a billion years passed before galaxies proliferated across the cosmos. Astronomers have long wondered: How did this dramatic transition from darkness to light come about?

Updated from the December 2001 issue



EARLIEST COSMIC STRUCTURE most likely took the form of a network of filaments. The first protogalaxies, small-scale systems about 30 to 100 light-years across, coalesced at the nodes of this network. Inside the protogalaxies, the denser regions of gas collapsed to form the first stars (*inset*).

After decades of study, researchers have recently made great strides toward answering this question. Using sophisticated computer simulation techniques, cosmologists have devised models that show how the density fluctuations left over from the big bang could have evolved into the first stars. In addition, observations of distant quasars have allowed scientists to probe back in time and catch a glimpse of the final days of the "cosmic dark ages."

The new models indicate that the first stars were most likely quite massive and luminous and that their formation was an epochal event that fundamentally changed the universe and its subsequent evolution. These stars altered the dynamics of the cosmos by heating and ionizing the surrounding gases. The earliest stars also produced and dispersed the first heavy elements, paving the way for the eventual formation of solar systems like our own. And the collapse of some of the first stars may have seeded the growth of supermassive black holes that formed in the hearts of galaxies and became the spectacular power sources of quasars. In short, the earliest stars made possible the emergence of the universe that we see today—everything from galaxies and quasars to planets and people.

The Dark Ages

THE STUDY of the early universe is hampered by a lack of direct observations. Astronomers have been able to examine much of the universe's history by training their telescopes on distant galaxies and quasars that emitted their light billions of years ago. The age of each object can be determined by the redshift of its light, which shows how much the universe has expanded since the light was produced. The oldest galaxies and quasars that have been observed so far date from about a billion years after the big bang (assuming a present age for the universe of 13.7 billion years). Researchers will need better telescopes to see more distant objects dating from still earlier times.

Cosmologists, however, can make deductions about the early universe based on the cosmic microwave background radiation, which was emitted about 400,000 years after the big bang. The uniformity of this radiation indicates that matter was distributed very smoothly at that time. Because there were no large luminous objects to disturb the primordial soup, it must have remained smooth and featureless for millions of years afterward. As the cosmos expanded, the background radiation redshifted to longer wavelengths and the universe grew increasingly cold and dark. Astronomers have no observations of this dark era. But by a billion years after the big bang, some bright galaxies and quasars had already appeared, so the first stars must have formed sometime before. When did these first luminous objects arise, and how might they have formed?

Many astrophysicists, including Martin Rees of the University of Cambridge and Abraham Loeb of Harvard University, have made important contributions toward solving these problems. The recent studies begin with the standard cosmological models that describe the evolution of the universe following the big bang. Although the early universe was remarkably smooth, the background radiation shows evidence of small-scale density fluctuations—clumps in the primordial soup. These clumps would gradually evolve into gravitationally bound structures. Smaller systems would form first and then merge into larger agglomerations. The denser regions would take the form of a network of filaments, and the first star-forming systems—small protogalaxies—would coalesce at the nodes of this network. In a similar way, the protogalaxies would then merge to form galaxies, and the galaxies would congregate into galaxy clusters. The process is ongoing: although galaxy formation is now mostly complete, galaxies are still assembling into clusters, which are in turn aggregating into a vast filamentary network that stretches across the universe.

According to the models, the first small systems capable of forming stars should have appeared between 100 million and 250 million years after the big bang. These protogalaxies would have been 100,000 to one million times more massive than the sun and would have measured 30 to 100 light-years across. These properties are similar to those of

It seems safe to conclude that the FIRST STARS IN THE UNIVERSE WERE TYPICALLY MANY TIMES more massive and luminous than the sun.

Overview/*The First Stars*

- Computer simulations show that the first stars should have appeared between 100 million and 250 million years after the big bang. They formed in small protogalaxies that evolved from density fluctuations in the early universe.
- Because the protogalaxies contained virtually no elements besides hydrogen and helium, the physics of star formation favored the creation of bodies that were many times more massive and luminous than the sun.
- Radiation from the earliest stars ionized the surrounding hydrogen gas. Some stars exploded as supernovae, dispersing heavy elements throughout the universe. The most massive stars collapsed into black holes. As protogalaxies merged to form galaxies, the black holes possibly became concentrated in the galactic centers.

the molecular gas clouds in which stars are currently forming in the Milky Way, but the first protogalaxies would have differed in fundamental ways. For one, they would have consisted mostly of dark matter, the putative elementary particles that are believed to make up 90 percent of the universe's mass. In present-day large galaxies, dark matter is segregated from ordinary matter: over time, ordinary matter concentrates in the galaxy's inner region, whereas the dark matter remains scattered throughout an enormous outer halo. But in the protogalaxies, the ordinary matter would still have been mixed with the dark matter.

The second important difference is that the protogalaxies would have contained no significant amounts of any elements besides hydrogen and helium. The big bang produced hydrogen and helium, but most of the heavier elements are created only by the thermonuclear fusion reactions in stars, so they would not have been present before the first stars had formed. Astronomers use the term "metals" for all these heavier elements. The young metal-rich stars in the Milky Way are called Population I stars, and the old metal-poor stars are called Population II stars. The stars with no metals at all—the very first generation—are sometimes called Population III stars.

In the absence of metals, the physics of the first star-forming systems would have been much simpler than that of present-day molecular gas clouds. Furthermore, the cosmological models can provide, in principle, a complete description of the initial conditions that preceded the first generation of stars. In contrast, the stars that arise from molecular gas clouds are born in complex environments that have been altered by the effects of previous star formation. Several research groups have used computer simulations to portray the formation of the earliest stars.

A team consisting of Tom Abel, Greg Bryan and Michael L. Norman (now at Pennsylvania State University, Columbia University and the University of California at San Diego, respectively) has made the most realistic simulations. In collaboration with Paolo Coppi of Yale University, we have done simulations based on simpler assumptions but intended to explore a wider range of possibilities. Toru Tsuribe, now at Osaka University in Japan, has made similar calculations using more powerful computers. The work of Fumitaka Nakamura and Masayuki Umemura (now at Niigata and Tsukuba universities in Japan, respectively) has yielded instructive results. All these studies have produced similar descriptions of how the earliest stars might have been born.

Let There Be Light!

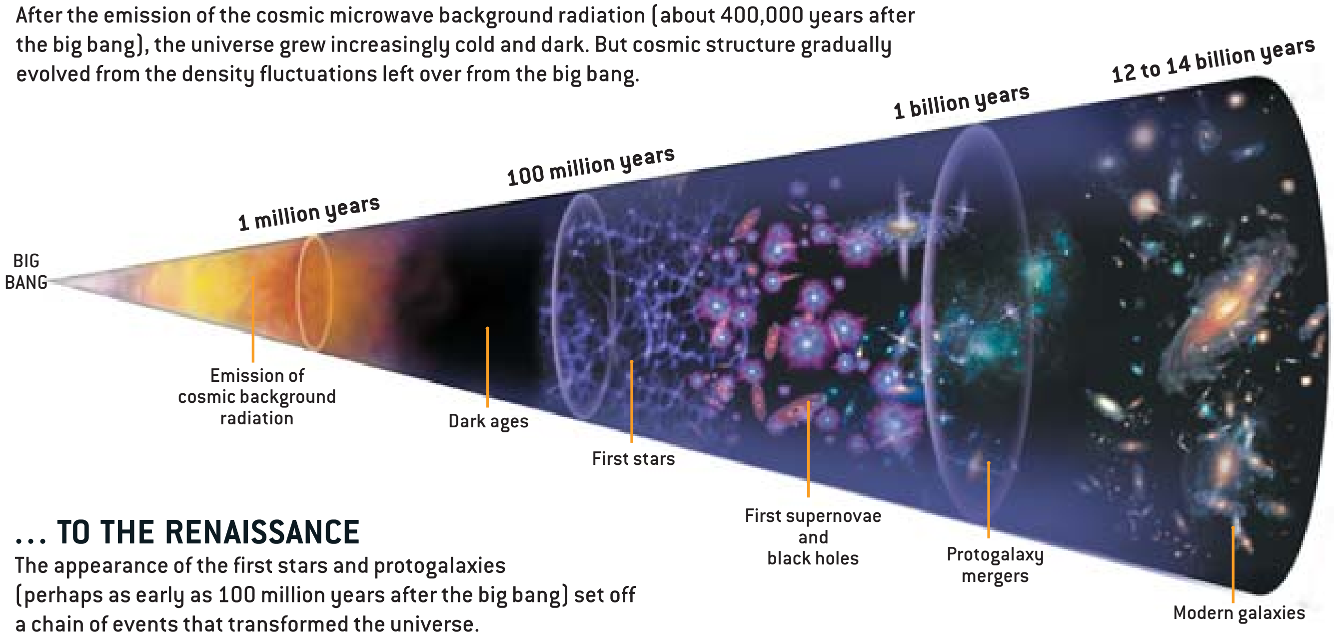
THE SIMULATIONS show that the primordial gas clouds would typically form at the nodes of a small-scale filamentary network and then begin to contract because of their gravity. Compression would heat the gas to temperatures above 1,000 kelvins. Some hydrogen atoms would pair up in the dense, hot gas, creating trace amounts of molecular hydrogen. The hydrogen molecules would then start to cool the densest parts of the gas by emitting infrared radiation after they collided with hydrogen atoms. The temperature in the densest parts would drop to 200 to 300 kelvins, reducing the gas pressure in these regions, allowing them to contract into gravitationally bound clumps.

This cooling plays an essential role in allowing the ordinary matter in the primordial system to separate from the dark matter. The cooling hydrogen would set-

COSMIC TIMELINE

FROM THE DARK AGES ...

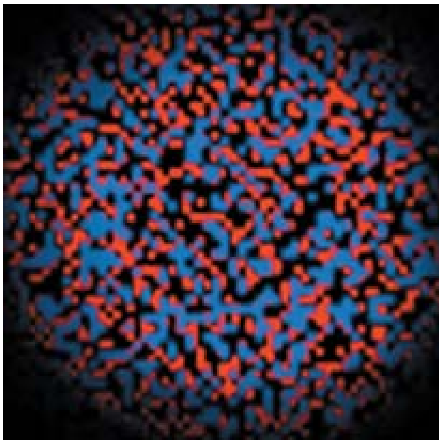
After the emission of the cosmic microwave background radiation (about 400,000 years after the big bang), the universe grew increasingly cold and dark. But cosmic structure gradually evolved from the density fluctuations left over from the big bang.



... TO THE RENAISSANCE

The appearance of the first stars and protogalaxies (perhaps as early as 100 million years after the big bang) set off a chain of events that transformed the universe.

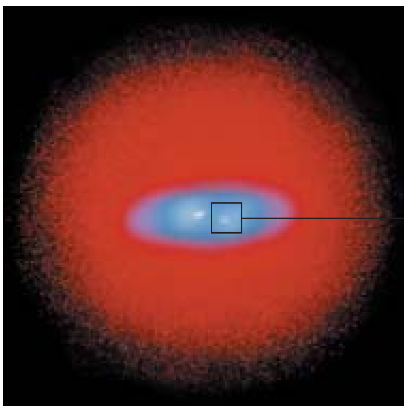
THE BIRTH AND DEATH OF THE FIRST STARS



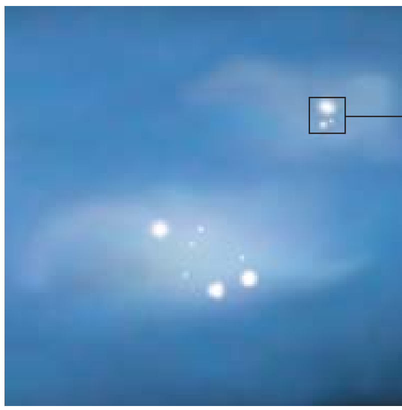
PRIMEVAL TURMOIL

The process that led to the creation of the first stars was very different from present-day star formation. But the violent deaths of some of these stars paved the way for the emergence of the universe that we see today.

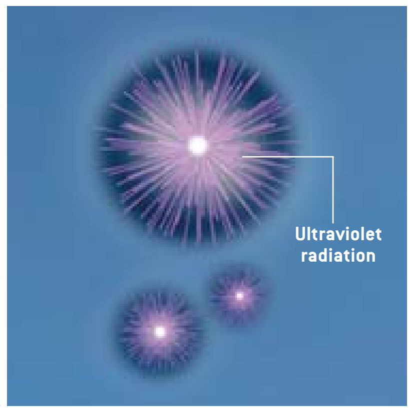
1 The first star-forming systems—small protogalaxies—consisted mostly of the elementary particles known as dark matter (*shown in red*). Ordinary matter—mainly hydrogen gas (*blue*)—was initially mixed with the dark matter.



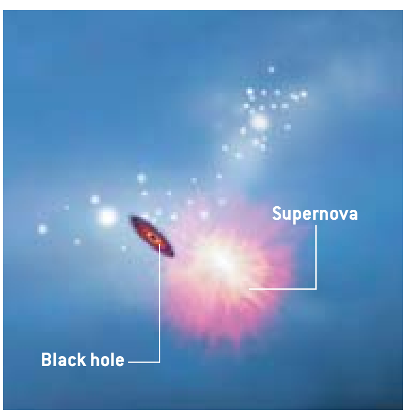
2 The cooling of the hydrogen allowed the ordinary matter to contract, whereas the dark matter remained dispersed. The hydrogen settled into a disk at the center of the protogalaxy.



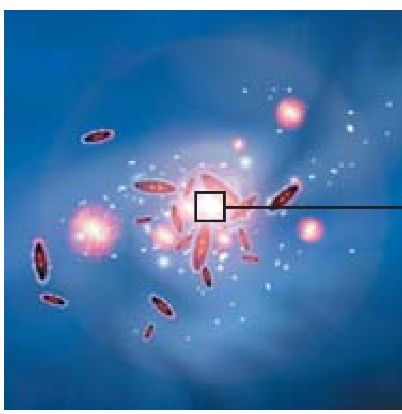
3 The denser regions of gas contracted into star-forming clumps, each hundreds of times as massive as the sun. Some of the clumps of gas collapsed to form very massive, luminous stars.



4 Ultraviolet radiation from the stars ionized the surrounding neutral hydrogen gas. As more and more stars formed, the bubbles of ionized gas merged and the intergalactic gas became ionized.



5 A few million years later, at the end of their brief lives, some of the first stars exploded as supernovae. The most massive stars collapsed into black holes.



6 Gravitational attraction pulled the protogalaxies toward one another. The collisions most likely triggered star formation, just as galactic mergers do now.

7 Black holes possibly merged to form a supermassive hole at the protogalaxy's center. Gas swirling into this hole might have generated quasarlike radiation.

tle into a flattened rotating configuration that was clumpy and filamentary and possibly shaped like a disk. But because the dark-matter particles would not emit radiation or lose energy, they would remain scattered in the primordial cloud. Thus, the star-forming system would come to resemble a miniature galaxy, with a disk of ordinary matter and a halo of dark matter. Inside the disk, the densest clumps of gas would continue to contract, and eventually some of them would undergo a runaway collapse and become stars.

The first star-forming clumps were much warmer than the molecular gas clouds in which most stars currently form. Dust grains and molecules containing heavy elements cool the present-day clouds much more efficiently to temperatures of only about 10 kelvins. The minimum mass that a clump of gas must have to collapse under its gravity is called the Jeans mass, which is proportional to the square of the gas temperature and inversely proportional to the square root of the gas pressure. The first star-forming systems would have had pressures similar to those of present-day molecular clouds. But because the temperatures of the first collapsing gas clumps were almost 30 times higher, their Jeans mass would have been almost 1,000 times larger.

In molecular clouds in the nearby part of the Milky Way, the Jeans mass is roughly equal to the mass of the sun, and the masses of the prestellar clumps are about the same. If we scale up, we can estimate that the masses of the first star-forming clumps would have been 500 to 1,000 solar masses. The computer simulations mentioned above showed the formation of clumps with masses of several hundred solar masses or more.

Our group's calculations suggest that the predicted masses of the first star-forming clumps are not very sensitive to the assumed cosmological conditions. The predicted masses depend primarily on the physics of the hydrogen molecule and only secondarily on the cosmological model or simulation technique. One reason is that molecular hydrogen cannot cool the gas below 200 kelvins, making this a lower limit to the temperature of the first star-forming clumps. Another is that the cooling from molecular hydrogen becomes inefficient at the higher densities encountered when the clumps begin to collapse. At these densities the hydrogen molecules collide with other atoms before they have time to emit an infrared photon; this raises the gas temperature and slows the contraction until the clumps have built up to at least a few hundred solar masses.

Did the first collapsing clumps form stars with similarly large masses, or did they fragment and form many smaller stars? The research groups have pushed their calculations to the point at which the clumps are well on their way to forming stars, and none of the simulations has yet revealed any tendency for the clumps to fragment. This agrees with our understanding of present-day star formation; the fragmentation of clumps is typically limited to the formation of binary systems (two stars orbiting around each other). Fragmentation seems even less likely to occur in the primordial clumps, because the inefficiency of molecular hydrogen cooling would keep the Jeans mass high. The simulations, however, have not yet determined the final outcome of collapse with certainty, and the formation of binary systems cannot be ruled out.

Precise estimates of just how massive the first stars might have been are difficult because of feedback effects. In general, a star forms from the "inside out," by accreting gas from the surrounding clump onto a central protostellar core. But when does this accretion process shut off? As the star grows in mass, it produces intense radiation and matter outflows that may blow away some of the gas in the collapsing clump. Yet these effects depend strongly on the presence of heavy elements, and therefore they should be less important for the earlier stars. In collaboration with Loeb of Harvard, one of us (Bromm) has recently used numerical simulations to study the accretion onto a primordial protostar. The calculations show that a Population III star grows to roughly 50 solar masses within the first 10,000 years after the initial core forms. Although we could not follow the accretion further because of numerical limitations, it is likely that the star continues to grow, perhaps to 100 to 200 solar masses. It seems safe to conclude that the first stars were typically many times more massive and luminous than the sun.

The Cosmic Renaissance

WHAT EFFECTS did these first stars have on the rest of the universe? An important property of stars with no metals is that they have higher surface temperatures than stars with compositions like that of the sun. The production of nuclear energy at the center of a star is less efficient without metals, and the star would have to be hotter and more compact to produce enough energy to counteract gravity. Because of the more compact structure, the surface layers of the star would also be hotter. In collaboration with Loeb and Rolf-Peter Kudritzki of the University of Hawaii Institute for Astronomy, Bromm devised theoretical models of such stars with masses between 100 and 1,000 solar masses. The models showed that the stars had surface temperatures of 100,000 kelvins—about 17 times higher than the sun's surface temperature. Thus, the first starlight in the universe would have been mainly ultraviolet radiation from very hot stars, and it would have begun to heat and ionize the neutral hydrogen and helium gas around these stars soon after they formed.

We call this event the cosmic renaissance. Although astronomers cannot yet

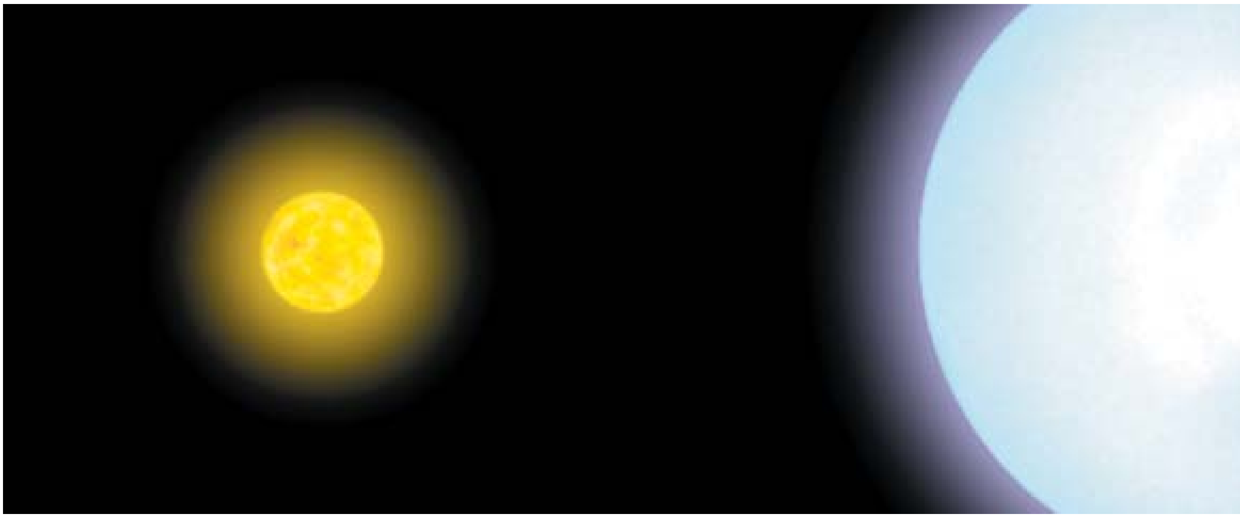
THE AUTHORS

RICHARD B. LARSON and *VOLKER BROMM* have worked together to understand the processes that ended the "cosmic dark ages" and brought about the birth of the first stars. Larson, a professor of astronomy at Yale University, joined the faculty there in 1968 after receiving his Ph.D. from the California Institute of Technology. His research interests include the theory of star formation as well as the evolution of galaxies. Bromm earned his Ph.D. at Yale in 2000 and is now an assistant professor of astronomy at the University of Texas at Austin, where he focuses on the emergence of cosmic structure. The authors acknowledge the many contributions of Paolo Coppi, professor of astronomy at Yale, to their joint work on the formation of the first stars.

STAR STATS

COMPARING CHARACTERISTICS

Computer simulations have given scientists some indication of the possible masses, sizes and other characteristics of the earliest stars. The lists below compare the best estimates for the first stars with those for the sun.



SUN

MASS: **1.989×10^{30} kilograms**
RADIUS: **696,000 kilometers**
LUMINOSITY: **3.85×10^{23} kilowatts**
SURFACE TEMPERATURE: **5,780 kelvins**
LIFETIME: **10 billion years**

FIRST STARS

MASS: **100 to 1,000 solar masses**
RADIUS: **4 to 14 solar radii**
LUMINOSITY: **1 million to 30 million solar units**
SURFACE TEMPERATURE: **100,000 to 110,000 kelvins**
LIFETIME: **3 million years**

estimate how much of the gas in the universe condensed into the first stars, even as little as one part in 100,000 could have been enough for these stars to ionize much of the remaining gas. Once the first stars started shining, a growing bubble of ionized gas would have formed around each one. As more and more stars formed over hundreds of millions of years, the bubbles of ionized gas would have merged, and the intergalactic gas would have become completely ionized.

Scientists from the California Institute of Technology and the Sloan Digital Sky Survey have found evidence for the final stages of this ionization process. They observed strong absorption of ultraviolet light in the spectra of quasars that date from about 900 million years after the big bang. The results suggest that the last patches of neutral hydrogen gas were being ionized at that time. A different probe has recently provided clues to the earliest stages of reionization, already occurring only 200 million years after the big bang. In an important breakthrough, NASA's Wilkinson Microwave Anisotropy Probe (WMAP) has measured the fundamental properties of the universe with high precision. These include the age of the universe—precisely 13.7 billion years—the proportions of dark and luminous matter, and dark energy in the cosmos. The biggest surprise: scrutinizing the subtle patterns that were imprinted into the photons of the cosmic microwave background, WMAP has indicated that ultraviolet radiation from the first stars ionized atomic hydrogen and helium, providing an abundance of free electrons early in cosmic history. Microwave background photons were polarized as they interacted with these electrons. An early generation of massive Population III stars seems to be required to account for the surprising strength of the polarization patterns.

Helium requires more energy to ionize than hydrogen does, but if the first stars were as massive as predicted, they would have ionized helium at the same time. On the other hand, if the first stars were not quite so massive, the helium must have been ionized later by energetic radiation from sources such as quasars. Future observations of distant objects may help determine when the universe's helium was ionized.

If the first stars were indeed very massive, they would also have had relatively short lifetimes—only a few million years. Some of the stars would have exploded as supernovae, expelling the metals they produced. Stars that are between 100 and 250 times as massive as the sun are predicted to blow up completely in energetic explosions, and some of the first stars most likely had masses in this range. Because metals are much more effective than hydrogen in cooling star-forming clouds and allowing them to collapse into stars, the production and dispersal of even a small amount could have had a major effect on star formation.

Working in collaboration with Andrea Ferrara of the Astrophysical Observatory of Arcetri in Italy, we have found that when the abundance of metals in star-forming clouds rises above one thousandth of the metal abundance in the sun, the metals rapidly cool the gas to the temperature of the cosmic background radiation. (This temperature declines as the universe expands, falling to 19 kelvins one billion years after the big bang and to 2.7 kelvins today.) This efficient cooling allows the formation of stars with smaller masses and may also considerably boost the rate at which stars are born. It is possible that the pace of star formation did not accelerate until after the first metals had been produced. In this case, the second-generation stars might have been the ones primarily responsible for lighting up the universe and bringing about the cosmic renaissance.

At the start of this active period of star birth, the cosmic background temperature would have been higher than in present-day molecular clouds (10 kelvins). Until the temperature dropped to that level—which happened about two billion years after the big bang—the process of star formation may still have favored massive stars. As a result, many such stars may have formed during the early stages of galaxy building by successive mergers of protogalaxies. A similar phenomenon may occur in the modern universe when two galaxies collide and trigger a starburst—a sudden increase in the rate of star formation—producing relatively large numbers of massive stars.

Puzzling Evidence

THIS HYPOTHESIS about early star formation might help explain some puzzling features of the present universe. One unsolved problem is that galaxies contain fewer metal-poor stars than would be expected if metals were produced at a rate proportional to the star formation rate. This discrepancy might be resolved if early star formation had produced relatively more massive stars; on dying, these stars would have dispersed large amounts of metals, which would have then been incorporated into most of the low-mass stars that we now see.

Another puzzling feature is the high metal abundance of the hot x-ray-emitting intergalactic gas in clusters of galaxies. This observation could be accounted for most easily if there had been an early period of rapid formation of massive stars and a correspondingly high supernova rate that chemically enriched the intergalactic gas. This case also dovetails with the recent evidence suggesting that most of the ordinary matter and metals in the universe lies in the diffuse intergalactic medium rather than in galaxies. To produce such a distribution of matter, galaxy formation must have been a spectacular process, involving intense bursts of massive star formation and barrages of supernovae that expelled most of the gas and metals out of the galaxies.

Stars that are more than 250 times more massive than the sun do not explode at the end of their lives; instead they collapse into massive black holes. Several of the simulations mentioned above predict that some of the first stars would have had masses this great. Because the first stars formed in the densest parts of the universe, any black holes resulting from their collapse would have become incorporated, via successive mergers, into systems of larger and larger size. It is possible that some of these black holes became concentrated in the inner part of large galaxies and seeded the growth of the supermassive black holes that are now found in galactic nuclei.

Furthermore, astronomers believe that the energy source for quasars is the gas whirling into the black holes at the centers of large galaxies. If smaller black holes had formed at the centers of some of the first protogalaxies, the accretion of matter into the holes might have generated "mini quasars." Because these objects could have appeared soon after the first stars, they might have provided an additional source of light and ionizing radiation at early times.

Thus, a coherent picture of the universe's early history is emerging, although certain parts remain speculative. The formation of the first stars and protogalaxies began a process of cosmic evolution. Much evidence suggests that the period of most intense star formation, galaxy building and quasar activity occurred a few billion years after the big bang and that all these phenomena have continued at declining rates as the universe has aged. Most of the cosmic structure building has now shifted to larger scales as galaxies assemble into clusters.

In the coming years, researchers hope to learn more about the early stages of the story, when structures started developing on the smallest scales. Because the first stars were most likely very massive and bright, instruments such as the James Webb Space Telescope—the planned successor to the Hubble Space Telescope—might detect some of these ancient bodies. Then astronomers may be able to observe directly how a dark, featureless universe formed the brilliant panoply of objects that now give us light and life. SA

The formation of the first stars and protogalaxies BEGAN A PROCESS OF COSMIC EVOLUTION.

MORE TO EXPLORE

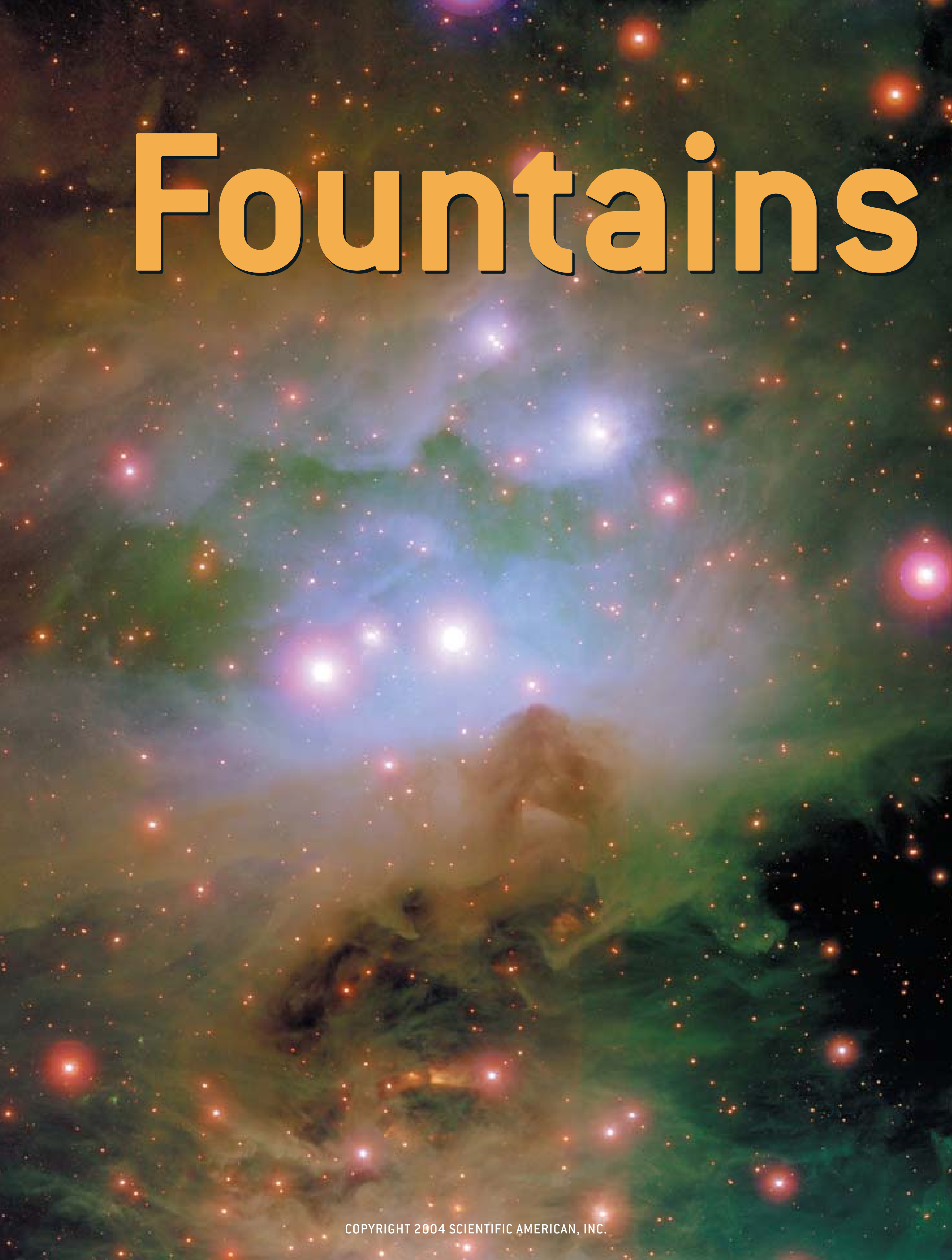
Before the Beginning: Our Universe and Others. Martin J. Rees. Perseus Books, 1998.

The First Sources of Light. Volker Bromm in *Publications of the Astronomical Society of the Pacific*, Vol. 116, pages 103–114; February 2004. Available at **www.arxiv.org/abs/astro-ph/0311292**

The First Stars. Volker Bromm and Richard B. Larson in *Annual Reviews of Astronomy and Astrophysics*, Vol. 42, pages 79–118; September 2004. Available at **www.arxiv.org/abs/astro-ph/0311019**

Graphics from computer simulations of the formation of the first luminous objects can be found at **http://cfa-www.harvard.edu/~vbromm/**

Fountains



# of Youth

Early Days in the Life of a Star

To make a star, gas and dust must fall inward. So why do astronomers see stuff streaming outward?

By Thomas P. Ray

Go out on a winter's night in the Northern Hemisphere and look due south around midnight. You will see the constellation of Orion the Hunter, probably the best-known group of stars after the Big Dipper. Just below Orion's Belt, which is clearly marked by three prominent stars in a line, is the Sword of Orion, and in the center of the sword is a faint fuzzy patch. This region, the Orion Nebula, is a giant stellar nursery embracing thousands of newborn stars.

Orion is a convenient place to study the birth of stars because it is relatively close by—a mere 1,500 light-years away—and has a good mix of low- and high-mass stars. It also contains a vast quantity of gas and dust in the form of a so-called molecular cloud. Such clouds are known to provide the raw material for new stars. What is now happening in Orion probably replicates what took place in our part of the galaxy five billion years ago, when the sun and its planets first came into being.

Understanding how stars and planets form is one of astronomy's quintessential subjects yet, until recently, one of the most poorly understood. Twenty years ago astronomers knew more about the first three minutes of the universe than they did about the first three billion days of our solar system. Only in the past decade have they started to get answers. Infant stars, it turns out, look like scaled-down versions of the heart of a quasar, with powerful jets of material flung outward by sweeping magnetic fields. These stellar fountains of youth not only make for spectacular pictures but also help to resolve paradoxes that have long dogged astronomers.

STELLAR BIRTHING GROUND in the Orion Nebula (*opposite page*) has given rise to hundreds of new stars. Surrounding it is an invisible but immense molecular cloud—a million suns' worth of gas and dust in a volume 300 light-years across. Young stars in Orion are swaddled in disks of material about the size of our solar system (*above*); around some, planets may even now be forming.

C. ROBERT O'DELL, MARK McCAUGHREAN AND JOHN BALLY *Hubble Space Telescope* AND NASA (*disks*); GARY BERNSTEIN *University of Pennsylvania*, MEGAN C. NOVICKI *University of Michigan* AND LUCENT TECHNOLOGIES (*nebula*)

The Journeywork of the Stars

THE THEORY OF HOW STARS and planets form has a venerable history. Just over 200 years ago French mathematician Pierre-Simon Laplace put forward the idea that the solar system was created from a spinning cloud of gas. He proposed that gravity pulled most of the gas to the center, thereby creating the sun. At the same time, some of the material, because of its spin, could not be absorbed by the young sun and instead settled into a disk. Eventually these dregs became the planets. According to modern numerical simulations of the process, once the spinning cloud starts to collapse, it proceeds quickly to the formation of one or more stars, a protoplanetary disk, and a leftover envelope of gas (individual atoms and molecules) and dust (very large clumps of atoms).

Laplace's model was not universally accepted. The uncertainty was mainly observational: testing the model was well beyond the astronomical capabilities of, say, 30 years ago, for two reasons. First, the leftover cloud of gas and dust blocks our view of the very region that must be studied. Second, protoplanetary disks subtend minute angles on the sky: if the distance between the sun and Pluto (six billion kilometers) is

FROM DUST TO A STAR

A star begins to coalesce when a disturbance, such as a nearby supernova explosion, causes a cloud of gas and dust to collapse.

Gas and dust clump at the center, surrounded by an envelope of material and a swirling disk. Magnetic forces direct jets along the axis.

Material continues to rain onto the disk. Roughly a tenth of it streams out in an uneven flow, shoving aside ambient gas.

representative of the scale of the disks, conventional ground-based telescopes can resolve them to a distance of only 200 light-years, less than halfway to the nearest star-forming region. Simply building bigger telescopes without special instrumentation does not help, because the blurring of detail occurs in the atmosphere.

Theoretical problems also stymied astronomers. Sunlike stars at the youthful age of 100,000 years rotate once every few days and are four or five times bigger than the mature sun. As such stars contract, they should spin faster, just like ice skaters pulling in their arms. Yet the sun has evidently slowed down, currently taking a month to rotate once. Something must have drained away its angular momentum. But what?

Another puzzle is how molecular clouds survive so long. Gravity is trying to force them to collapse, and without support they should implode within about a million years. But clouds seem to have endured for a few tens of millions of years. What holds them up? Thermal pressure is woefully inadequate because the clouds are far too cold, just 10 or 20 kelvins. Turbulence might do the trick, but what would generate it? In giant molecular clouds such as Orion, winds and shock waves produced by embedded massive stars would stir things up, but many smaller, sedate clouds have no massive stars.

The first observational obstacle yielded in the late 1970s, when astronomers began to observe star-forming regions at wavelengths that penetrate the dust shroud. Studying regions such as the Orion molecular cloud at millimeter wavelengths—a previously unexplored part of the spectrum—astronomers identified dense, cold clumps typically measuring a light-year across. Such clumps, known as molecular cores, contain as much as a few suns' worth of gas and quickly became identified with Laplace's spinning clouds.

As is often the case in astronomy, new mysteries immediately emerged. Although a few of the molecular cores seem to be in the process of collapsing, most of them seemed stabilized by means that are not entirely understood. What triggers their eventual collapse is equally uncertain, but it may involve some outside push from, for example, a nearby supernova explosion. Or turbulence may simply die away, letting gravity take over. The biggest conundrum concerns the direction in which material is moving. According to Laplace's hypothesis, stars arise from gravitational accretion, so astronomers expected to see signs of gas plummeting toward the cores.

To their astonishment, they discovered that gas, in the form of molecules, is actually moving outward. Usually two giant lobes of molecular gas were found lying on either side of a young star. These lobes, typically a few light-years in length, have masses similar to or even larger than that of the young star itself, and they move apart at speeds of tens of kilometers per second.

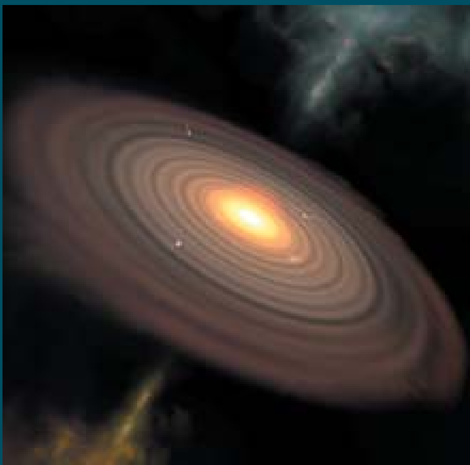
Jetting from the Crib

THE MOLECULAR LOBES bear a strange resemblance to the vastly larger lobes of hot plasma seen near active galaxies such as quasars. Astronomers had known for years that jets produce these lobes. Squirting outward at velocities close to the speed of light, jets from active galaxies can stretch for millions of light-years. Might a miniature version of these jets also drive the molecular lobes in star-forming regions?

This idea harked back to a discovery in the early 1950s by astronomers George H. Herbig and Guillermo Haro. Herbig, then working at Lick Observatory in northern California, and Haro, at Tonantzintla Observatory in Mexico, independently found some faint fuzzy

THE AUTHOR

THOMAS P. RAY is astronomy professor at the Dublin Institute for Advanced Studies, having also worked at the University of Sussex in England and the Max Planck Institute for Astronomy in Heidelberg, Germany. Ray has been the principal or co-investigator on numerous Hubble observations of jets from young stars and is helping design the James Webb Space Telescope, Hubble's successor. His other interests include quasars, comets, archaeoastronomy (the study of sites such as Stonehenge), sailing and Guinness.



Disk material agglomerates into planetesimals. The envelope and the jets dissipate. By this point, one million years have passed.

The high pressure and temperature at the center of the star trigger nuclear fusion. The planetesimals have assembled into planets.

patches in Orion. Now known as Herbig-Haro objects, these small clouds were initially thought to be specific sites of star formation. (Some popular astronomy books still repeat this error.) In 1975, however, Richard D. Schwartz, then at the University of California at Santa Cruz, realized that the spectrum of a Herbig-Haro object closely resembles that of the material left over from a supernova. From the Doppler shifting of the spectral lines, he found that Herbig-Haro objects are moving at speeds up to a few hundred kilometers per second.

That is slower than the motion of a typical supernova remnant, but Schwartz reckoned that the principles are the same—namely, that the Herbig-Haro objects are heated gas flowing away from a star. The heat, as in supernova remnants, comes from the motion of the gas itself; shock waves convert some of the kinetic energy of motion into thermal energy and then into radiation. Schwartz's idea gained further support when astronomers looked at photographs of Herbig-Haro objects taken a number of years apart. They were indeed moving. By extrapolating backward in time, astronomers deduced their source. Invariably it was a star only a few hundred thousand years old.

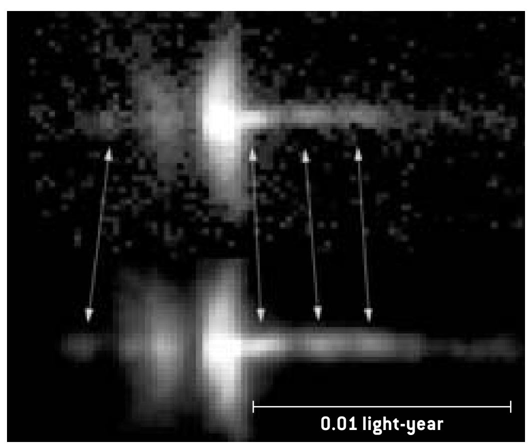
Verification of this connection came with another technological revolution: the charge-coupled device (CCD), the light-sensitive chip found in camcorders and digital cameras. For astronomers, CCDs offer greater sensitivity and contrast than the traditional photographic plates. In 1983 Reinhard Mundt and Josef Fried of the Max Planck Institute for Astronomy in Heidelberg, Germany, made the first CCD observations of stellar jets. Subsequent work by Mundt, Bo Reipurth of the European Southern Observatory in Santiago, Chile, and others (including me) showed that jets from young stars stretch for several light-years. These jets are closely related to Herbig-Haro objects. In fact, some such objects turned out to be nothing more than the brightest parts of jets. Others were discovered to be bow shocks caused by jets as they plow their way supersonically through ambient gas, like the shock wave that surrounds a bullet zinging through the air. The jets typically have a temperature of about 10,000 kelvins and contain 100 atoms per cubic centimeter—denser than their surroundings but still thinner by a factor of 10,000 than the best vacuum in labs on Earth. Near the star the jets are narrow, opening with an angle of a few degrees, but farther from the star they fan out, reaching a diameter wider than Pluto's orbit.

Out of the Way

HOW ARE THE JETS and Herbig-Haro objects, which are mostly made up of atoms and ions, related to the molecular flows? The leading explanation is that a molecular lobe consists of ambient gas that got in the way of the jet and was accelerated.

None of these observations reach the heart of the matter, however: the disk around the nascent star. Astronomers had long been gathering circumstantial evidence for disks. In the early 1980s the Infrared Astronomical Satellite discovered that many new stars had excess infrared radiation over and above what should be produced by the star alone. Warm dust in a disk seemed the most likely source. Around the same time, millimeter-wave telescopes began to measure the mass of gas and dust around these stars, typically finding 0.01 to 0.1 solar mass—just the right amount of material needed to form planetary systems. In the mid-1980s Edward B. Churchwell of the University of Wisconsin–Madison and his colleagues observed the Orion Nebula at radio wavelengths. They found sources comparable in size to our own solar system and suggested that the sources were clouds of hot gas that had evaporated from a disk.

Sighting the disks themselves, however, ran up against the second observational obstacle: their comparatively small size. For that, astronomers had to await the clarity afforded by the Hubble Space Telescope and by ground-based instruments equipped with adaptive optics. In 1993 C. Robert O'Dell of Rice University and his collaborators ob-



TIME-LAPSE IMAGES of a hatchling star, Herbig-Haro 30, taken a year apart show pockets of gas moving away from the center. These jets are clearly perpendicular to the dark disk that hides the star.

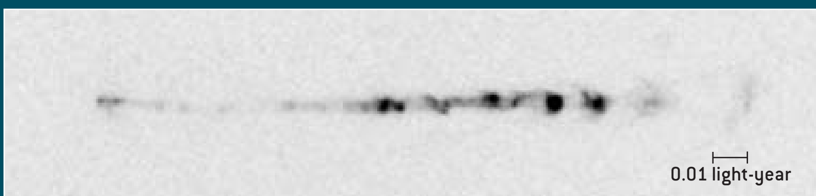
ALFRED T. KAMAJIAN (*illustrations*); CHRISTOPHER J. BURROWS *Space Telescope Science Institute* AND NASA (*photograph*)

JET ACTION

Mysterious though their detailed mechanisms may be, jets always involve the same basic physical process: a balance of power between gravity, magnetic fields and angular momentum. Gravity tries to pull matter toward the center of mass, but because of centrifugal forces, the best it can do is gather material into a swirling disk. Narrow streams of gas shoot out along magnetic field lines, the direction in which matter can most easily move. The escaping matter carries away angular momentum, thereby allowing less footloose matter to settle inward.

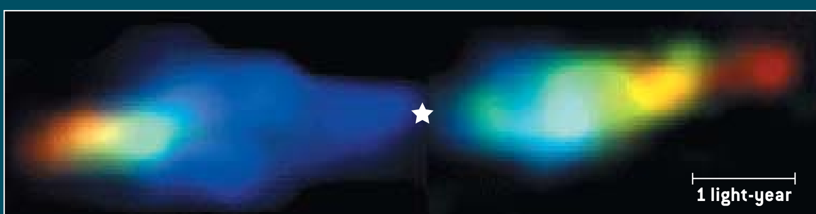
WILLIAM B. SPARKS *Space Telescope Science Institute*

In the core of the active galaxy Messier 87, the driving force is thought to be a black hole a billion times more massive than the sun.



THOMAS P. RAY

On a much smaller scale, a newborn star whips up and sprays out a current of gas known as Herbig-Haro 34. The jet may push ambient molecular gas outward.



MICHEL FICH *University of Waterloo* AND GERALD MORIARTY-SCHIEVEN *Joint Astronomy Center, Hawaii*

Observing the star-forming region NGC 2264 at millimeter wavelengths, astronomers see two lobes of molecular gas moving at tens of kilometers per second. Red indicates the fastest velocities, violet the slowest.

JON A. MORSE *Space Telescope Science Institute* AND NASA

Complex jet patterns, as evident in Herbig-Haro 47, can arise because of variations in the outflow rate and the gravitational effect of companion stars.

served Orion with Hubble and finally saw the disks that Laplace had predicted [*see illustrations on page 13*]. Their material, where it was buffeted by the intense radiation and winds from nearby massive stars, was seen to be evaporating. O'Dell christened these disks "proplyds" for protoplanetary disks. The name may actually be a misnomer because some disks will evaporate within a million years, probably before planets can form. But similar disks in milder environments should indeed survive long enough to give birth to planets.

With the discovery of all the basic components of a modern version of Laplace's theory—spinning clouds, outflows, disks—astronomers could begin to study the relationships among them. My colleagues and I, along with another group led by Christopher J. Burrows, then at the Space Telescope Science Institute, turned the Hubble telescope on Herbig-Haro 30, which consists of a pair of oppositely directed jets. To our surprise, the images revealed two small cusp-shaped nebulae where the source of the jets should have been. Cutting across the nebulae was a dark band. It soon became clear that we were looking at a disk perpendicular to the jets. As seen from our edge-on view, the disk obscures the central star. The nebulae are dust clouds illuminated by starlight. Jets stream outward, culminating in the Herbig-Haro objects. The jigsaw puzzle of star formation was coming together.

In active galaxies, disks are crucial to the formation of jets. But how does this process work for an embryonic star? An intriguing coincidence has provided a crucial clue. All the jets and flows located near Herbig-Haro 30, with one odd exception, have roughly the same orientation. In fact, they are aligned with the magnetic field of the parent cloud. This seems to support ingenious suggestions—made by Ralph E. Pudritz and Colin A. Norman, both then at the University of Cambridge, and by Frank H. Shu of the University of California at Berkeley—for how magnetic fields could drive an outflow from a young star.

Astronomy abounds with examples of magnetic fields guiding ionized gas. For example, auroras are caused by charged particles that stream down Earth's magnetic field lines and hit the upper atmosphere. In the same way, ionized particles from a circumstellar disk could attach themselves to the field lines of either the disk or the star. Because the disk is spinning, the particles would experience a centrifugal force and would thus be flung out along the field lines. More matter would flow in to replace what was lost, and so the process would continue. Although most of the matter would end up being accreted by the star, some 10 percent might be ejected. In computer simulations the process may proceed in fits and starts, which would account for the knotty structures seen in many jets.

Nebulous No More

THE REALIZATION that jets are integral to star formation may solve several of the theoretical puzzles. As particles travel outward, they carry angular momentum away from their source—which would partially explain why mature stars such as the sun rotate so slowly. Jets may also churn up the surrounding cloud, supplying the necessary turbulent support to slow down its collapse.

At the same time, many questions remain. For example, only about 50 percent of evolved young stars are found to have disks. The others may have had disks in the past, but these disks could have already coalesced into planets. Observers, however, have been unable to confirm this. Another problem in star formation is the distribution of stellar masses. Why is the ratio of high- to low-mass stars pretty much the same irrespective of location in the galaxy? This ratio seems to be a fundamental property of the way molecular clouds fragment, but for unknown reasons. Researchers know little about the early life of high-mass stars—partly because they are rarer, partly because they evolve faster and are difficult to catch in the act of forming. We do know, however, that some high-mass young stars are surrounded by disks and produce jets.

With these caveats, astronomers can now sketch out nature's recipe for stars. They form in interstellar clouds that consist largely of the ashes of earlier generations of stars. The dust was manufactured in the cool winds and outer atmospheres of stars as they approached the ends of their lives. The clouds are also laced with heavy elements such as iron and oxygen that were forged deep in the nuclear furnaces of bygone stars. Magnetic fields or turbulent motions hold up the clouds, but eventually they collapse under their own weight, perhaps because the magnetic fields leak away, the turbulence dissipates or a supernova goes off nearby. As the material falls in, the clouds fragment into cloudlets, each of which settles into a primitive star system. In massive molecular cores, such as those that gave rise to the cluster in the Orion Nebula, these systems are spaced every few light-weeks (as opposed to light-years) apart. Most stars in the galaxy, including the sun, probably formed in such clusters.

Jets carry away angular momentum and allow the accretion to continue. Our sun must once have had narrow jets that stretched for several light-years. What turned them off is not certain. The store of infalling material may simply have run out. Some of it may have been driven away by the outflows; if so, the jets may have served to limit the sun's final mass. Around the same time, large dust grains were beginning to stick together to form planetesimals, the building blocks of the planets. The planetesimals swept up any remaining gas, further choking off the jets. The outflows from the sun and its stellar contemporaries blew away the leftover gas and dust that threaded the space between them. This weakened the gravitational glue that bound them together, and over a few million years the stars dispersed. Today the nearest star to the sun is about four light-years away.

Two centuries after Laplace put forward his nebular hypothesis, the pieces are beginning to fall into place. Studies of young stars suggest not only that planet formation is going on today but that planets are very common throughout our own and other galaxies. SA

MORE TO EXPLORE

Young Stars and Their Surroundings. C. Robert O'Dell and Steven V. W. Beckwith in *Science,* Vol. 276, No. 5317, pages 1355–1359; May 30, 1997.

Jets: A Star Formation Perspective. Thomas P. Ray in *Astrophysical Jets, Open Problems.* Edited by Silvano Massaglia and Gianluigi Bodo. Gordon and Breach Scientific Publishers, 1998.

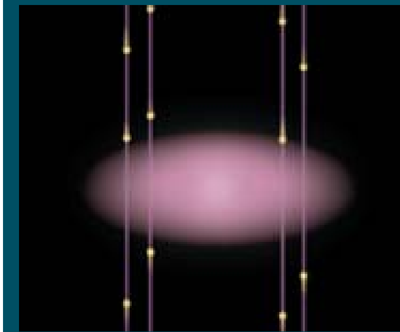
The Origin of Stars and Planetary Systems. Edited by Charles J. Lada and Nikolaos D. Kylafis. Kluwer Academic Publishers, 1999.

Protostars and Planets IV. Edited by Vince Mannings, Alan P. Boss and Sara S. Russell. University of Arizona Press, 2000.

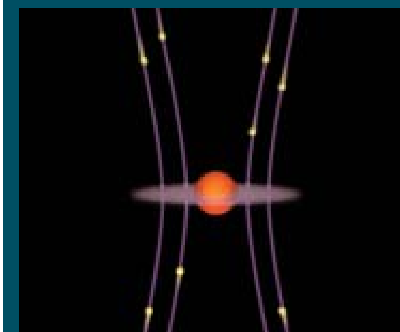
For links to World Wide Web sites, visit **astro.caltech.edu/~lah/starformation.html**

ALFRED T. KAMAJIAN

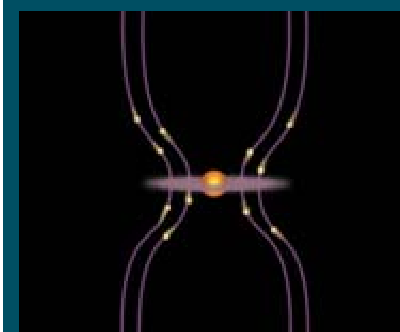
MAGNETIC, PERIPATETIC



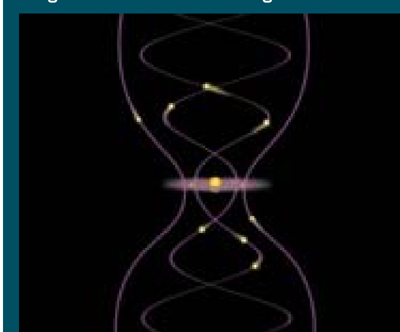
The generation of jets may begin when material—a mixture of ions, atoms, molecules and dust—rains onto the circumstellar disk along magnetic field lines.



As the disk contracts under gravity, the lines (which are frozen into the material) are pulled in, taking on an hourglass shape.



When the field lines are bent to an angle of 30 degrees from the perpendicular, centrifugal force overcomes gravity and flings material outward along the lines.



The inertia of the swirling material twists the field lines into a helix, which helps to channel the outward-flowing material in a vertical direction.

Companions to Young Stars

The surprising finding that even the youngest stars commonly exist in sets of two or three has revised thinking about the birth of star systems

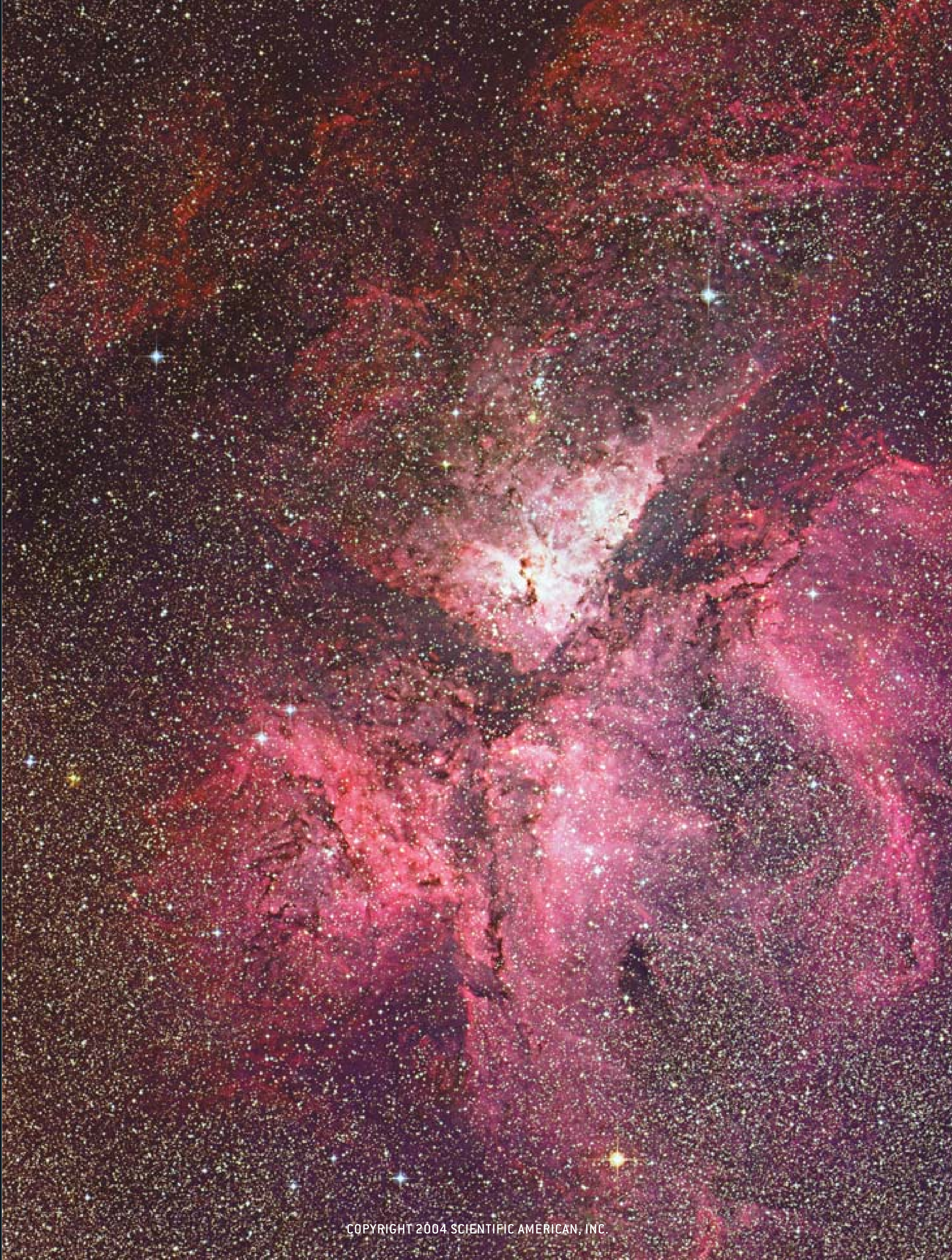
By Alan P. Boss

A minor revolution in astronomy occurred on April 6, 1992. It did not take place at a mountaintop observatory but happened at an unlikely location—the Callaway Gardens Inn on Georgia's Pine Mountain (elevation: 820 feet). Astronomers had gathered there for an international meeting on the normally slow-paced research topic of double stars, a field where discoveries often require decades to allow for many of these systems to complete their orbits. While azaleas flowered outside in the spring rain, astronomers inside presented results pointing to the startling conclusion that young stars frequently have stellar companions. This realization was the product of painstaking observations by many different people using a host of clever techniques and new devices. That morning in Georgia, the separate works of these numerous researchers appeared magically to dovetail.

RHO OPHIUCHUS molecular cloud harbors colorful reflection nebulae and numerous stars in the process of formation. Because these stellar nurseries lie relatively close to Earth, observations of them can provide important insights into the birth of double stars.

DAVID MALIN *ROE/Anglo-Australian Observatory*

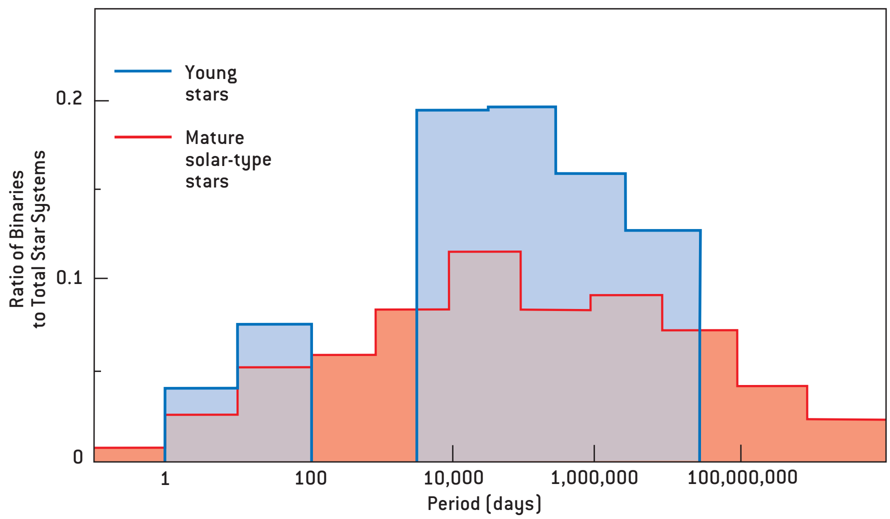
Updated from the October 1995 issue



The finding that binary systems are at least as common for young stars as for older ones might seem reasonable enough, but to astronomers it came as a shock. Most notions of double star formation had predicted that stellar companions are produced or captured well after a star has formed; hence, the youngest stars would be expected to exist singly in space. Such theories no longer bear weight. There remains, however, at least one idea for the formation of double stars that holds up to the recent observations. It may be the sole explanation for why binary star systems are so abundant in the universe.

The sun, which is a mature star, has no known stellar companions, even though most stars of its age are found in groups of two or more. In 1984 Richard A. Muller of Lawrence Berkeley National Laboratory and his colleagues hypothesized that the sun is not truly a single star but that it has a distant companion orbiting it with a period of about 30 million years. He reasoned that gravitational forces from this unseen neighbor could disturb material circling in the outermost reaches of the solar system, sending a shower of comets toward the inner planets every time the star neared. Muller suggested that this effect might explain periodic mass extinctions: comets generated by the sun's companion would hit the earth every 30 million years or so and—as with the demise of the dinosaurs—would have wiped out much of life on our planet. Because its approach would have sparked such widespread destruction, Muller called the unseen star "Nemesis."

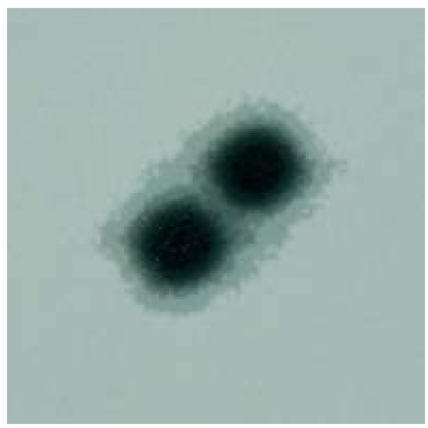
Most scientists have not accepted Muller's interesting idea. For one, the closest known stars (the Alpha Centauri triple star system, at a distance of 4.2 light-years) are much too far away to be bound to the sun by gravity. In fact, there is no astronomical evidence that the sun is anything other than a single star whose largest companion (Jupiter) is one thousandth the mass of the sun itself. But living on a planet in orbit around a solitary sun gives us a distorted view of the cosmos; we tend to think that single stars are the norm and that double stars must be somewhat odd. For stars like the sun, this turns out to be far from true.



YOUNG BINARIES are at least as ubiquitous as mature double stars. For all orbital periods yet determined, young doubles found in star-forming regions (*blue*) are even more common than solar-type binaries that have been surveyed in the sun's neighborhood (*red*).

Doubles, Anyone?

IN 1990 Antoine Duquennoy and Michel Mayor of the Geneva Observatory completed an exhaustive, decade-long survey of nearby binary stars. They considered every star in the sun's "G-dwarf" class within 72 light-years, a sample containing 164 primary stars that are thought to be representative of the disk of our galaxy. Duquennoy and Mayor found that only about one third of these systems could be considered true single stars; two thirds had companions more massive than one hundredth the mass of the sun, or about 10 Jupiters.



GLASS-I proved to be a young double star when imaged by an infrared camera at a wavelength of 0.9 micron.

Binary star systems have widely variable characteristics. Stars of some double G-dwarf systems may be nearly touching one another; others can be as far apart as a third of a light-year. Those in contact may circle each other in less than a day, whereas the most widely sep-

Overview/*Twin Suns*

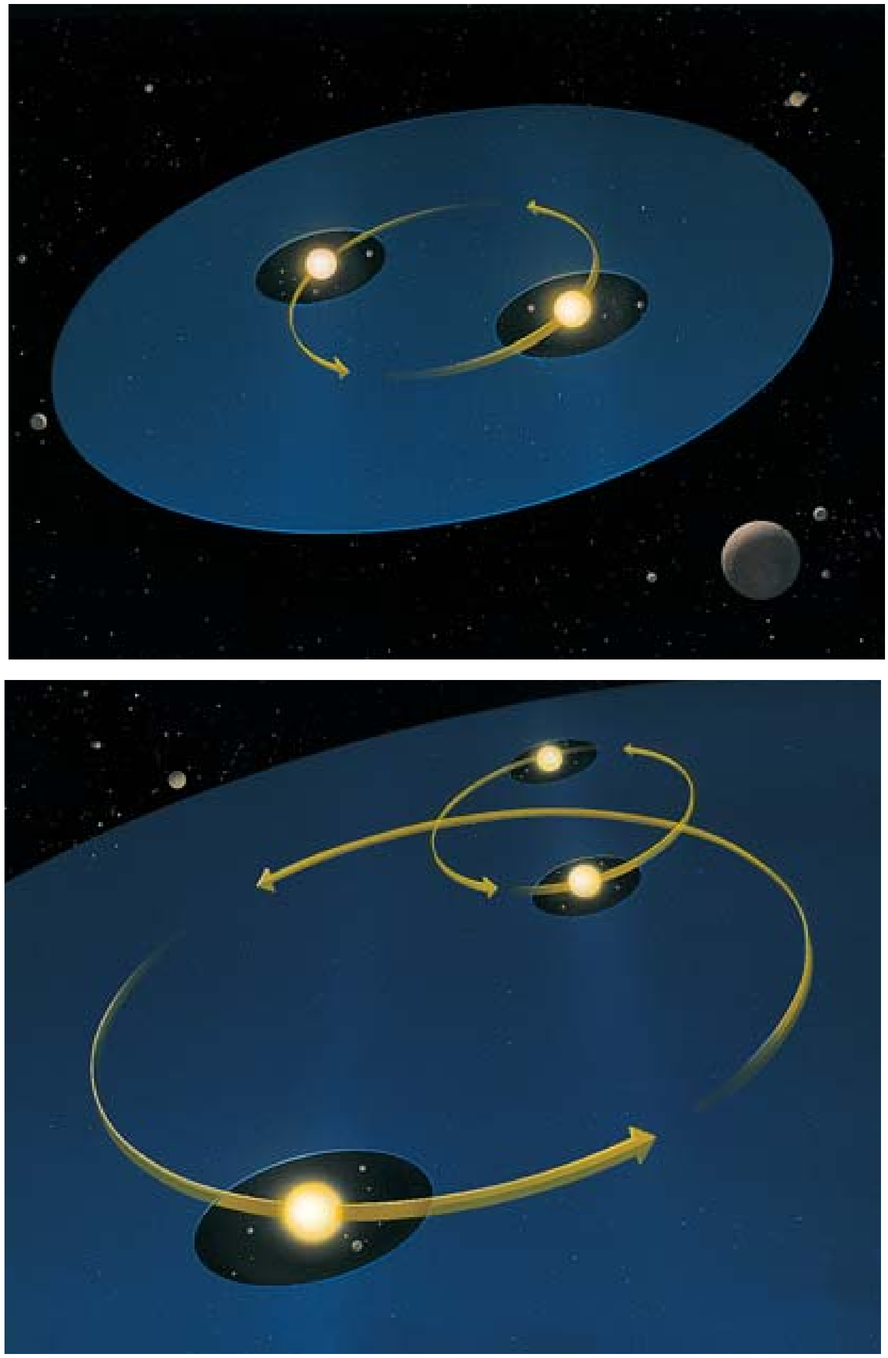
- In our galaxy, two thirds of mature stars have one or more companions. But astronomers have been shocked to find that binary systems are just as common for young stars.
- The collapse of protostellar gas clouds into binary systems occurs quickly—over a few hundred thousand years during a stellar lifetime lasting several billion years.
- Whether binary, triple or quadruple systems eventually form depends on the initial shape of the cloud and the precise amount of thermal and rotational energy in it. Planets may readily form around these systems, too.

CAREY BALLARD (*top*); WOLFGANG BRANDNER *University of Würzburg* (*bottom*) arated double stars may take tens of millions of years to complete a single orbit. Duquennoy and Mayor showed that triple and quadruple G-dwarf stars are considerably rarer than double stars. They counted 62 distinct doubles, seven triples and two quadruple groupings. They further determined that each of the triple and quadruple sets had a hierarchical structure, composed of a relatively close double orbited by either a more distant single star (forming a triple system) or another close double star (forming a quadruple system). The separation between distant pairs needs to be at least five times the gap of the close doubles for the group to survive for long. Arrangements having smaller separations are named Trapezium systems, after a young quadruple system in the Orion nebula. These arrangements are orbitally unstable—they will eventually fly apart. For instance, if the three stars of a triple system come close enough together, they will tend to eject the star of lowest mass, leaving behind a stable pair.

Double stars thus seem to be the rule rather than the exception. This conclusion does not, however, mean that planets must be rare. A planet could travel around a double star system provided that it circles either near one of the two stars or far away from both of them. Imagine living on such a world orbiting at a safe distance from a tightly bound binary, where the two stars complete an orbit every few days. The daytime sky would contain a pair of suns separated by a small distance. Sunrises and sunsets would be fascinating to watch as first one and then the other glowing orb crossed the horizon. Other strange celestial configurations might also occur. If, for example, the planet orbited in the same plane as did two stars of equal mass, the two suns periodically would appear to merge as they eclipsed each other, briefly halving the amount of combined sunlight reaching the planet.

Stellar Nurseries

THE SUN FORMED about 4.6 billion years ago and has about five billion years remaining of its so-called main-sequence lifetime. After it reaches the end of its

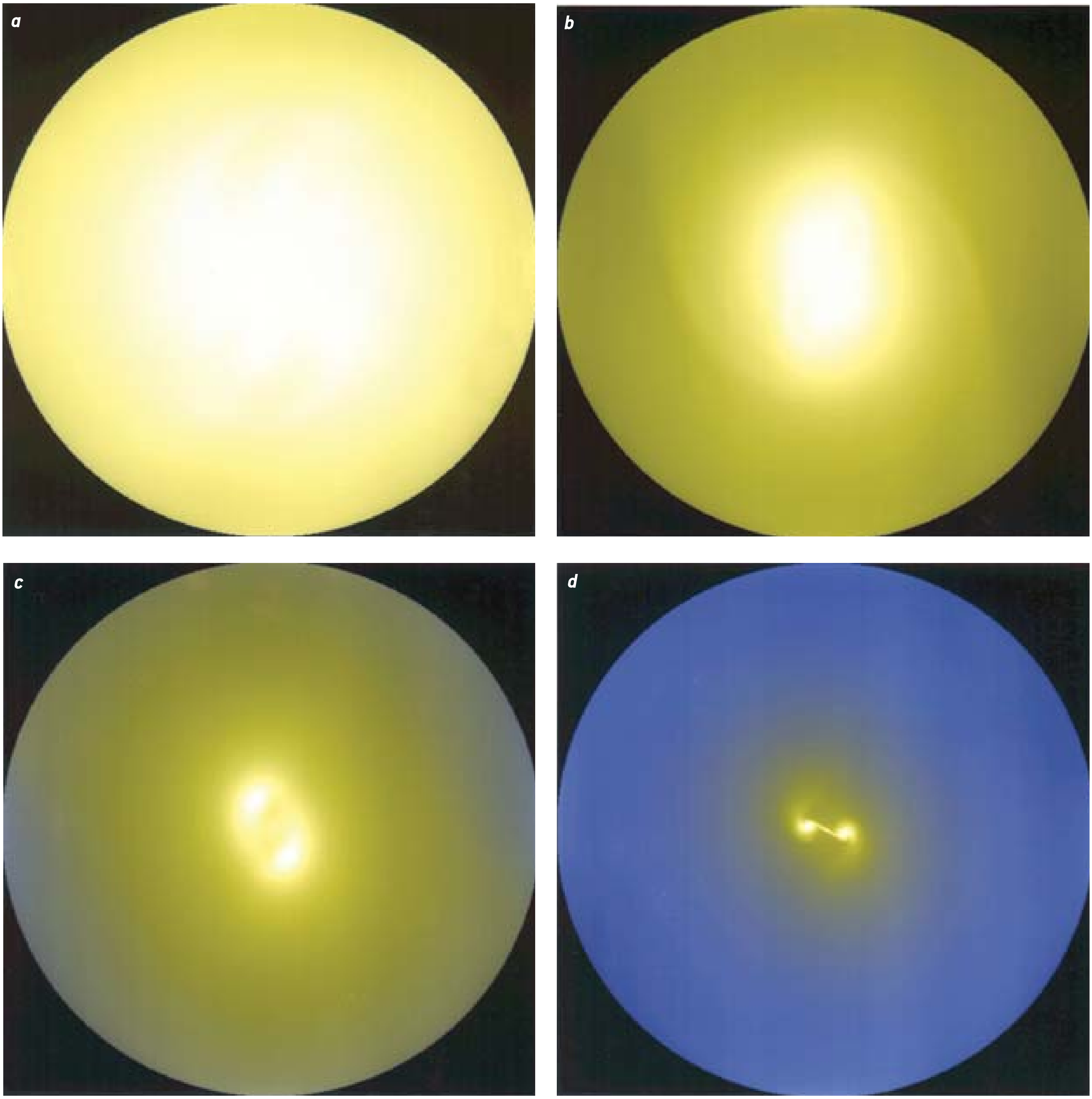


PLANETS in double or triple star systems would be excluded from special regions (*blue*) within which they could not orbit stably. Inside this zone, a planet would eventually be tossed out by gravitational interactions. For a double system (*top*), planets could reside either near each of the stars or far from them both. For a triple system (*bottom*), planets could orbit close to either of the paired stars, in a more extended region around the single member or far from all three.

ALFRED T. KAMAJIAN

THE AUTHOR

ALAN P. BOSS began modeling the formation of stellar and planetary systems as a physics graduate student at the University of California, Santa Barbara, where he received a doctorate in 1979. After two years at the NASA Ames Research Center, he joined the department of terrestrial magnetism at the Carnegie Institution of Washington (where, despite its name, terrestrial magnetism has not been studied for decades). Boss chairs the International Astronomical Union committee that maintains the organization's list of extrasolar planets (available at www.ciw.edu/boss/IAU/div3/wgesp/planets.html).



BINARY STARS can form from the collapse of a molecular cloud. Computer simulation of a slightly prolate, magnetically supported cloud (*a*) shows that the cloud becomes increasingly prolate once magnetic support is lost and the collapse begins (*b*), leading to an intermediate bar-shaped entity (*c*) that fragments into a binary protostar (*d*).

ALAN P. BOSS

main sequence, it will expand to become a red giant that will engulf the inner planets. This configuration will be somewhat akin to one that occurred early in the sun's history, when it extended far beyond its present radius. At that time, before it had contracted to its current size, the sun was similar to the T Tauri class of stars that can be seen in those regions of our galaxy where stars are now forming. During its T Tauri stage, the sun's radius was about four times greater than its present measurement of some 700,000 kilometers. And even earlier still, the protosun must have extended out to about 1.5 billion kilometers, or 10 times the distance between the earth and the sun (that span, 150 million kilometers, is known as an astronomical unit, or AU).

Present-day T Tauri stars offer astronomers an opportunity to learn what the sun was like early in its evolution. The nearest T Tauri stars are in two locations, known as the the Taurus molecular cloud and the Rho Ophiuchus molecular cloud, both about 460 light-years from Earth. The fact that young

EARTHLINGS HAVE A DISTORTED VIEW OF the cosmos. Double stars are not odd; THEY ARE ALMOST AS COMMON AS SINGLES.

stars are always embedded in such dusty concentrations of gas gives us convincing testimony on the nature of their origin—stars are born from the contraction and collapse of the dense cores of molecular hydrogen clouds.

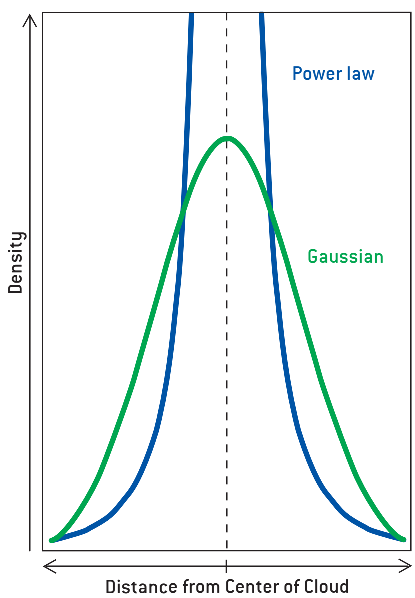
Because young stars are typically enshrouded by dust, astronomers usually have difficulty viewing them in visible light, no matter how powerful the telescope. But these sites can be detected readily using infrared wavelengths that are characteristic of the emission from heated dust grains surrounding the nearby star. Progress in understanding the formation of stars has thus been dependent to a large extent on the development of detectors capable of sensing infrared radiation. At the 1992 meeting in Georgia, the first results were presented from several different infrared surveys specifically designed to detect companions to the T Tauri stars in Taurus and Ophiuchus.

Subsequently, Andrea M. Ghez of the University of California at Los Angeles and her colleagues Gerry F. Neugebauer and Keith Matthews, both at the California Institute of Technology, used a new indium antimony array camera on the five-meter Hale telescope to photograph the regions around known T Tauri stars at the near-infrared wavelength of 2.2 microns. (Visible light has a wavelength between about 0.4 and 0.7 micron.) Using a so-called speckle imaging technique to minimize the noise introduced by fluctuations in Earth's atmosphere above the telescope, Ghez and her colleagues found that almost half of the 70 T Tauri stars in their sample showed stellar companions. For the limited range of separations considered, about 10 to 400 AU, this study indicated that for the youngest systems, binaries are twice as common as for main-sequence stars.

Christoph Leinert of the Max Planck Institute for Astronomy in Heidelberg also conducted a near-infrared speckle imaging survey. He found that 43 of the 106 T Tauri stars examined had nearby companions, again implying that binaries were much more common in these stars than in G-dwarf stars like our sun.

Hans Zinnecker, now at Potsdam Astrophysics Institute, Wolfgang Brandner, now at European Southern Observatory, and Bo Reipurth, now at the University of Hawaii Institute for Astronomy, used a high-resolution digital camera with the European New Technology Telescope to image 160 T Tauri stars at an infrared wavelength of one micron. After analyzing the data, they uncovered 28 companions lying from 100 to 1,500 AU from the T Tauri stars, about a third more than circle around older, solar-type stars in that distance range.

Michal J. Simon of the Stony Brook University, N.Y., along with Wen Ping Chen of the National Central University in Taiwan and their colleagues, reported a novel way to find young double stars. When the moon passes over, or occults, a distant star system, careful monitoring of the light received can reveal the presence of two or more sources, as first one and then another star slips behind the sharp edge of the lunar face. Simon and Chen's measurements detected companions much closer to T Tauri stars than was possible with infrared imaging. Their work again showed that a large fraction of the entities are binaries. Robert D. Mathieu of the University of Wisconsin–Madison employed a more traditional means for detecting close double stars, the same method as that used by Duquennoy and Mayor. Mathieu used spectroscopic measurements of the periodic Doppler shift to show that some T Tauri stars have companions located within 1 AU. Once more, closely spaced binaries proved more common for young T Tauri systems than for solar-type stars.



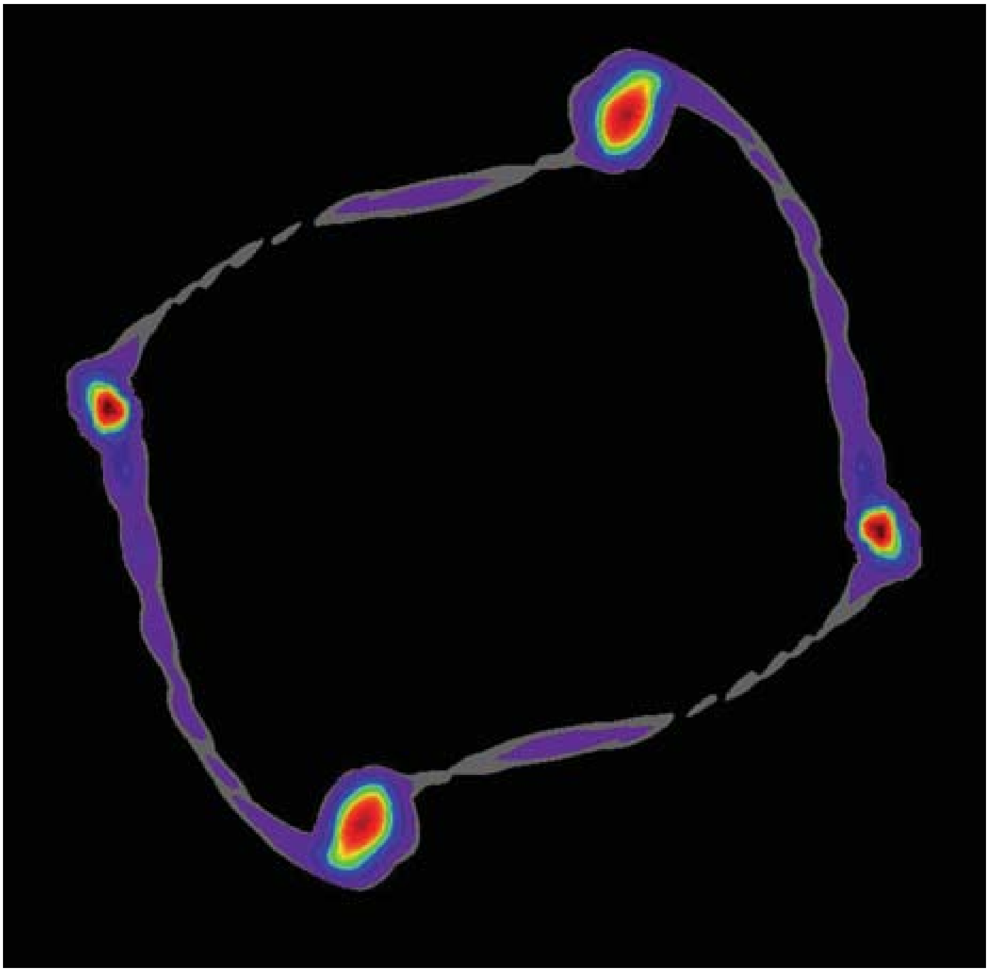
DENSITY in star-forming clouds was thought to follow a power law that concentrates mass tightly, but better data now suggest it matches a bell-shaped Gaussian curve that allows binaries to form.

CAREY BALLARD

Search for a Theory

HOW DID ALL THESE stellar companions come to be? Why did they form so abundantly and so early in their evolution? The wealth of observations of young stars requires that binary stars must form well before even their pre-main-sequence (T Tauri) phase. Moreover, the finding that binaries are so common demands that the mechanism generating them—whatever it is—must be very efficient.

In principle, a double star system could arise from two stars that pass close enough together that one forces the other into a stable orbit. The celestial mechanics of such an event, however, requires the intervention of a third object to remove the excess energy of motion



COLLAPSE of an initially oblate-shaped, magnetically supported cloud leads to fragmentation into a multiple protostar system. Here four protostars have formed with separations on the order of a few AU. This system is highly unstable and will probably decay into a close binary system and two ejected single stars.

ALAN P. BOSS

between the two stars and ultimately leave them trapped in a gravitationally bound system. Nevertheless, such three-body encounters are too rare to account for very many binary stars. Cathy J. Clarke and James E. Pringle of the University of Cambridge studied a more likely way that companion stars might have paired up. They investigated the gravitational coupling that occurs between two young stars that still had flattened disks of dust and gas surrounding them. That geometry would be far more common than three-body encounters and could, in theory, remove enough energy from the stars' motions. But in their analysis they found that such interactions are much more likely to end up ripping apart the circumstellar disks than to result in one star's neatly orbiting with the other. So this embellishment seems to help little in explaining the extensive existence of binary star systems.

Failure of the capture mechanism has forced most astronomers to think about processes that might form binary stars more directly. Consideration of this notion goes back over a century. In 1883 Lord Kelvin proposed that double stars result from "rotational fission." Based on studies of the stability of bodies in rapid rotation, Kelvin suggested that as a star contracted, it would spin faster and faster until it broke up into a binary star.

Astronomers now know that pre-main-sequence stars contract considerably as they approach the hydrogen-burning main sequence, but T Tauri stars do not rotate fast enough to become unstable. Furthermore, Kelvin's fissioning would act too late to explain the frequency of binaries among young stars. Richard H. Durisen of Indiana University and his colleagues showed that fission fails on theoretical grounds as well—a reasonable calculation of this instability shows that the ejected matter would end up as trailing spiral arms of gas rather than as a separate cohesive star.

In contrast to the century-old fission theory, the more recent idea for creating binary stars is called fragmentation. This concept supposes that binary stars are born during a phase when dense molecular clouds collapse under their own gravity and become protostars. The obscuring gas and dust then clear away, and a newly formed binary star (of the T Tauri class) emerges. In contrast to older theories of the birth of binary systems, fragmentation fully agrees with the latest observations of young stars.

The protostellar collapse that enables fragmentation occurs relatively suddenly in the scale of a several-billion-year stellar lifetime; the event takes place in a few hundred thousand years. This violent transformation of a diffuse cloud into a compact star thus offers a special opportunity for a single object to break into several distinct members. Astrophysicists have identified two mechanisms that might operate on the transition. Very cold clouds can fragment directly into binary systems, whereas warmer clouds with substantial rotation can first settle into thin disks and then later break up as they gain more mass or become progressively flattened.

Cloudy Ideas

A KEY OBJECTION to the fragmentation theory involved the distribution of matter in protostellar clouds. It was previously thought that this material was distributed according to a so-called power law. That is, there would be an extremely high concentration of material near the center of the cloud and a rapid decrease in density with increasing distance. This objection was apparently removed, however, by high-resolution radio observations made in 1994 using submillimeter wavelengths. Derek Ward-Thompson, now at the University of Cardiff in Wales, and his colleagues determined the distribution of material inside several precollapse clouds. They found that the density follows a classic Gaussian (bell-shaped) distribution gravitational contraction. The conditions appear to be nothing out of the ordinary for the clouds found in stellar nurseries.

Whether a binary, triple or quadruple system eventually forms depends on many details, including the three-dimensional shape of the original cloud, how lumpy it is, and the precise amount of thermal and rotational energy available. In general, prolate, or football-shaped, clouds tend to form bars that fragment into binary systems, whereas

FOOTBALL-SHAPED GAS CLOUDS TEND TO FORM binary systems. Pancake-shaped clouds FRAGMENT INTO SEVERAL STARS.

more oblate, or pancake-shaped, clouds flatten to disks that later fragment into several members.

The collapse is thought to occur in two separate steps. The first phase generates protostars with a radius on the order of 1 AU or so. These bodies then undergo a second collapse to form the final protostars that have stellar dimensions. Fragmentation is believed to be possible only during the first collapse phase and appears to be capable of generating most of the entire range of separations observed in young binary stars. The very closest systems appear to be the result of the orbital decay of multiple protostar systems.

rather than a power law. Hence, matter would be less tightly concentrated toward a central point when the star system began to form.

Elizabeth A. Myhill, then at the University of California at Los Angeles, and I had shown separately that the high density at the center of a cloud that follows a power law makes it almost impossible for a second or third star to coalesce. It proves much easier for fragmentation to occur with an initial Gaussian distribution.

Astrophysicists can predict whether multiple fragments will ultimately form by solving the set of equations that govern the flow of gas, dust and radiation in a protostellar cloud. The calculations are sufficiently complex to require accurate software and a powerful computer for their solution.

I began modeling the collapse of dense clouds with Gaussian density profiles in 1986 and found that fragmentation could readily occur provided certain conditions were met. As long as a Gaussian cloud has sufficient rotation to give the binary system the angular momentum it requires and the precollapse material is cold enough (less than 10 kelvins) to make its thermal energy less than about half its gravitational energy, the cloud will fragment during its

Dwarfs and Giants

WHAT ABOUT FINDING companions of even lower mass? Duquennoy and Mayor produced evidence that as many as 10 percent of solar-type stars are bound to brown dwarfs—that is, they have stellar companions with masses ranging from 0.01 to 0.08 times the mass of the sun. Brown dwarfs are too small to ignite hydrogen the way the sun does, but they could still be massive enough to burn deuterium soon after formation. After that, their radiation would cease, and they would become cool and extremely difficult to detect. Brown dwarf stars have been found to be commonplace in the solar neighborhood, though seldom as companions to sunlike stars. The best place to find a brown dwarf is in orbit around another brown dwarf.

The original publication of this article occurred at a propitious time. October 1995 began the era of the discovery of planets around binary as well as sunlike stars: Mayor and his colleague Didier Queloz announced they had found a half–Jupiter mass planet orbiting around the solar twin 51 Pegasi.

In the nine years since then, well over 100 extrasolar planets have been discovered, all of them gas giants similar to Jupiter. As the effort continues to expand, the race is on to find the first Earth-like planet outside our solar system, and it could indeed be found around a binary system. SA

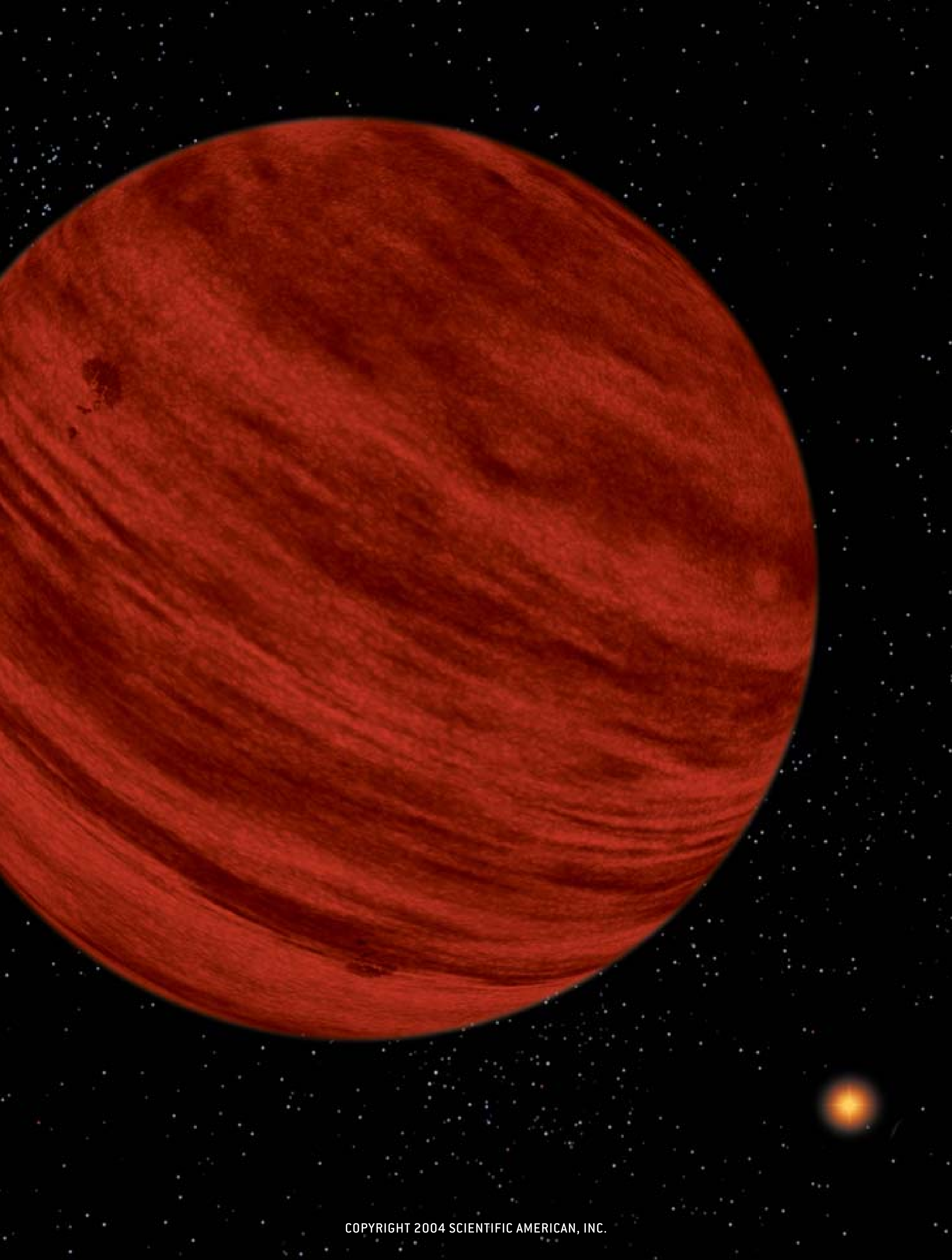
MORE TO EXPLORE

Formation of Binary Stars. Alan P. Boss in *The Realm of Interacting Binary Stars.* Edited by J. Sahade, G. E. McCluskey, Jr., and Y. Kondo. Kluwer Academic Publishers, 1993.

Stellar Multiple Systems: Constraints on the Mechanism of Origin. P. Bodenheimer, T. Ruzmaikina and R. D. Mathieu in *Protostars & Planets,* Vol. 3. Edited by E. H. Levy and J. I. Lunine. University of Arizona Press, 1993.

A Lunar Occultation and Direct Imaging Survey of Multiplicity in the Ophiuchus and Taurus Star-Forming Regions. M. Simon, A. M. Ghez, C. Leinert, L. Cassar, W. P. Chen, R. R. Howell, R. F. Jameson, K. Matthews, G. Neugebauer and A. Richichi in *Astrophysical Journal,* Vol. 443, No. 2, Part 1, pages 625–637; April 20, 1995.

Pre-Main-Sequence Binary Stars. R. D. Mathieu in *Annual Review of Astronomy and Astrophysics,* Vol. 32, pages 465–530; 1995.



THE DISCOVERY OF
BROWN DWARFS

Less massive than stars but more massive than planets, brown dwarfs were long assumed to be rare. New sky surveys, however, show that the objects may be as common as stars By Gibor Basri

A brown dwarf is a failed star. A star shines because of the thermonuclear reactions in its core, which release enormous amounts of energy by fusing hydrogen into helium. For the fusion reactions to be sustained, though, the temperature in the star's core must reach at least three million kelvins. And because core temperature rises with gravitational pressure, the star must have a minimum mass: about 75 times the mass of the planet Jupiter, or about 7 percent of the mass of our sun. A brown dwarf just misses that mark—it is heavier than a gas-giant planet but not quite massive enough to be a star.

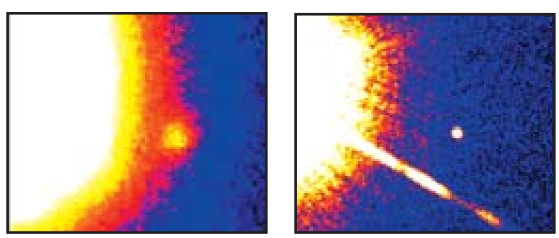
For decades, brown dwarfs were the "missing link" of celestial bodies: thought to exist but never observed. In 1963 University of Virginia astronomer Shiv Kumar theorized that the same process of gravitational contraction that creates stars from vast clouds of gas and dust would also frequently produce smaller objects. These hypothesized bodies were called black stars or infrared stars before the name "brown dwarf" was suggested in 1975 by astrophysicist Jill C. Tarter, now director of research at the SETI Institute in Mountain View, Calif. The name is a bit misleading; a brown dwarf actually appears red, not brown. But the name "red dwarf" was already taken. (It is used to describe stars with less than half the sun's mass.)

In the mid-1980s astronomers began an intensive search for brown dwarfs, but their early efforts were unsuccessful. It was not until 1995 that they found the first indisputable evidence of their existence. That discovery opened the floodgates; since then, researchers have detected hundreds of the objects. Now observers and theorists are tackling a host of intriguing questions: How many brown dwarfs are there? What is their range of masses? Is there a continuum of objects all the way down to the mass of Jupiter? And did they all originate in the same way?

The search for brown dwarfs was long and difficult because they are so faint. All astrophysical objects—including stars, planets and brown dwarfs—emit light during their formation because of the energy released by gravitational contraction. In a star, the glow caused by contraction is eventually supplanted by the thermonuclear radiation from hydrogen fusion; once it begins, the star's size and luminosity stay constant, in most cases for billions of years. A brown dwarf, however, cannot sustain hydrogen fusion, and its light steadily fades as it shrinks [*see box on page 31*]. The light from brown dwarfs is primarily in the near-infrared part of the spectrum. Because brown dwarfs are faint from the start and dim with time, some scientists speculated that they were an important constituent of "dark matter," the mysterious invisible mass that greatly outweighs the luminous mass in the universe.

Astronomers assumed that a good place to look for very faint objects would be close to known stars. More than half the stars in our galaxy are in binary pairs—two stars orbiting their common center of gravity—and researchers suspected that some stars that seemed to be alone might actually have a brown dwarf as a companion. One advantage of such a search is that astronomers do not have to survey large sections of sky for brown dwarfs—they can focus their telescopes on small areas near known stars.

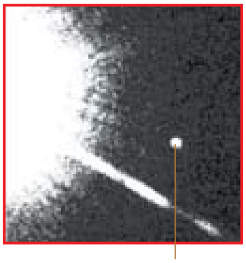
The strategy looked good early on. A likely candidate ap-

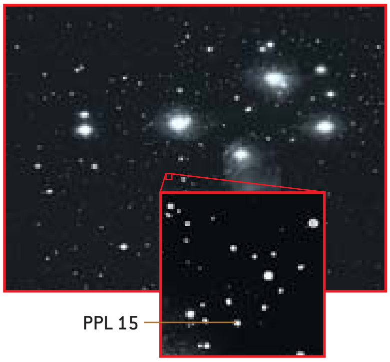


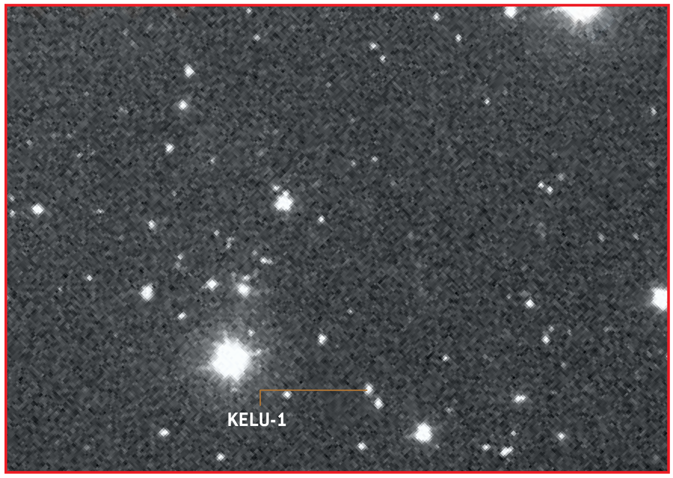
BROWN DWARF GLIESE 229B gives off a red glow in this artist's conception (*opposite page*). The object is believed to be slightly smaller than Jupiter but about 10 times hotter and 30 to 40 times more massive. It was discovered in 1995 as a companion to the red dwarf star Gl 229A (*shown in background*). Astronomers detected the brown dwarf in images from the Palomar Observatory's 1.5-meter telescope (*near right*) and from the Hubble Space Telescope (*far right*) that show the object as a faint spot next to the red dwarf. Gl 229B is actually more than six billion kilometers from its companion star—farther than Pluto is from our sun.

PAT RAWLINGS (*opposite page*); T. NAKAJIMA *California Institute of Technology* AND S. DURRANCE *Johns Hopkins University* (*left*); S. KULKARNI *Caltech*, D. GOLIMOWSKI *Johns Hopkins University* AND NASA (*right*)

FINDING BROWN DWARFS: SEARCH METHODS







OBSERVING FAINT OBJECTS such as brown dwarfs requires special strategies. One approach is to focus telescopes on areas near known stars and to look for companions; astronomers used this method to find Gl 229B (*above left*). Another strategy is to concentrate on young star clusters, because brown dwarfs are brightest when they are young. Scientists searched the 120-million-year-old Pleiades cluster (*above center*) to find the brown dwarf PPl 15 (*center inset*) as well as many others. Last, astronomers can find "field" brown dwarfs by imaging large sections of sky with instruments that are sensitive to faint, red sources. The discovery of the first field brown dwarf, Kelu-1 (*above right*), was announced in 1997.

peared in 1988, when Eric Becklin and Benjamin Zuckerman of the University of California at Los Angeles reported the discovery of GD 165B, a faint red companion to a white dwarf. White dwarfs are unrelated to brown dwarfs: they are the corpses of moderately massive stars and are smaller, hotter and much heavier than brown dwarfs. GD 165B may indeed be a brown dwarf, but astronomers have been unable to say for certain because the object's inferred mass is close to the 75-Jupiter-mass boundary between low-mass stars and brown dwarfs.

Another advantage of looking for brown dwarfs as companions to stars is the brown dwarf itself doesn't necessarily have to be observed. Researchers can detect them with the same method used to find extrasolar planets: by observing their periodic effects on the motions of the stars they are circling. Astronomers determine the variations in the stars' velocities by measuring the Doppler shifts in the stars' spectral lines. It is actually easier to detect brown dwarfs than planets by this technique because of their greater mass.

Nevertheless, famed planet hunter Geoffrey W. Marcy of San Francisco State University and the University of California at Berkeley found no brown dwarfs in a survey of 70 low-mass stars conducted in the late 1980s. In the mid-1990s Marcy discovered half a dozen extrasolar gas-giant planets in a survey of 107 stars similar to our sun but still saw no clear-cut evidence of brown dwarfs. The failure of these efforts gave rise to the term "brown dwarf desert" because the objects appeared to be much less common than giant planets or stars.

Only one of the early Doppler-shift searches detected a brown dwarf candidate. In a 1988 survey of 1,000 stars, David W. Latham of the Harvard-Smithsonian Center for Astrophysics found a stellar companion at least 11 times as massive as Jupiter. The Doppler-shift method, though, provides only a lower limit on a companion's mass, so Latham's object could be a very low mass star instead of a brown dwarf. This issue will remain unresolved until scientists can determine stellar positions more precisely.

Meanwhile other astronomers pursued a different strategy that took advantage of the fact that brown dwarfs are brightest when they are young. The best place to look for young objects is in star clusters. The stars in a cluster all form at the same time but have very different lifetimes. The most massive stars shine for only a few million years before running out of hydrogen fuel and leaving the main-sequence phase of their lifetimes, whereas low-mass stars can keep shining for billions, even trillions, of years. The standard method for estimating the age of a cluster amounts to finding its most massive main-sequence star. The age of the cluster is roughly the lifetime of that star.

Once researchers locate a young cluster and determine its age, they need only look for the faintest, reddest (and therefore coolest) objects in the cluster to identify the brown dwarf candidates. Theory provides the expected surface temperature and luminosity of objects of various masses for a given age, so by measuring these properties astronomers can estimate each candidate's mass. Several teams began the search, imaging the areas of sky containing young clusters and picking out faint red objects.

The research teams made a series of announcements of brown dwarf candidates in young clusters, including the star-forming region in the Taurus constellation and the bright cluster called the Pleiades (better known as the Seven Sisters). Unfortunately, closer scrutiny showed that none of the candidates was really a brown dwarf. Some turned out to be red giant stars located thousands of light-years behind the cluster; because these background stars are so distant, they appear faint even though they are quite luminous. Others were low-mass stars behind or in front of the cluster. Some of the "discoveries" made it into the press, but the later retractions were

T. NAKAJIMA *Caltech* AND S. DURRANCE *Johns Hopkins University* (*left*); SPACE TELESCOPE SCIENCE INSTITUTE (*top, center*); JOHN STAUFFER *Harvard-Smithsonian Center for Astrophysics* (*bottom, center*); EUROPEAN SOUTHERN OBSERVATORY (*right*) not given much play. This led to further skepticism among astronomers toward all brown dwarf announcements and reinforced the widespread view that the objects were rare.

Looking for Lithium

IN 1992 RAFAEL REBOLO, Eduardo L. Martín and Antonio Magazzú of the Astrophysics Institute in Spain's Canary Islands proposed a clever new method to help distinguish low-mass stars from brown dwarfs. Called the lithium test, it exploits the fact that below a mass of about 60 Jupiter masses, a brown dwarf never achieves the conditions necessary to sustain lithium fusion in its core. This nuclear reaction occurs at a slightly lower temperature than hydrogen fusion does; as a result, stars quickly consume whatever lithium they originally had. Even the lowest-mass star burns all its lithium in about 100 million years, whereas all but the most massive brown dwarfs retain their lithium forever. Thus, the continued presence of lithium is a sign that the object has a substellar mass.

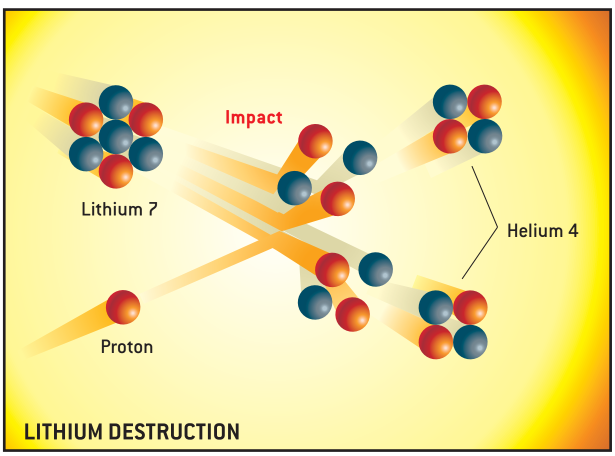
The spectral lines produced by lithium are fairly strong in cool red objects. The Canary Islands group looked for these lines in all the coolest objects in the sky that are also bright enough to provide a spectrum of the needed quality. None showed evidence of lithium. In 1993 another team—consisting of myself, Marcy and James R. Graham of Berkeley—began to apply the lithium test to fainter objects using the newly built 10-meter Keck telescope on Mauna Kea in Hawaii. We, too, met with failure at first, but our luck changed when we focused on the Pleiades cluster.

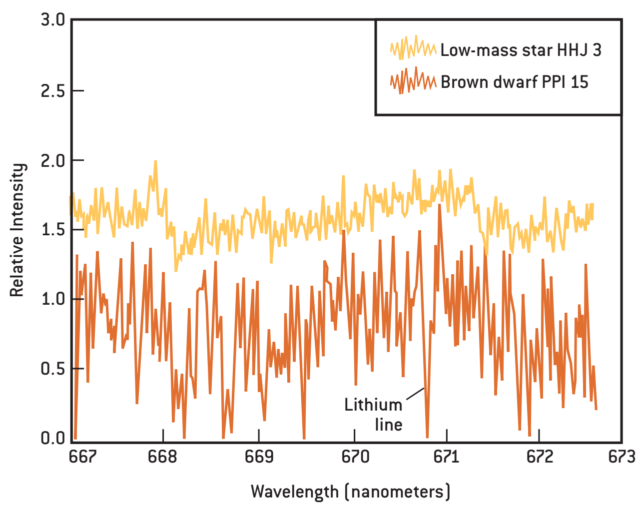
A group of British astronomers had just conducted one of the broadest, deepest surveys of the cluster. They found several objects that by all rights should have had substellar masses. They showed that these objects shared the proper motion of the cluster across the sky and thus had to be members of the cluster rather than background stars. We went right to the faintest one, an object called HHJ 3, expecting to find lithium. It was not present. But Harvard-Smithsonian Center astronomer John Stauffer supplied us with another target. He, too, had been surveying the Pleiades for low-mass objects and had detected an even fainter candidate, dubbed PPl 15 (the 15th good candidate in the Palomar Pleiades survey). At last, we were successful: for the first time we detected lithium in an object for which its presence implied a substellar mass. We reported the discovery at the June 1995 meeting of the American Astronomical Society. Our results indicated that the cluster was about 120 million years old, giving PPl 15 an inferred mass at the upper end of the brown dwarf range.

In one of the interesting convergences that seem to occur regularly in science, other research teams also reported strong evidence of brown dwarfs in 1995. The Canary Islands group had also been conducting a deep survey of the Pleiades cluster and had detected two objects even fainter than PPl 15: Teide 1 and Calar 3, both named after Spanish observatories. Each had an inferred mass just below 60 Jupiter-masses. By the end of the year I had teamed up with the Canary Islands group, and we confirmed the expected presence of lithium in both objects. The astronomical community retained some skepticism about these objects for the first few months—after all, they still looked like stars—until further discoveries made it clear that now the brown dwarfs were for real.

At the same time, a very different

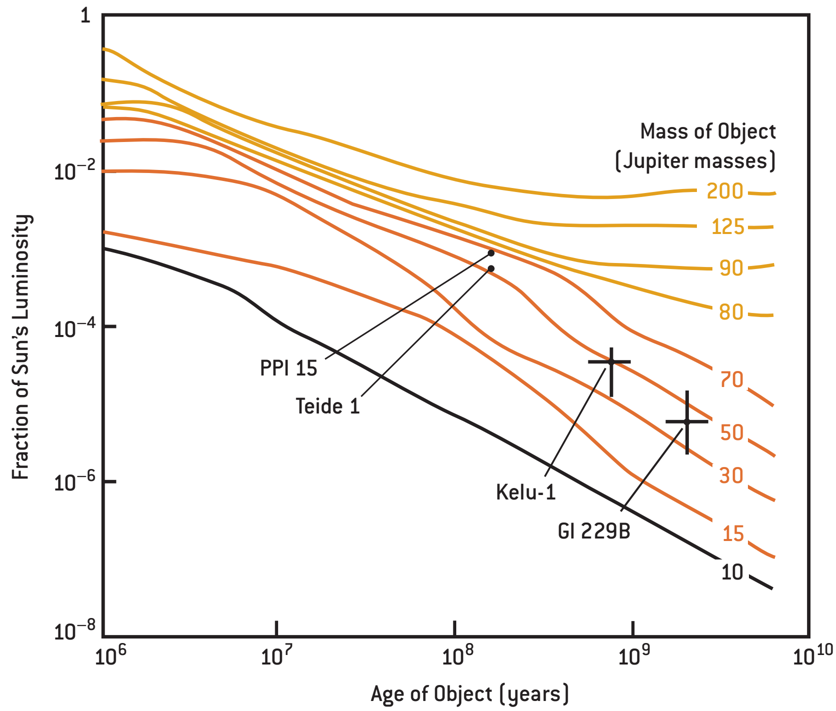
CONFIRMING THE DISCOVERIES: THE LITHIUM TEST





ANALYZING THE SPECTRA of faint objects can reveal whether they are stars or brown dwarfs. All stars destroy the lithium in their cores; in this reaction, a proton collides with the isotope lithium 7, which then splits into two helium atoms (*left*). In contrast, all but the most massive brown dwarfs cannot achieve the core temperature needed for lithium destruction, so they retain the element forever. The spectrum of HHJ 3 (*right, yellow line*), a low-mass star in the Pleiades cluster, shows no sign of lithium. The spectrum of brown dwarf PPl 15 (*red line*), however, has a strong absorption line indicating the presence of the light metallic element.

BRYAN CHRISTIE (*left*); LAURIE GRACE (*right*)



LUMINOSITY HISTORY of low-mass stars (*yellow lines*), brown dwarfs (*red lines*) and planets (*black line*) shows that only stars are massive enough to achieve a stable luminosity. The light from brown dwarfs and planets fades as they age. Data from brown dwarfs (*black crosses*) indicate how old and heavy they are.

search bore spectacular fruit. A group from the California Institute of Technology and Johns Hopkins University had been looking for brown dwarf companions of nearby low-mass stars. The astronomers had equipped the Palomar 1.5-meter telescope with an instrument that blocked most of the light of the primary star, allowing a faint nearby companion to be more easily seen. In 1993 they observed several brown dwarf candidates. They took second images a year later. Because the targets are relatively close to our solar system, their movements through the galaxy are perceptible against the background stars. If a candidate is truly a companion, it will share this motion. One of the companions confirmed was 1,000 times fainter than its primary, the low-mass star Gliese 229A. Because the primary was already known to be faint, the companion's luminosity had to be well below that of the faintest possible star. The group kept quiet until it obtained an infrared spectrum of the object.

At a meeting of the Cambridge Workshop on Cool Stars, Stellar Systems and the Sun in October 1995, the Caltech/Johns Hopkins group announced the discovery of Gl 229B, the brown dwarf companion to Gl 229A. It was clearly substellar by virtue of its faintness, and the clincher was the detection of methane in its spectrum. Methane is common in the atmospheres of the giant planets, but all stars are too hot to allow it to form. Its strong presence in Gl 229B guaranteed that this object could not be a star. At the same meeting the Canary Islands group reported the observation of several new brown dwarf candidates in the Pleiades cluster, suggesting that these objects might be fairly numerous. In addition, a group led by Michel Mayor of the Geneva Observatory in Switzerland announced the discovery of the first extrasolar planet, a gas giant circling the star 51 Pegasi. In one morning, the frustrating search for substellar objects came to a dramatic conclusion.

Most astronomers view Gl 229B as the first indisputable brown dwarf discovered because it is a million times fainter than the sun and has a surface temperature under 1,000 kelvins—far below the minimum temperature that even the faintest star would generate (around 1,800 kelvins). It has reached this state because it is a few billion years old. It is probably 30 to 40 times more massive than Jupiter. In contrast, PPl 15, Teide 1 and Calar 3 in the Pleiades are more massive (from 50 to 70 Jupiter masses) and also much hotter (with surface temperatures between 2,600 and 2,800 kelvins), primarily because they are much younger.

Once the methods for detecting brown dwarfs had been proved, the discoveries came at an increasing pace. Several groups returned to the Pleiades. The Canary Islands group, now including Maria Rosa Zapatero Osorio of the Laboratory for Space Astrophysics and Theoretical Physics, near Madrid, discovered a Pleiades brown dwarf only 35 times more massive than Jupiter—the lightest brown dwarf found in the cluster. More important, the Canary Islands group conducted the first useful assessment of the number of brown dwarfs in the Pleiades by counting the most likely candidates in a small surveyed area and then extrapolating the tally for the entire cluster. The results indicated comparable numbers of stars and brown dwarfs in the Pleiades. If true in general, this would mean that our galaxy alone contains about 100 billion brown dwarfs. But it also means that brown dwarfs are not the dominant constituent of the universe's mass, because they are much lighter than stars. The hope that they would provide an answer to the dark matter mystery has faded.

Other researchers focused on how the brown dwarfs are distributed by mass. What is the lowest mass a brown dwarf can attain? Is there a continuum of objects down to the planetary range—below 13 Jupiter masses—or is there a gap

THE AUTHOR

GIBOR BASRI is a professor in the astronomy department at the University of California, Berkeley. He received his Ph.D. in astrophysics from the University of Colorado at Boulder in 1979. Basri has studied solar-type stars, low-mass stars and star formation using a variety of telescopes, including the Lick and Keck observatories, as well as spaceborne instruments, including the Hubble Space Telescope. He is keen on promoting an interest in science among groups currently underrepresented in the field.

LAURIE GRACE

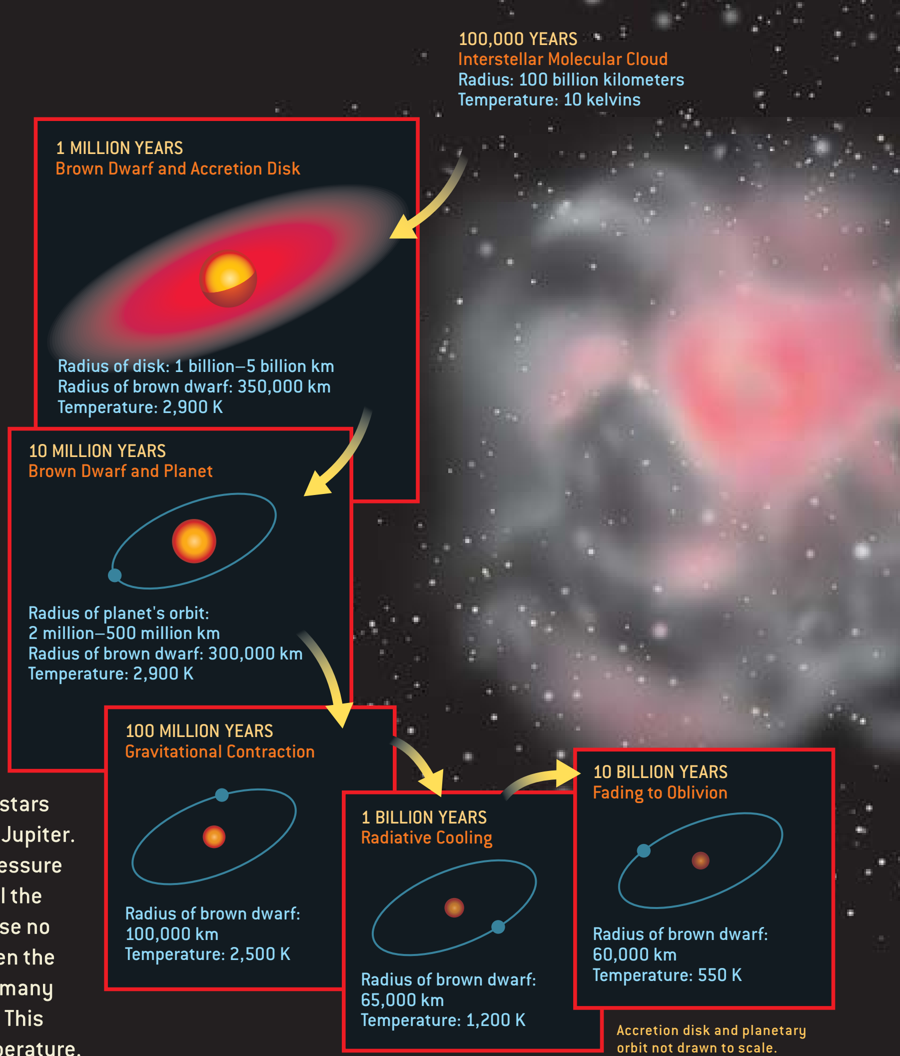
The Life Cycle of Brown Dwarfs

The early lives of brown dwarfs and stars follow the same pattern. Both are believed to originate from the gravitational collapse of interstellar clouds of gas and dust. These clouds are composed primarily of hydrogen and helium, but they also initially contain small amounts of deuterium and lithium that are remnants of the nuclear reactions that took place a few minutes after the big bang.

As young stars and brown dwarfs contract, their cores grow hotter and denser, and the deuterium nuclei fuse into helium 3 nuclei. (Deuterium fusion can occur in brown dwarfs because it requires a lower temperature—and hence a lower mass—than hydrogen fusion.) The outpouring of energy from these reactions temporarily halts the gravitational contraction and causes the objects to brighten. But after a few million years the deuterium runs out, and the contraction resumes. Lithium fusion occurs next in stars and in brown dwarfs more than 60 times as massive as Jupiter.

During the contraction of a brown dwarf, thermal pressure rises in its core and opposes the gravitational forces. All the electrons are freed from their nuclei by the heat. Because no two electrons can occupy the same quantum state, when the core is very dense the low-energy states are filled, and many electrons are forced to occupy very high energy states. This generates a form of pressure that is insensitive to temperature. Objects supported in this manner are called degenerate. One consequence of this process is that all brown dwarfs are roughly the size of Jupiter—the heavier brown dwarfs are simply denser than the lighter ones.

In stars the cores do not become degenerate. Instead hydrogen fusion provides the pressure that supports the star against its own gravity. Once fusion begins in earnest, the star stops contracting and achieves a steady size, luminosity and temperature. In high-mass brown dwarfs, hydrogen fusion begins but then sputters out. As degeneracy pressure slows the collapse of brown dwarfs, their luminosity from gravitational contraction declines. Although very low mass stars can shine for trillions of years, brown dwarfs fade steadily toward oblivion. This makes them increasingly difficult to find as they age. In the very distant future, when all stars have burned out, brown dwarfs will be the primary repository of hydrogen in the universe. —G.B.



BROWN DWARF IS BORN from the contraction of a vast cloud of gas and dust. After a million years the object is a glowing ball of gas, possibly surrounded by an accretion disk from which an orbiting planet could later arise. (So far no planets have been detected around brown dwarfs; their existence and possible orbits are strictly hypothetical.) Over time the brown dwarf shrinks and cools. The radii and surface temperatures shown here are for an object of 40 Jupiter masses.

BRYAN CHRISTIE

between the lightest brown dwarf and the heaviest planet because they are formed by different mechanisms?

The best place to answer these questions is in newly forming star clusters, where even very low mass brown dwarfs are still bright enough to see. Surveys of various nearby star-forming regions have turned up a number of objects that seem cool and dim enough to lie near, or even below, the fusion limit, as predicted by models. The models are not very reliable for such objects, however, so there has been some skepticism, and some objects have turned out to not actually be in the star-forming regions. Recently, though, work by me, Subhanjoy Mohanty of the Harvard-Smithsonian Center and our collaborators has taken a more direct tack. We do not rely on evolutionary models but deduce the surface gravity of confirmed members by studying their spectral lines (which broaden as gravity and pressure increase). We confirm very low masses in some cases. Thus, it appears that brown dwarfs are produced in all possible masses between planets and stars [*see box on next page*].

Continuing searches for brown dwarfs around solar-type stars have confirmed the initial impression that they are fairly rare in this situation. Brown dwarfs appear to be more common, though, as companions to lower-mass stars. In 1998 Rebolo and his co-workers discovered one orbiting the young star G196-3. Despite its youth, this brown dwarf is already quite cool, which means it must be light, perhaps only 20 Jupiter masses.

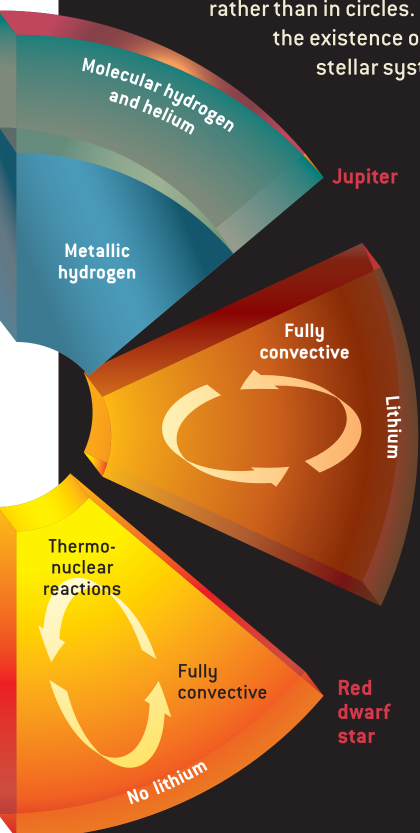
The first binary system involving two brown dwarfs was identified by Martín and me. We determined that the Pleiades

Planets versus Brown Dwarfs

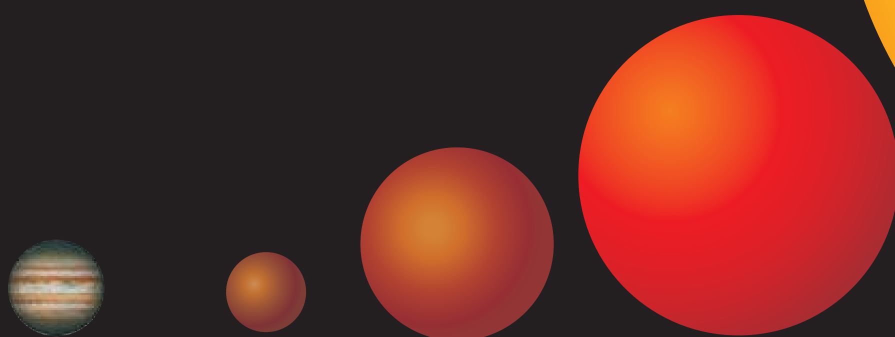
Is there a fundamental difference between the largest planets and the smallest brown dwarfs? The classical view is that planets form in a different way than brown dwarfs or stars do. Gas-giant planets are thought to build up from planetesimals—small rocky or icy bodies—amid a disk of gas and dust surrounding a star. Within a few million years these solid cores attract huge envelopes of gas. This model is based on our own solar system and predicts that all planets should be found in circular orbits around stars and that gas-giant planets should travel in relatively distant orbits.

These expectations have been shattered by the discovery of the first extrasolar giant planets. Most of these bodies have been found in close orbits, and most travel in eccentric ovals rather than in circles. Some theorists have even predicted the existence of lone planets, thrown out of their stellar systems by orbital interactions with sibling planets. This makes it very hard for observers to distinguish planets from brown dwarfs on the basis of how or where they formed or what their current location and motion is. We can find brown dwarfs by themselves or as orbital companions to stars or even other brown dwarfs. The same may be true for giant planets.

An alternative approach is gaining adherents: to distinguish between planets and brown dwarfs based on whether the object has ever managed to produce any nuclear fusion reactions. In this view, the dividing line is set at about 13 Jupiter masses. Above that mass, deuterium fusion occurs in the object. The fact that brown dwarfs seem to be less common than planets—at least as companions to more massive stars—suggests that the two types of objects may form by different mechanisms. A mass-based distinction, however, is much easier to observe. —*G.B.*



CONTINUUM OF OBJECTS from planets to stars (*below*) shows that older brown dwarfs, such as Gliese 229B, are fairly similar to gas-giant planets in size and surface temperature. Younger brown dwarfs, such as Teide 1, more closely resemble low-mass stars, such as Gliese 229A. Brown dwarfs and low-mass stars are fully convective, meaning that they mix their contents (*left*). Thermonuclear reactions in the stars' cores destroy all their lithium, so its presence is a sign that the object may be a brown dwarf.



Name	Jupiter	Gliese 229B	Teide 1	Gliese 229A	Sun
Type of object	Gas-giant planet	Brown dwarf	Brown dwarf	Red dwarf star	Yellow dwarf star
Mass (Jupiter masses)	1	30–40	55	300	1,000
Radius (kilometers)	71,500	65,000	150,000	250,000	696,000
Temperature (kelvins)	100	1,000	2,600	3,400	5,800
Age (years)	4.5 billion	2–4 billion	120 million	2–4 billion	4.5 billion
Hydrogen fusion	No	No	No	Yes	Yes
Deuterium fusion	No	Yes	Yes	Yes	Yes

We now have an idea of how brown dwarfs look AS THEIR ATMOSPHERES COOL TO almost planetary temperatures.

brown dwarf PPl 15 is really a close pair of brown dwarfs, with an orbital period of six days. Together with German astronomer Wolfgang Brandner, we also resolved the first nearby binary pair of very cool objects with the Hubble Space Telescope. Such systems should provide the first real dynamical masses of brown dwarfs within a few years.

Several subsequent studies have shown that the fraction of very cool objects that turn out to be double in space telescope images is about 20 percent (the fraction of stellar binaries at any separation is about 50 percent). These observations suggest that the brown dwarf desert is only a lack of brown dwarfs as companions to more massive stars. When looking near low-mass objects (either stars or brown dwarfs), the likelihood of finding a brown dwarf companion is much greater. This variance probably results from the process that gives birth to binary systems, which is still poorly understood. Apparently this process is less likely to produce a system in which the primary object is more than about 10 times the mass of the secondary. Remarkably, the brown dwarfs are never found with separations more than about the size of our solar system, even though that is only the median separation for stars.

Brown Dwarfs Everywhere

ASTRONOMERS FOUND still more brown dwarfs using another search technique: looking for them at random locations in the sky. These "field" brown dwarfs are easily lost among the myriad stars of our galaxy. To locate such objects efficiently, one must image large sections of sky with great sensitivity to faint red sources. The first field brown dwarf was announced by Maria Teresa Ruiz of the University of Chile in 1997. She dubbed it "Kelu-1" from a South American Indian word for "red" and noted that it shows lithium. At about the same time, the Deep Near-Infrared Survey (DENIS)—a European project that scans the southern hemisphere of the sky—found three similar objects. Researchers quickly confirmed that one contains lithium.

The Two Micron All Sky Survey (2MASS), managed by the University of Massachusetts at Amherst, has detected even more field brown dwarfs. The team, including J. Davy Kirkpatrick of NASA's IPAC center in Pasadena, Calif., has found hundreds of new extremely cool objects and confirmed lithium in over 50. Most of these objects have surface temperatures greater than 1,500 kelvins and so must be younger than about a billion years. They are relatively bright and easier to observe than older objects.

The hunt for older field brown dwarfs was frustrated until the summer of 1999, when the Sloan Digital Sky Survey (which uses optical detectors) turned up two brown dwarfs containing methane in their atmospheres. The presence of methane indicates a surface temperature below 1,300 kelvins and hence an age greater than one billion to two billion years. At the same time, the 2MASS group reported the observation of four similar objects. Most of the brown dwarfs in our galaxy should be methane-bearing, because the majority formed long ago and should have now cooled to that state. Thus, these discoveries are just the tip of the iceberg. Adam Burgasser of Caltech and others have now been able to collect a large enough sample of older brown dwarfs to give us a preliminary idea of how they look as their atmospheres cool to near-planetary temperatures.

Further study of very cool objects has yielded clues to the composition and evolution of brown dwarf atmospheres. Their optical spectra lack the molecules of titanium oxide and vanadium oxide that dominate the spectra of low-mass stars. These molecules do not appear, because their constituent heavy elements condense into hard-to-melt dust grains. The primary optical spectral lines are from molecules of hydrides, instead of oxides, and from neutral alkali metals. The dust grains form into clouds, whose height appears to drop as the objects cool. There is some evidence the clouds may not always cover the whole object, so one might be able to study "weather" on brown dwarfs. The dust clouds sink below the visible surface in the methane objects, whose optical spectra then are dominated by sodium and potassium. These spectral lines are broadened even more than in high-pressure street lamps, making the color of the objects magenta (not brown!).

The initial discovery phase for brown dwarfs is now almost over. Astronomers have good methods for detecting them and many targets for detailed study. Indeed, the 2MASS and DENIS teams have found that the number of field brown dwarfs in the surveyed areas is similar to the number of low-mass stars in those areas. Brown dwarfs seem to be nearly as common as stars.

Over the next few years, scientists will get a better handle on the basic facts about brown dwarfs: their numbers, masses and distribution in our galaxy. Researchers will also try to determine how they form as binary or solo objects and what processes take place as their atmospheres cool. It is remarkable that these nearby and common objects, as abundant as stars, have only now begun to reveal their secrets. SA

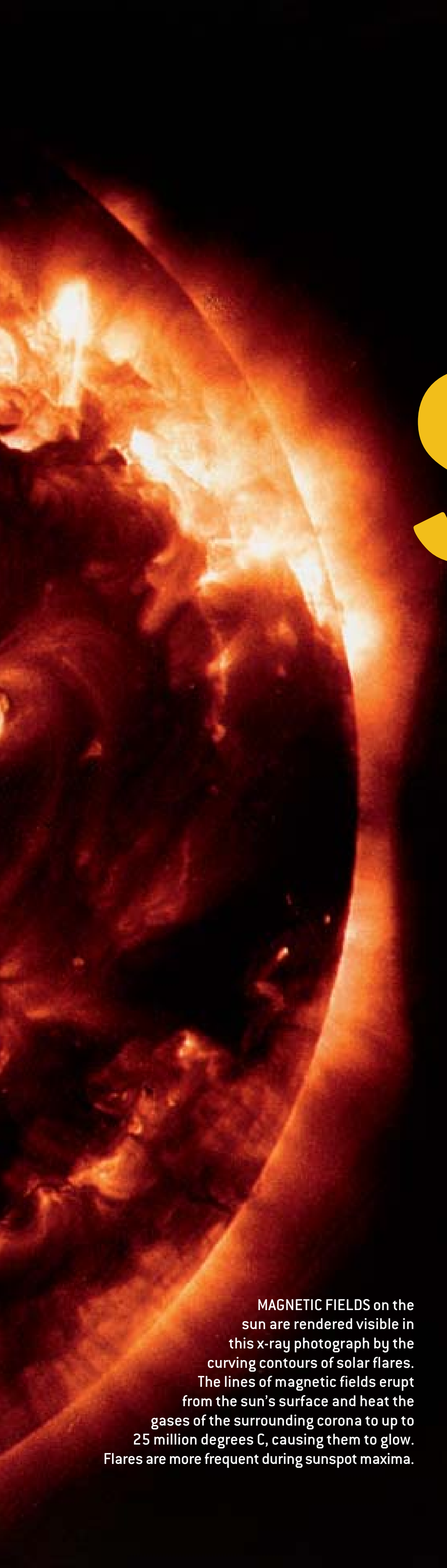
MORE TO EXPLORE

Brown Dwarfs: A Possible Missing Link between Stars and Planets. S. R. Kulkarni in *Science,* Vol. 276, pages 1350–1354; May 30, 1997.

Brown Dwarfs and Extrasolar Planets. Edited by R. Rebolo, E. L. Martín and M. R. Zapatero Osorio. Astronomical Society of the Pacific Conference Series, Vol. 134, 1998.

More on brown dwarfs is available at **astron.berkeley.edu/~basri/bdwarfs/**

BRYAN CHRISTIE



MAGNETIC FIELDS on the sun are rendered visible in this x-ray photograph by the curving contours of solar flares. The lines of magnetic fields erupt from the sun's surface and heat the gases of the surrounding corona to up to 25 million degrees C, causing them to glow. Flares are more frequent during sunspot maxima.

The Stellar Dynamo

Sunspot cycles—on other stars—are helping astronomers study the sun's variations and the ways they might affect Earth

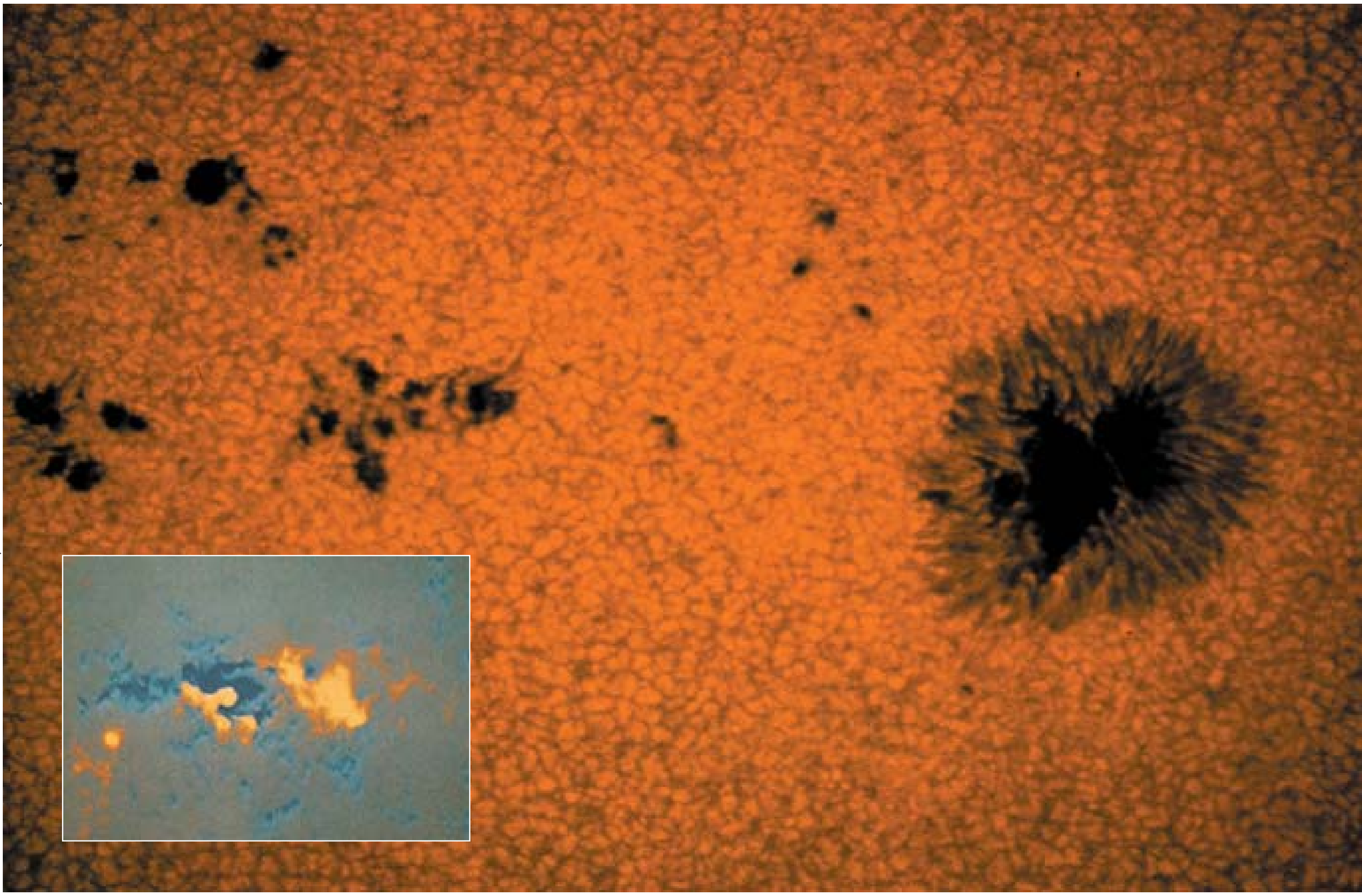
By Elizabeth Nesme-Ribes, Sallie L. Baliunas and Dmitry Sokoloff

In 1801, musing on the vagaries of English weather, astronomer William Herschel observed that the price of wheat correlated with the disappearance of sunspots. But the pattern soon vanished, joining what scientists at large took to be the mythology connecting earthly events with solar ones. That the sun's brightness might possibly vary, and thereby affect Earth's weather, remained speculative.

Thus, in the mid-1980s, when three solar satellites—Solar Maximum Mission, Nimbus 7 and Earth Radiation Budget—reported that the sun's radiance was declining, astronomers assumed that all three instruments were failing. But the readings then perked up in unison, an occurrence that could not be attributed to chance. The sun was cooling off and heating up; furthermore, the variation was connected with the number of spots on its face.

In recent years one of us (Baliunas) has observed that other stars undergo rhythmic changes much like those of our sun. Such studies are helping refine our understanding of the "dynamo" that drives the sun and other stars. Moreover, they have revealed a strong link between "star spots" and luminosity, confirming the patterns discovered in our sun. And yet astrophysicists, including the three of us, are still debating the significance of the sun's cycles and the extent to which they might influence Earth's climate.

NATIONAL SOLAR OBSERVATORY/SACRAMENTO PEAK; NATIONAL SOLAR OBSERVATORIES/KITT PEAK (*inset*)

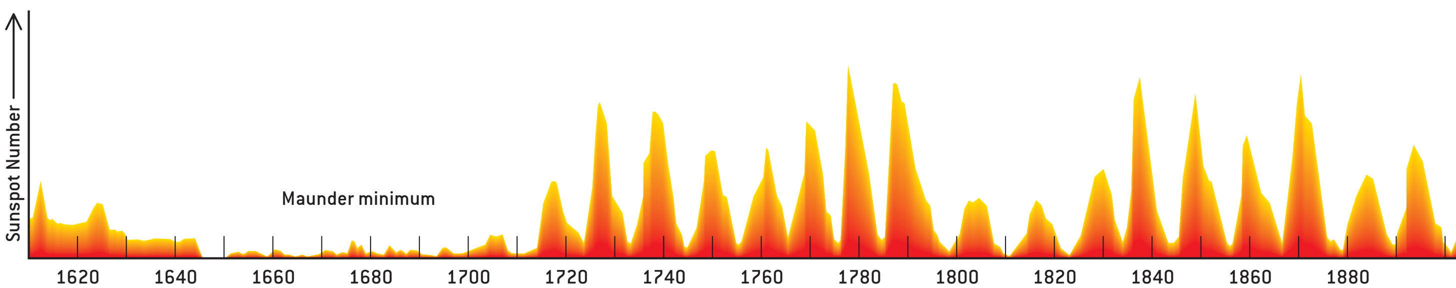


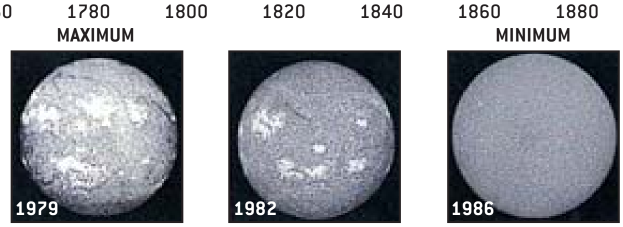
SUNSPOTS are relatively cool regions formed where magnetic fields emerge from the sun, thereby suppressing the upwelling of hot gases from the interior. Elsewhere on the surface, tightly coiled cells of cyclonically flowing gases show up as granules. Near a sunspot the magnetic fields organize the gaseous flow into lines resembling iron filings near a bar magnet. The magnetogram (*inset*) shows field lines emerging at one sunspot (*yellow*) and reentering at another (*blue*); such sunspot pairs are common.

Sunspots

THE EARLIEST KNOWN sunspot records are Chinese documents that go back 2,000 years, preserving observations made by the naked eye. From 1609 to 1611 Johannes Fabricius, Thomas Harriot, Christoph Scheiner and Galileo Galilei, among others, began telescopic studies of sunspots. These records, as German astronomer Samuel Heinrich Schwabe announced in 1843, displayed a prominent periodicity of roughly 10 years in the number of observed sunspot groups. By the 20th century George Ellery Hale of the Mount Wilson Observatory in California found those dark surface irregularities to be the seat of intense magnetic fields, with strengths of several thousand gauss. (Earth's magnetic field is, on the average, half a gauss.)

Sunspots appear dark because they



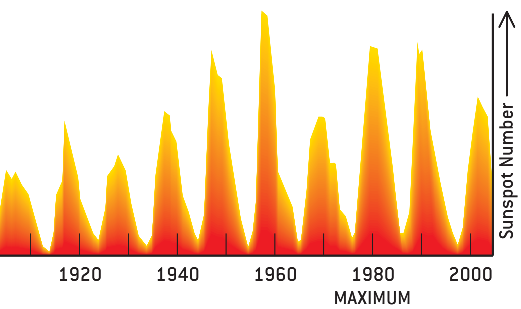


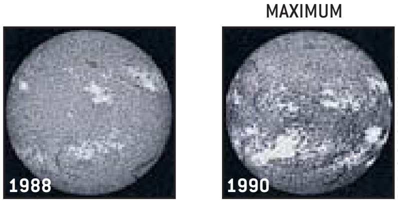
ELEVEN-YEAR CYCLES of sunspot activity were interrupted between 1645 and 1715 by a period of quiescence. This dearth of sunspots, called the Maunder minimum, coincided with unusually cool temperatures across northern Europe, indicating that solar fluctuations influence Earth's climate. The regular pulsing of the sun's activity (*right*) was observed over one cycle at the Paris Observatory. These photographs were taken in violet light emitted by ionized calcium; the technique that produced them is now used to study the magnetic activity of other stars.

SUNSPOTS ARE 2,000 DEGREES C COOLER than the surrounding surface; they would GLOW ORANGE-RED AGAINST A NIGHT SKY.

are 2,000 degrees Celsius cooler than the surrounding surface of the sun; they would glow orange-red if seen against the night sky. The spots form when strong magnetic fields suppress the flow of the surrounding gases, preventing them from carrying internal heat to the surface. Next to the sunspots are often seen bright areas called plages (after the French word for "beach"). The magnetic field lines tend to emerge from the surface at one spot to reenter the sun at another, linking the spots into pairs that resemble the two poles of a bar magnet that is oriented roughly east-west.

At the start of each 11-year cycle, sunspots first appear at around 40 degrees latitude in both hemispheres; they form closer to the equator as the cycle progresses. At sunspot minimum, patches of intense magnetism, called active regions, are seen near the equator. Aside from the sunspots, astronomers have observed that the geographic poles of the sun have weak overall magnetic fields of a few gauss. This large-scale field has a "dipole" configuration, resembling the field of a bar magnet. The leading sunspot in a pair—the one that first comes into view as the sun rotates from west to east—has the same polarity as the pole of its hemisphere; the trailing sunspot has the opposite polarity. Moreover, as Hale and Seth B. Nicholson had discovered by 1925, the polarity patterns reverse every 11 years, so that the total magnetic cycle takes 22 years to complete. But the sun's behavior has not always been so regular. In 1667, when the Paris Observatory was founded, astronomers there began systematic observations of the sun, logging more than 8,000 days of observation over the next 70 years. These records showed very little sunspot activity. This important finding did not raise much interest until the sunspot cycle was discovered, prompting Rudolf Wolf of Zürich Observatory to scrutinize the records. Although he rediscovered the sunspot lull, Wolf's finding was criticized on the grounds that he did not use all the available documents.





NOAA; PETER SAMEK *Slim Films*

PARIS OBSERVATORY

During the late 1880s, first Gustav F. W. Spörer and then E. Walter Maunder reported that the 17th-century solar anomaly coincided with a cold spell in Europe. That astonishing observation lay neglected for almost a century, with many astronomers assuming that their predecessors had not been competent enough to count sunspots. It was only in 1976 that John A. Eddy of the High Altitude Observatory in Boulder, Colo., reopened the debate by examining the Paris archives and establishing the validity of what came to be known as the Maunder minimum.

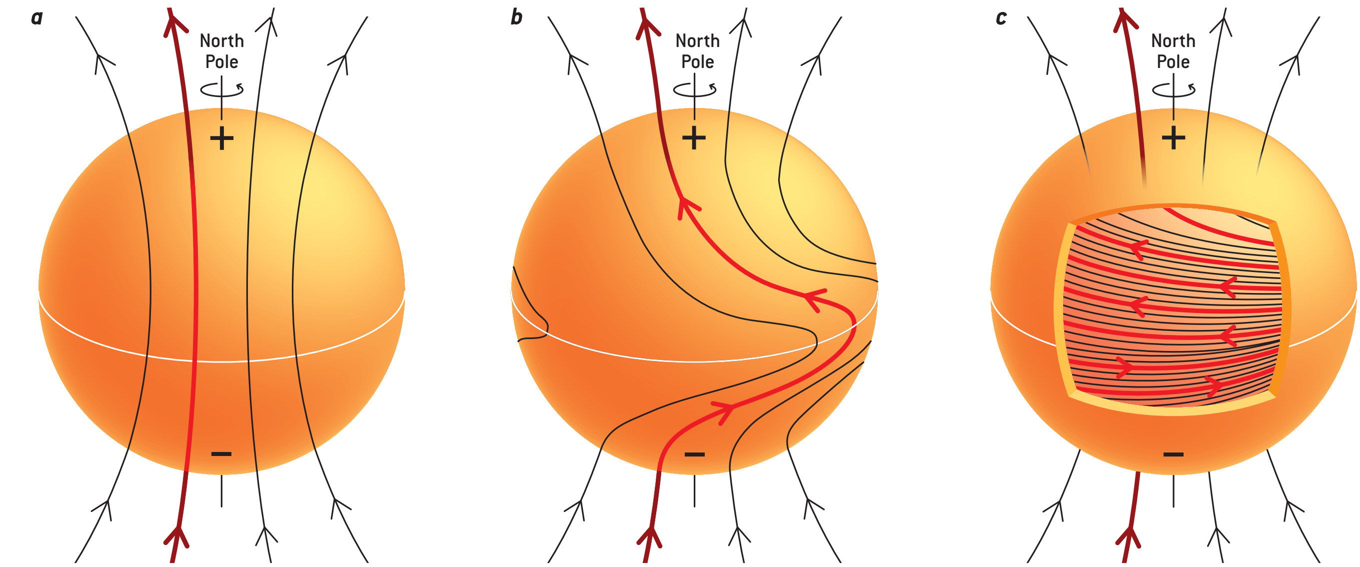
Minze Stuiver, while at Yale University, Hans Suess of the University of California at San Diego and others had discovered that the amount of carbon 14 in tree rings increased during the dearth of sunspots. This radioactive element is created when galactic cosmic rays transmute nitrogen in the upper atmosphere. Their findings suggested that when the magnetic fields in the solar wind—the blast of particles and energy that flows from the sun—are strong, they shield Earth from cosmic rays, so that less carbon 14 forms; the presence of excess carbon 14 indicated a low level of magnetic activity on the sun during the Maunder phase. Eddy thus reinforced the connection between the paucity of sunspots and a lull in solar activity.

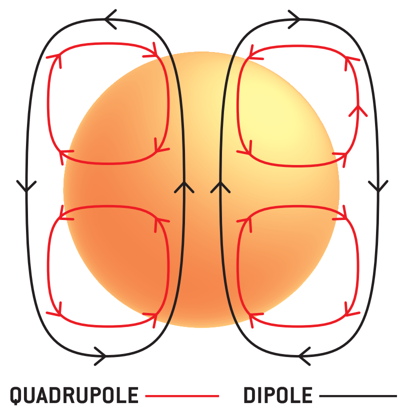
Aside from the rarity of sunspots during the Maunder minimum, the Paris archives brought to light another oddity: from 1661 to 1705, the few sunspots that astronomers sighted were usually in the southern hemisphere. They were also traveling much more slowly across the sun's face than present-day sunspots do. Only at the beginning of the 18th century did the sun assume its modern appearance, having an abundance of sunspots rather evenly distributed between the two hemispheres.

The Solar Dynamo

THE MAGNETIC ACTIVITY of the sun is believed to reside in its convective zone, the outer 200,000 kilometers where churning hot gases bring up energy from the interior. The fluid forms furious whorls of widely different sizes: the best known is an array of convective cells or granules, each 1,000 kilometers across at the surface but lasting only a few minutes. There are also "supergranules" that are 30,000 to 50,000 kilometers across and even larger flows. Rotation gives rise to Coriolis forces that make the whorls flow counterclockwise in the northern hemisphere (if one is looking down at the surface) and clockwise in the southern hemisphere; these directions are called cyclonic.

Whether similar cyclones exist underneath the surface is not known. Deep within, the convective zone gives way to the radiative zone, where the energy is transported by radiation. The core of the sun, where hydrogen fuses into helium to fuel all the sun's activity, seems to rotate rigidly and slowly compared with the surface.





SOLAR DYNAMO generates the sun's magnetic field and also causes it to change orientation every 11 years. Suppose that the initial magnetic field (*a*) resembles that of a bar magnet with its north pole (+) near the sun's geographic north pole. The magnetic field lines are carried along with the electrically charged gases. The faster flow at the equator therefore distorts the field lines (*b*) until they wrap tightly (*c*) around the sun. But the field lines then resist the stretching and unwind,

PETER SAMEK *Slim Films*

The first description of how the sun's gases conspire to create a magnetic field was proposed in 1955 by Eugene N. Parker of the University of Chicago. Because of the high temperature, the atoms of hydrogen and helium lose their electrons, thereby giving rise to an electrically charged substance, or plasma. As the charged particles move, they generate magnetic fields. Recall that the lines describing magnetic fields form continuous loops, having no beginning or end—their density (how closely together the lines are packed) indicates the intensity of the magnetic field, whereas their orientation reveals the direction. Because plasma conducts electricity very efficiently, it tends to trap the field lines: if the lines were to move through the plasma, they would generate a large, and energetically expensive, electric current.

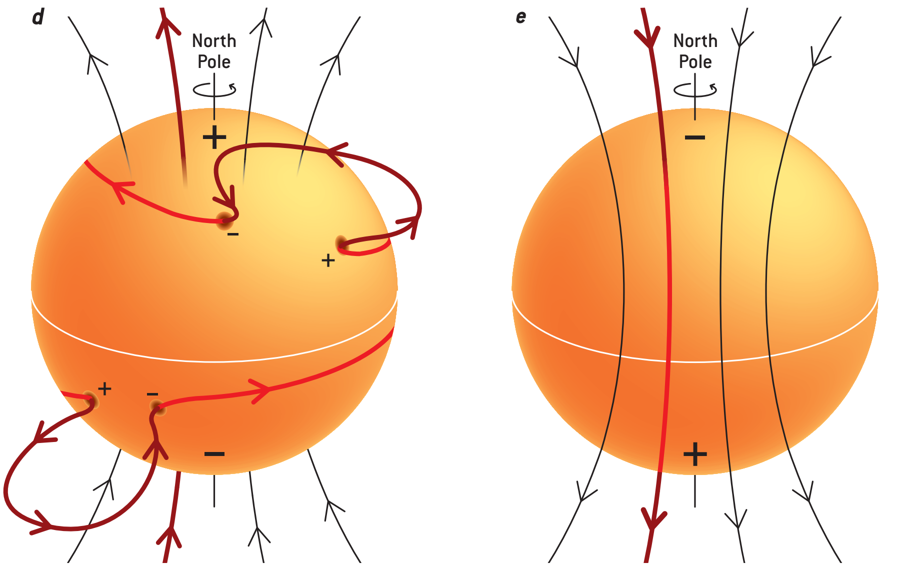
Thus, the magnetic fields are carried along with the plasma and end up getting twisted. The entwined ropes wrap together fields of opposite polarity, which tend to cancel each other. But the sun's rotation generates organizational forces that periodically sort out the tangles and create an overall magnetic field. This automatic engine, which generates magnetism from the flow of electricity, is the solar dynamo.

The dynamo has two essential ingredients: the convective cyclones and the sun's nonuniform rotation. During the mid-1800s, Richard C. Carrington, an English amateur astronomer, found that the sunspots near the equator rotate faster, by 2 percent, than those at midlatitudes. Because the spots are floating with the plasma, the finding indicates that the sun's surface rotates at varying speeds. The rotation period is roughly 25 days at the equator, 28 days at a latitude of 45 degrees and still longer at higher latitudes. This differential rotation should extend all the way through the convective zone.

Now suppose that the initial shape of the sun's field is that of a dipole oriented roughly north-south. The field lines get pulled forward at the equator by the faster rotation and are deformed in the east-west direction. Ultimately, they lie parallel to the equator and float to the surface, erupting as pairs of sunspots.

But Coriolis forces tend to align the cyclones and thereby the sunspots, which are constrained to follow the plasma's gyrations. The cyclones arrange the sunspots so that, for example, a trailing sunspot in the northern hemisphere lies at a slightly higher latitude than a leading one. As the equatorial field lines are stretched, they eventually unwind and drift outward. The trailing sunspot reaches the pole first, effectively reversing the magnetic field there. (Recall that the trailing spot has a polarity opposite that of the nearest pole.) Those field lines that initially extended far beyond the sun reconnect into loops and are blown away by the solar wind. In this manner, the overall magnetic field flips, and the cycle begins again.

There is, however, a caveat. This simple picture seems to be at odds with results from helioseismology, the science of sunquakes. The model requires the sun to rotate faster at the interior; in contrast, results from the Global Oscillation Network Group, an international collaboration of observatories, show that the rotation velocity near the equator decreases downward. Such experiments are providing accurate information on internal motions of the sun and thereby help-



moving up toward the surface and erupting as sunspot pairs (*d*). The sunspots drift toward the poles, with the trailing sunspot reaching first; as a result, the overall field flips (*e*). In addition to the dipole field above, the sun probably also has a "quadrupole" field (*opposite page, red*) whose "beating" with the dipole field was responsible for the Maunder minimum.

ing Mausumi Dikpati and Peter Gilman of the High Altitude Observatory and others to refine dynamo theory.

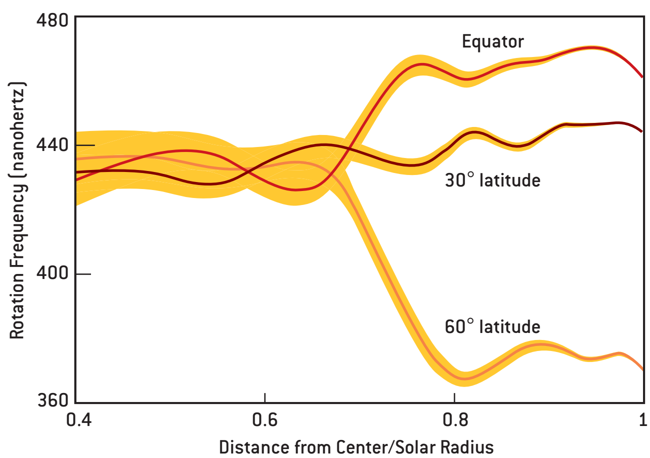
But what happened during the Maunder minimum? To explain this lull, two of us (Nesme-Ribes and Sokoloff) noted that apart from a dipole pattern, the magnetic field must also have a small quadrupole component, resembling the field of two bar magnets placed side by side. If the quadrupole oscillates at a slightly different rate than the dipole, the sunspots in one hemisphere are produced slightly earlier than those in the other hemisphere—precisely what we observe now. Furthermore, over the past four centuries, a few solar cycles showed different numbers of sunspots in the northern and southern hemispheres. This pattern seems to repeat every century or so, exactly what one would expect if the dipole "beats" with a weak quadrupole.

But suppose that the quadrupole field is as strong as the dipole. The equatorial field lines that result from stretching this combination will then cancel out in one hemisphere yet remain in the other. And the few spots that do appear will all be in one hemisphere, just as 17th-century astronomers noted during the Maunder minimum.

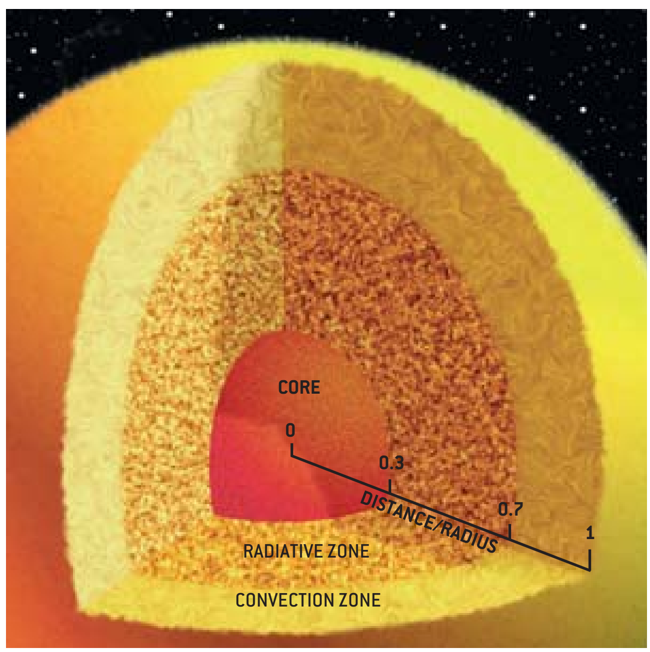
We can encapsulate the relation between the dipole and quadrupole fields in a "dynamo number" *D*. It is the product of the helicity, or spiraling motion, of the plasma and the local rate of change of rotation. When *D* is very small, the magnetic field tends to die out; as *D* increases, the quadrupole field shows up, with the dipole following. Beyond a critical value, both components of the field are steady. But as *D* increases further, the dynamo becomes periodic, increasing and decreasing; this is the regime in which the sun now lies. A weak quadrupole field, beating in phase with the dipole, leads to short and intense cycles; a stronger quadrupole field, if slightly out of phase with the dipole field, lengthens and weakens the sunspot cycle. Far beyond the critical dynamo number, chaos results.

Dynamic Stars

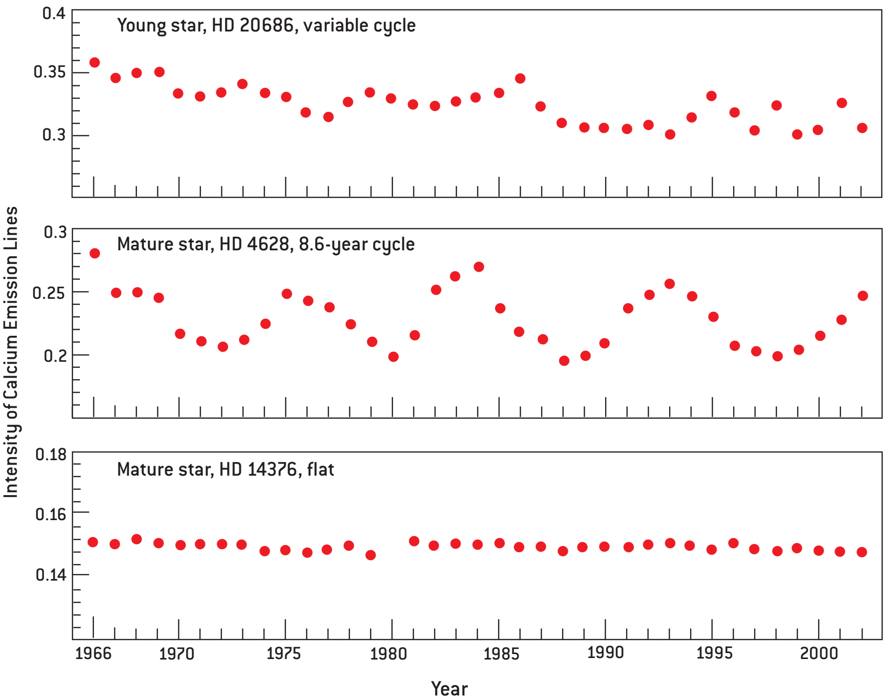
AS WE NOW KNOW, the sun's brightness increases with the magnetic activity



ROTATION of the sun's surface is faster at the equator and slower near the poles. This differential rotation (as measured by means of sunquakes by the Global Oscillation Network Group) extends through the outer layers. The sun's core, in which fusion generates the energy that ultimately powers the dynamo, most likely rotates at a constant angular velocity, like a rigid body.



PETER SAMEK *Slim Films*



INTERANNUAL MAGNETIC VARIABILITY of stars over the years is detected by way of violet calcium emission lines. Here activity from three nearby stars reveals the likely states of our own sun: the variable magnetic cycles of a young star (*top*); the steady cycles of a star at an age comparable to our sun's (*middle*); and the subsidence of a sunlike star into a Maunder-type minimum phase (*bottom*). The magnetic, and therefore sunspot, activity of other stars indicates that our sun is capable of far greater variability than it has shown in the past century.

MOUNT WILSON OBSERVATORY HK PROJECT; PETER SAMEK *Slim Films*

over a cycle: the bright plages overwhelm the dark sunspots. (Presumably, as the sun brightens and darkens, its total energy is temporarily channeled into different reservoirs—kinetic, magnetic, thermal or potential.) During the past 24 years of satellite observations, the sun's total energy output has varied roughly 0.1 percent between a brighter, magnetically active phase and a fainter, quiet one.

Because of the brevity of the satellite records, we do not know the variability of the sun's brightness over decades. Richard Willson and his colleagues at Columbia University's Center for Climate Systems Research recently found a slight, 0.05 percent increase in brightness at the observed solar minima in 1986 and 1996. Finding a longer-term value for brightness variability, however, is vital to evaluating the sun's influence on Earth. One possible way to answer this question is to examine "star spot" cycles on other stars.

It is not easy to map the features on the surface of stars. But as magnetic fields heat the outer layers of a star's atmosphere, they radiate the energy in certain spectral lines. For example, on our sun, the intensity of the two violet emission lines of calcium (having wavelengths of 396.7 and 393.4 nanometers) closely follows the strength and extent of the magnetic fields. Variations in these lines thus give us a measure of the changing surface magnetism of a star.

At Mount Wilson Observatory in 1966, Olin C. Wilson began a program of measuring the magnetic activity of roughly 100 so-called main-sequence stars—those that, like the sun, are burning hydrogen. (When the hydrogen runs out, a star expands into a red giant.) Most of these stars show obvious signs of magnetic activity, by way of variations in their violet calcium emission lines. The fluctuations vary greatly in amplitude and duration, depending primarily on the age and mass of the star.

All these stars have a dynamo number, *D*, higher than the critical value required for sustaining magnetic fields. For a young star of one or two billion years, the rotation period is fast, roughly 10 to 15 days. The resulting high value of *D* means that these young stars have erratic fluctuations in magnetic activity over intervals as short as two years and no well-defined cycles. The fluctuations sometimes repeat, however, having periods between two and 20 years or so that lengthen with age.

But as a star ages, it slows down—because its angular momentum is carried off by the magnetic wind—and *D* falls. Then a consistent dynamo cycle begins to appear, with a period of about six to seven years and sometimes even with two independent periods. Later on—for an even lower *D*—one period starts to dominate, lengthening with age from eight to 14 years. In addition, there are occasional Maunder minima. If rotation were to slow further, in the very oldest stars, we predict that the magnetic field should be steady. The Wilson sample contains a few very old stars, but they still show cycles, indicating that the steady dynamo would not be reached in 10 billion years—soon after which they will expand into red giants.

To focus on the solar dynamo, Baliunas and her collaborators restricted Wilson's broad sample of stars to those similar to our sun in mass and age. That group currently comprises 10- to 20-year records of 150 stars, depending on the criteria defining similarity to the sun. Many of these stars show prominent cycles similar in amplitude and period to those of the sun. About one quarter of

THE AUTHORS

ELIZABETH NESME-RIBES, SALLIE L. BALIUNAS and *DMITRY SOKOLOFF* all have been active in unraveling connections between the sun's variations and Earth's climate. Nesme-Ribes, who recently passed away, was an astronomer at the Paris Observatory and the National Center for Scientific Research in France. Apart from studying the solar dynamo, she conducted extensive searches into the 17th-century archives on sunspots at her home institution. Baliunas is a scientist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Mass. She observes the variations of sunlike stars at the Mount Wilson Observatory in Pasadena, Calif. Sokoloff is professor of mathematics in the department of physics at Moscow State University in Russia.

the records show that the stars are in a dead calm, suggesting a phase similar to our sun's Maunder minimum. This finding implies that sunlike stars spend a quarter of their lives in a lull—consistent with radiocarbon results.

We may have captured one star, HD 3651, in transition between the cyclic and Maunder minimum phases. HD 3651's cycles have weakened and lengthened dramatically (from 12 to 15 years) as its surface activity has rapidly dropped to very low levels. Sunlike stars such as HD 3651, observed over a few decades, offer us "snapshots" of the range of variability that our sun—and we—might experience over a timescale of centuries.

The brightness of these sunlike stars can also be compared with their magnetic activity. In 1984 thorough and precise photometric observations of some of the Wilson stars began at the Lowell and Sacramento Peak observatories. Since 1992 those of us at the Smithsonian Astrophysical Observatory and at Tennessee State University have used automated telescopes to observe some of these stars. Nearly all the older stars, like the sun, are brightest near the peak of the activity cycle. Some stars vary as little as our sun does—only 0.1 percent over the last 11-year cycle—but other sunlike stars have varied by as much as 0.6 percent in a cycle. Thus, the sun's current changes might be a poor indicator of the full range of fluctuations of which it is capable.

Over the decades, researchers have inferred the evolutionary history of a sunlike star from the collection of stellar records. A young star has a relatively rapid rotation period of several days and high, irregular levels of surface magnetism. Changes in brightness of several percent accompany the magnetic variations. The young star is, however, darkest during the peak of magnetic activity, presumably because the dark spots are so large that they, not the plages, dominate. As the sunlike star ages, it rotates more slowly, and the magnetic activity decreases. Maunder minima appear in these "older" stars; furthermore, radiance now peaks at sunspot maximum, with fluctuations of 1 percent or less over a cycle.

Influencing Earth

THE STAR-SPOT RESULTS point to a change in brightness of at least 0.4 percent between the cyclic phase and the Maunder minimum phase. This value corresponds to a decrease in the sun's net energy input of one watt per square meter at the top of Earth's atmosphere. Simulations performed at the Laboratory of Dynamic Meteorology in Paris and elsewhere suggest that such a reduction, occurring over several decades, is capable of cooling Earth's average temperature by 1 to 2 degrees C—enough to explain the observed cooling during the Maunder minimum.

But greenhouse gases generated by human activity may be warming our planet, by trapping heat that would otherwise radiate into space. This warming is equivalent to Earth's receiving radiation of two watts per square meter at the surface. The sun has apparently delivered to Earth no more or less than 0.5 to 1.0 watt per square meter over the past few centuries. Therefore, if direct heating is the only way in which the sun affects Earth's climate and is presumed to act the same as the enhanced greenhouse effect, the added greenhouse gases should already be dominating the climate, washing out any correlation with the sun's varying activity.

The sun's energy reaches Earth as radiation and particles and varies over many frequencies and periods. Yet the link between climate and solar magnetic activity seems rather persistent. The length of the sunspot cycle, for example, correlates closely with global temperatures over the past 240 years. Minima in solar magnetism, as traced by radiocarbon dating in tree rings and beryllium 10 in ice cores, coincide with roughly 1,500-year intervals of cooler climate, seen in environmental changes going back 10,000 years. In addition, the sunspot cycle correlates with stratospheric wind patterns, for reasons not yet well understood. All these pieces of evidence induce some scientists, including us, to argue that the sun must be influencing Earth by powerful indirect routes as well as the obvious ones.

UNRAVELING THE INFLUENCES OF THE SUN provides vital information on the role OUR STAR PLAYS IN CLIMATE CHANGE.

Variations in the sun's ultraviolet radiation, for example, may be changing the ozone content of our upper atmosphere, as well as its dynamics. Recent simulations also indicate that winds in the lower stratosphere can convey variations in solar radiance down to the troposphere, where they interact more directly with the weather system. Such matters are now the subject of vigorous debate. Unraveling the ways in which the sun warms Earth provides vital information concerning the role played by humankind—and the role played by the sun—in the process of climatic change. SA

MORE TO EXPLORE

The Variable Sun. Peter V. Foukal in *Scientific American,* Vol. 262, No. 2, pages 34–41; February 1990.

The Paradox of the Sun's Hot Corona. Bhola N. Dwivedi and Kenneth J. H. Phillips in *Scientific American,* Vol. 284, No. 6, pages 40–47; June 2001.

The Maunder Minimum and the Variable Sun-Earth Connection. Willie Wei-Hock Soon and Steven H. Yaskell. World Scientific Publishing, 2003.

THE FURY *of*

SHOCK WAVES FROM THE SUN CAN TRIGGER SEVERE TURBULENCE IN THE SPACE AROUND EARTH, ENDANGERING SATELLITES AND ASTRONAUTS IN ORBIT. A NOVEL SPACECRAFT IS SHOWING HOW SPACE STORMS DEVELOP BY JAMES L. BURCH

The tempest began on a date known for its violent events: Bastille Day, the anniversary of the beginning of the French Revolution. On the morning of July 14, 2000, the Space Environment Center in Boulder, Colo., detected a warning sign from the GOES-8 satellite, which monitors x-rays from the sun as well as weather conditions on Earth. At 10:03 Universal Time the center's forecasters saw a sharp jump in the intensity of x-rays emanating from active region 9077, a section of the sun's surface that had been roiling for the past week. The data indicated the onset of a solar flare, a brief but powerful burst of radiation.

The flare, which reached its maximum intensity at 10:24 UT, was also sighted by the Solar and Heliospheric Observatory (SOHO), a spacecraft stationed between the sun and Earth, about 1.5 million kilometers from our planet. Half an hour later, as the flare was waning, SOHO observed an even more ominous phenomenon: a bright, expanding cloud that surrounded the sun like a halo. It was a coronal mass ejection (CME), an eruption in the corona—the sun's outer atmosphere—throwing billions of tons of electrically charged particles into interplanetary space. The halo signature meant that the particles were heading directly toward Earth, at an estimated speed of 1,700 kilometers per second.

As the CME plowed into the solar wind—the flow of ionized gas continuously streaming from the sun—it created a shock wave that accelerated some charged particles to even higher velocities. In less than an hour a deluge of high-energy protons struck SOHO, temporarily blinding its instruments. The bombardment also damaged the spacecraft's solar arrays, causing a year's worth of degradation in 24 hours. But this torrent of particles was only the leading edge of the squall. The CME-driven shock wave arrived the next day, slamming into Earth's magnetic field at 14:37 UT. The impact marked the start of a severe geomagnetic storm, whose full fury was unleashed by the arrival, a few hours later, of the CME itself. According to the index of geomagnetic activity used by the Space Environment Center, the Bastille Day storm was the largest such event in nearly a decade.

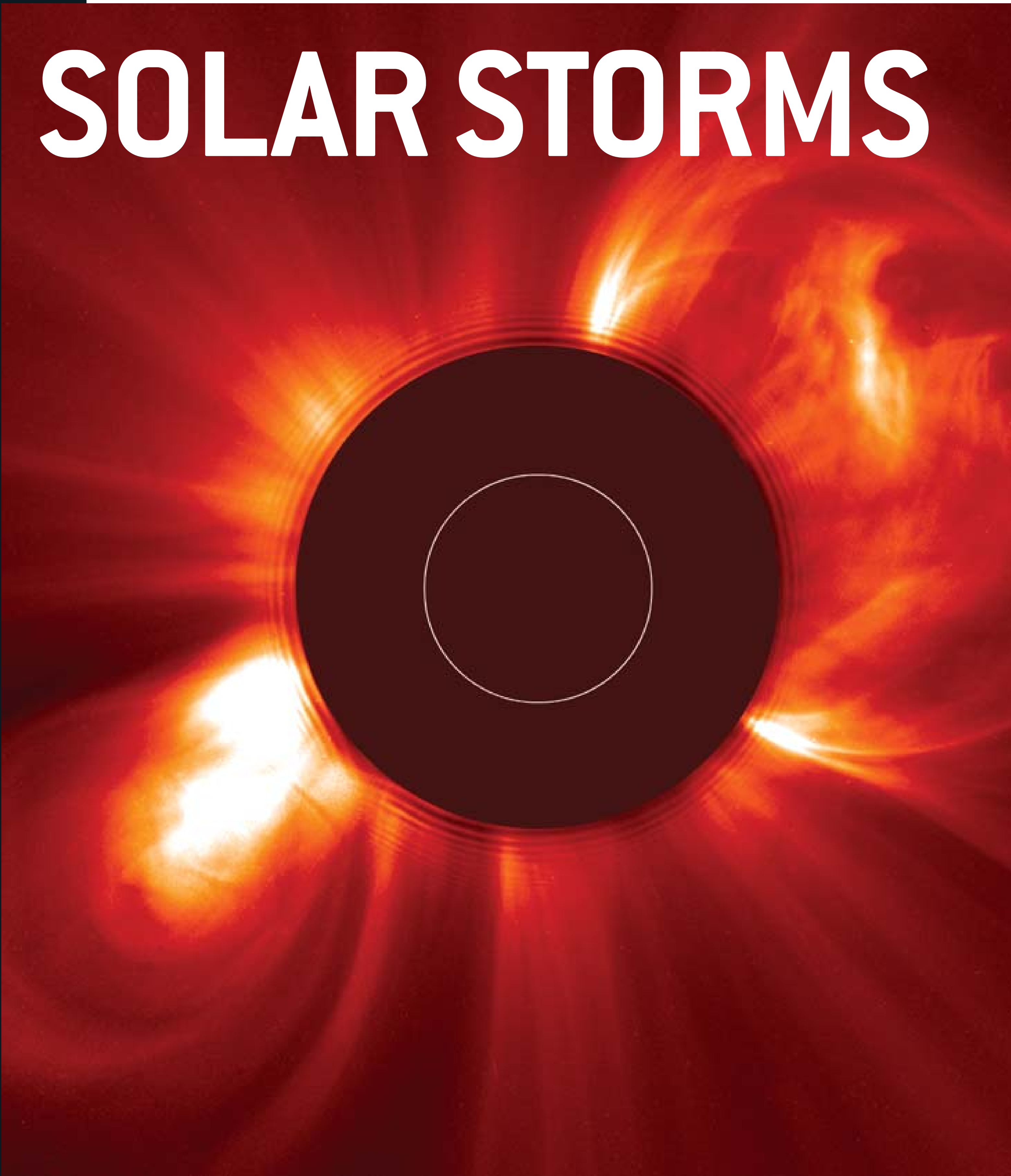
Most people on the ground were completely unaware of the celestial fireworks, but researchers were following the tempest closely, collecting data from instruments on Earth and in space. Among the satellites tracking the storm was the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE), which NASA had launched just four months earlier. IMAGE is the first satellite dedicated to obtaining global images of the magnetosphere, the region of space protected by Earth's magnetic field. By providing an overall picture of the activity in the magnetosphere, IMAGE does for space what the first weather satellites did for Earth's atmosphere.

In 1996 I had been selected by NASA to lead a team of engineers and scientists in developing the IMAGE spacecraft and analyzing the data that it transmits. As the Bastille Day storm progressed, we received astounding images of ions circling Earth and pictures of the brilliant aurora borealis—the northern lights—that occurred when the charged particles struck the upper atmosphere. The results will help scientists answer long-standing questions about how CMEs and the solar wind interact with Earth's magnetosphere. The findings may also have practical applications. Space storms can disable satellites, threaten the safety of astronauts and even knock out power grids on the ground [*see box on page 44*]. Indeed, the Bastille Day storm caused the loss of the Advanced Satellite for Cosmology and Astrophysics, an x-ray observatory launched in 1993 by the Japanese space research agency. In hopes of mitigating such effects in the future, scientists are keenly interested in improving the accuracy of space weather forecasts.

VIOLENT ERUPTION in the sun's outer atmosphere on November 8, 2000, spewed billions of tons of charged particles toward Earth. The event was observed by the Solar and Heliospheric Observatory (SOHO); the spacecraft's coronagraph uses a disk (*dark circle*) to block direct light from the sun (*white circle*) so that its atmosphere can be seen.

Updated from the April 2001 issue

SOLAR STORMS



It's Not the Heat or the Humidity

LIKE WEATHER ON EARTH, weather in space is extremely variable. Conditions can turn from quiet to stormy in a matter of minutes, and storms can last for hours or days. And just as terrestrial weather changes with the seasons, space weather, too, follows its own cycles. Solar magnetic activity, which causes flares and CMEs, rises and falls every 11 years, and therefore geomagnetic storms follow the same pattern. The Bastille Day storm took place during the solar maximum, the most active part of the cycle. Space weather also varies, though less dramatically, according to the sun's 27-day rotation period, as alternating streams of fast and slow solar wind sweep past our planet.

Space weather, however, arises from physical processes that are profoundly different from those responsible for terrestrial weather. The medium for terrestrial weather is the dense, electrically neutral gas in Earth's lower atmosphere, whose behavior is governed by the laws of fluid dynamics and thermodynamics. The medium for space weather, in contrast, is plasma—very sparse gases consisting of equal numbers of positively charged ions and negatively charged electrons. Unlike the atoms and molecules of the atmosphere, these plasma particles are subject to the influence of electric and magnetic fields, which guide and accelerate the particles as they travel through the space surrounding Earth.

Terrestrial weather is driven by the sun's radiation as it heats Earth's atmosphere, oceans and landmasses. But in the magnetosphere, weather results from the interaction between Earth's magnetic field and the solar wind. The solar wind has its own magnetic field, which travels with the outflowing plasma into interplanetary space. As the wind carries this interplanetary magnetic field (IMF) away from the sun, the field lines typically stretch out so that they are directed radially (pointing toward or away from the sun). Under certain conditions, however, the IMF's field lines can tilt out of the equatorial plane of the sun, taking on a northward or southward component. A strong and sustained southward IMF direction is a key factor in triggering geomagnetic storms. The IMF was oriented southward for many hours during the Bastille Day storm.

Protons are the dominant constituents of the solar wind, accounting for about 80 percent of its total mass. Helium nuclei make up about 18 percent, and trace quantities of heavier ions are also present. The average density of the solar wind at Earth's orbit is nine protons per cubic centimeter. The wind's average velocity is 470 kilometers per second, and the average strength of the IMF is six nanoteslas (about one five-thousandth the strength of Earth's magnetic field at the planet's surface). These properties, along with the orientation of the IMF, are highly variable, and it is this variability that ultimately explains the changing weather in space.

All the bodies in the solar system are immersed in the solar wind, which continues flowing outward until it encounters the ionized and neutral gases of interstellar space. The solar wind does not impinge directly on Earth and its atmosphere, however. The planet is shielded by its magnetic field, which forms a kind of bubble within the stream of charged particles emanating from the sun. The shape of this cavity—the magnetosphere—is determined by the pressure of the solar wind and by the IMF [*see illustration on opposite page*]. The wind compresses Earth's magnetic field on the dayside of the planet—the side facing the sun—and stretches the field on the nightside to form a long tail resembling that of a comet. This magnetotail extends more than one million kilometers past Earth, well beyond the orbit of the moon.

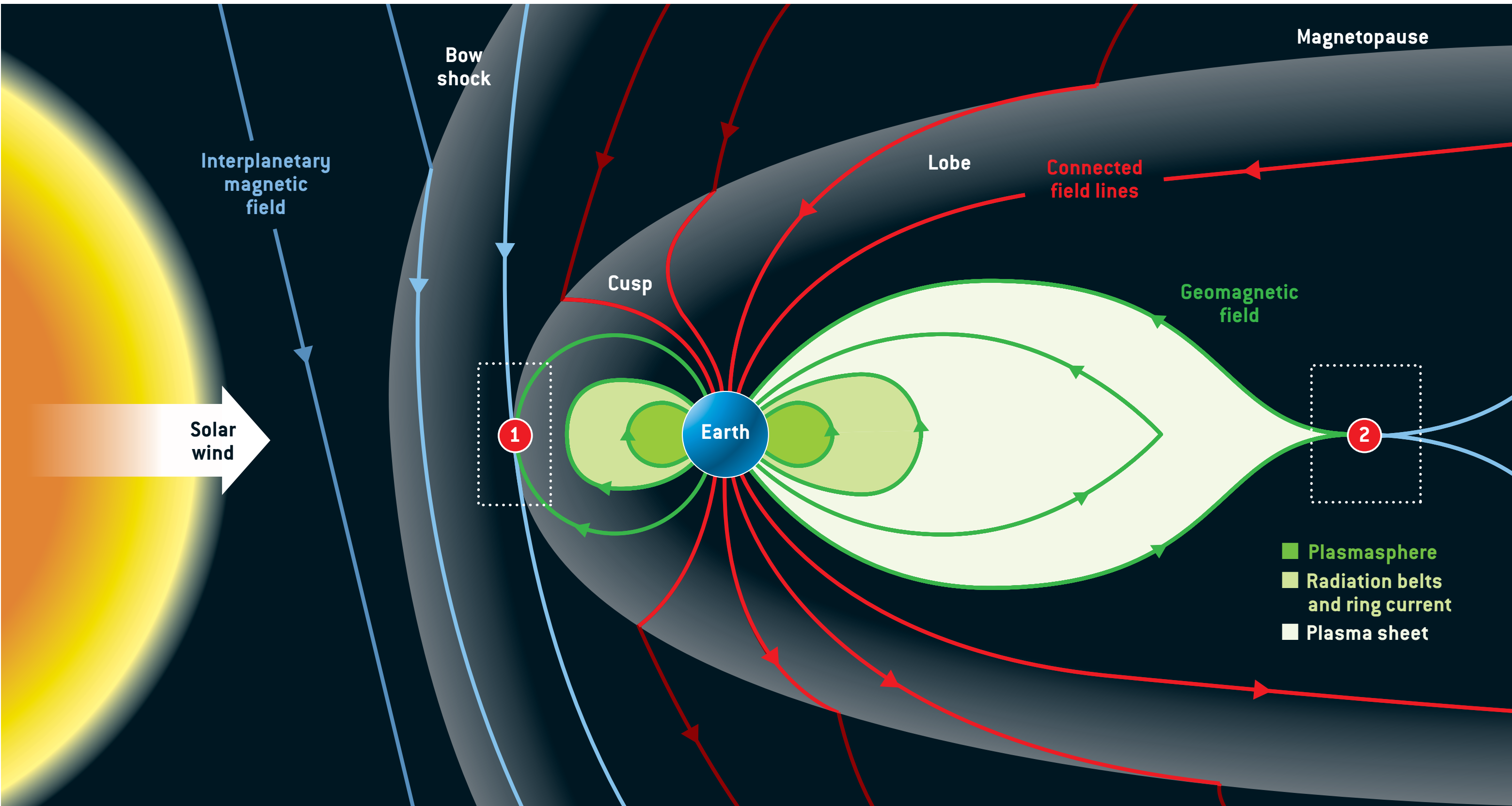
Between the solar wind and the magnetosphere is a thin boundary called the magnetopause, where the pressure of the geomagnetic field balances that of the solar wind. On Earth's dayside, this boundary is usually located about 64,000 kilometers from the planet's center, although the distance varies with changes in the solar-wind pressure. When the pressure increases, as occurred during the Bastille Day storm, the dayside magnetopause is pushed closer to Earth, sometimes by as much as 26,000 kilometers.

THE EFFECTS OF SPACE STORMS

During geomagnetic storms, charged particles swirl around Earth and bombard the upper atmosphere, particularly at the higher latitudes. The gusts of particles can have severe effects on:

- **POWER GRIDS.** As electrons cascade toward Earth, they create a strong current in the upper atmosphere called the auroral electrojet. This current causes fluctuations in the geomagnetic field, which can induce electrical surges in power lines on the ground. During an intense geomagnetic storm on March 13, 1989, a surge knocked out the Hydro-Quebec power grid, plunging large parts of Canada into darkness.
- **SATELLITES.** When particles strike a satellite, the craft's surface becomes charged. This buildup sometimes triggers sparks that can short-circuit the satellite's electronics. Also, space storms heat Earth's atmosphere, causing it to expand. If the atmospheric density at a satellite's orbit becomes high enough, friction will slow the craft and drag it downward. This process led to the premature fall of Skylab in 1979.
- **ASTRONAUTS.** A severe storm could expose the International Space Station to protons that could penetrate a spacesuit or even the station's walls. To protect its astronauts, NASA monitors space weather data. If an oncoming storm poses a risk, NASA will postpone or cancel any planned space walks and may order the astronauts to seek shelter in a shielded part of the station.

COURTESY OF NASA/ESA, SOHO/LASCO CONSORTIUM (*preceding page*)



DISTURBANCES IN THE MAGNETOSPHERE occur when the interplanetary magnetic field (IMF) carried by the solar wind turns southward. In a process called reconnection, the field lines of the IMF connect with the northward geomagnetic field lines at the dayside of the magnetopause (*1*). Energy and particles from the solar wind rush into the magnetosphere, enlarging its northern and southern lobes and narrowing the plasma sheet. Then the geomagnetic field lines themselves reconnect (*2*), accelerating ions and electrons toward Earth.

JOE ZEFF

Just as the passage of a supersonic jet through the atmosphere produces a shock wave, the encounter of the solar wind with the magnetosphere forms a shock wave—known as the bow shock—some 13,000 kilometers upstream (that is, closer to the sun) from the dayside magnetopause. The region of solar-wind plasma between the bow shock and the magnetopause is known as the magnetosheath. Because of its passage through the shock, the magnetosheath plasma is slower, hotter and more turbulent than the plasma farther upstream.

Satellite detectors have indicated that the charged particles surrounding Earth are a mix of plasma from the magnetosheath (mostly protons) and plasma that flows out of the upper atmosphere above the North and South poles (mostly protons and oxygen ions). The proportions of this mix vary according to whether the magnetosphere is in a quiet or a disturbed state. During geomagnetic storms, charged particles bombard Earth at high latitudes. The resulting electric currents heat the upper atmosphere, pumping increased amounts of protons and oxygen ions into the magnetosphere. This plasma is stored, together with the solar-wind plasma that has gained entry into the magnetosphere, in a great reservoir known as the plasma sheet, which extends for tens of thousands of kilometers on Earth's nightside.

At the heart of the study of space weather is a question: How do changes in the solar wind affect conditions in the space surrounding Earth? In other words, how can the wind overcome the barrier of the geomagnetic field and drive the motions of the plasma inside the magnetosphere?

Blowing in the Solar Wind

THE MOST ACCEPTED ANSWER was proposed in 1961 by British physicist James W. Dungey. In this process, called magnetic reconnection, the field lines of the IMF become temporarily interconnected with the geomagnetic field lines on the dayside of the magnetopause [*see illustration above*]. This tangling of the field lines allows large amounts of plasma and magnetic energy to be transferred from the solar wind to the magnetosphere.

Reconnection is most efficient when the IMF has a component that is directed southward—that is, opposite to the northward direction of Earth's magnetic field at the dayside of the magnetosphere. Under these circumstances, reconnection takes place along a wide equatorial belt, opening up nearly the entire outer boundary of the magnetosphere to the solar wind. For other orientations of the IMF, reconnection still happens, but it may be more localized in the higher latitudes, where the released energy mainly flows around the magnetosphere rather than into it.

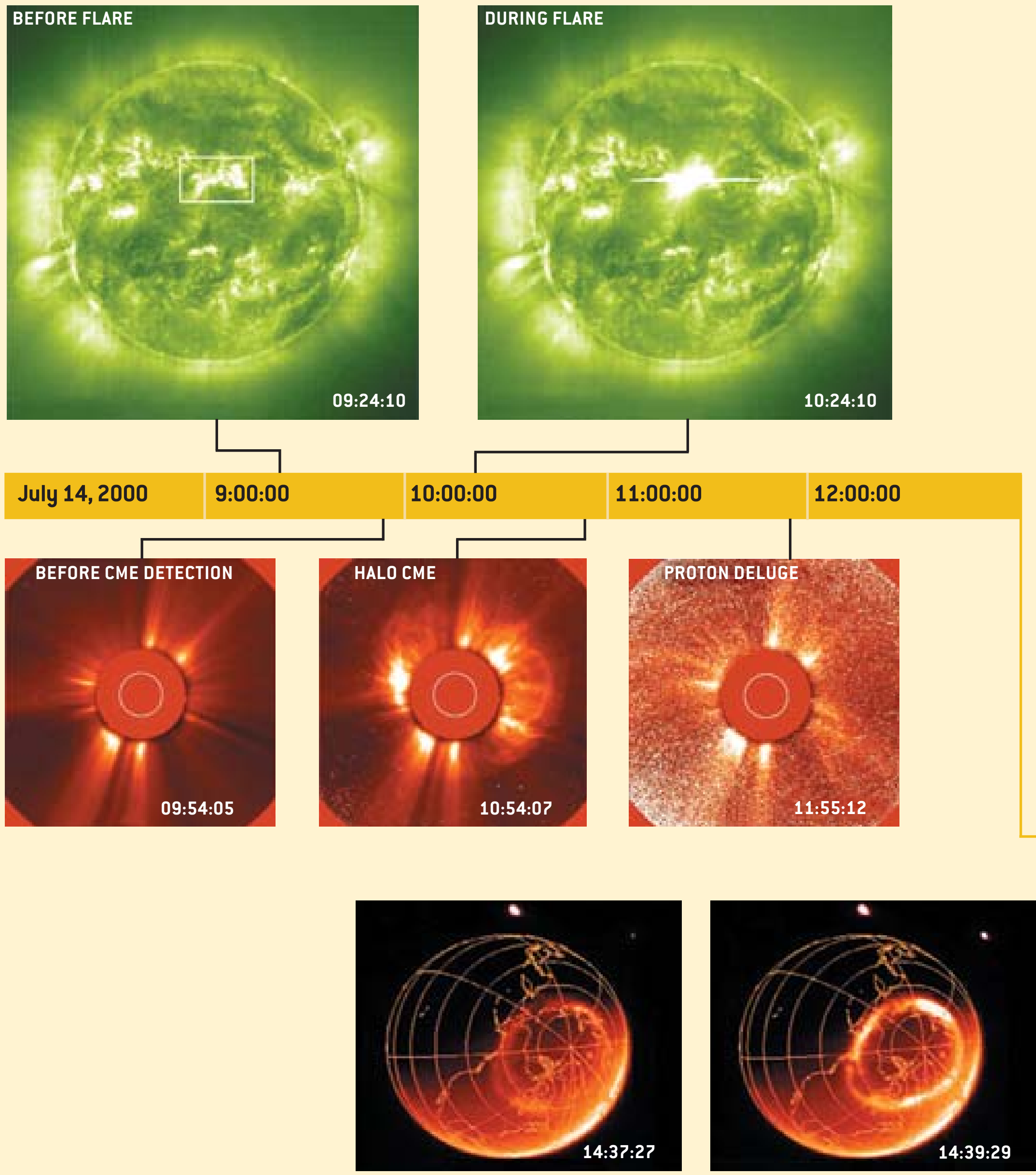
The transfer of magnetic energy from the solar wind radically alters the shape of the magnetosphere. When reconnection is initiated on the dayside magnetopause, the interconnected IMF and geomagnetic field lines are swept back by the solar wind over Earth's poles, pouring energy into the northern and southern lobes of the long magnetotail on the nightside. As the lobes swell with the added magnetic energy, the plasma sheet that lies between them begins to thin. The process continues until the field lines of the northern and southern lobes, which have opposite directions, are pressed together and themselves reconnect.

This second reconnection releases the solar wind's magnetic field, enabling it to continue its flow through the solar system. At the same time, it allows Earth's magnetic field lines, which have been stretched tailward during the loading of the lobes, to snap back into their normal configuration. The abrupt movement of the field lines heats and accelerates the ions and electrons in the plasma sheet, injecting them into the inner part of the magnetosphere. Some of these particles, traveling along geomagnetic field lines, dive into the upper atmosphere above Earth's poles, stimulating auroral emissions at x-ray, ultraviolet, visible and radio wavelengths as they collide with oxygen atoms and nitrogen molecules. The entire sequence of events—from dayside reconnection to nightside reconnection to auroras—is known as a magnetospheric substorm.

In addition to transferring magnetic energy to the tail lobes, dayside reconnection also intensifies the electric field across the magnetotail. The stronger field, in turn, increases the flow of ions and electrons from the plasma sheet to the inner magnetosphere. This flow feeds into Earth's ring current, which is carried by charged particles circling the planet above its equator at altitudes between 6,400 and 38,000 kilometers. During longer periods of dayside reconnection—which occur when the IMF's orientation remains consistently southward—the sustained enhancement of the earthward plasma flow greatly increases the

A SOLAR STORM IN ACTION

First warning of the Bastille Day storm was a solar flare on July 14, 2000. Images of the sun from SOHO's Extreme Ultraviolet Imaging Telescope (*top*) show active region 9077 (*in white box*) before and during the flare. At about the same time, SOHO's coronagraph observed a coronal mass ejection (CME) that soon deluged the spacecraft with high-speed protons, temporarily blinding its instruments (*middle*). The shock wave and CME slammed into Earth's magnetic field the next day, triggering auroras observed by the IMAGE spacecraft's Wideband Imaging Camera (*bottom*) and a sharp drop in geomagnetic field strength at the planet's surface (*on opposite page, middle*). In this graph, called the disturbance storm time index, zero represents the normal surface field strength. As the storm progressed, IMAGE's High Energy Neutral Atom instrument monitored the waxing and waning of the ring current around Earth's equator (*on opposite page, top*).



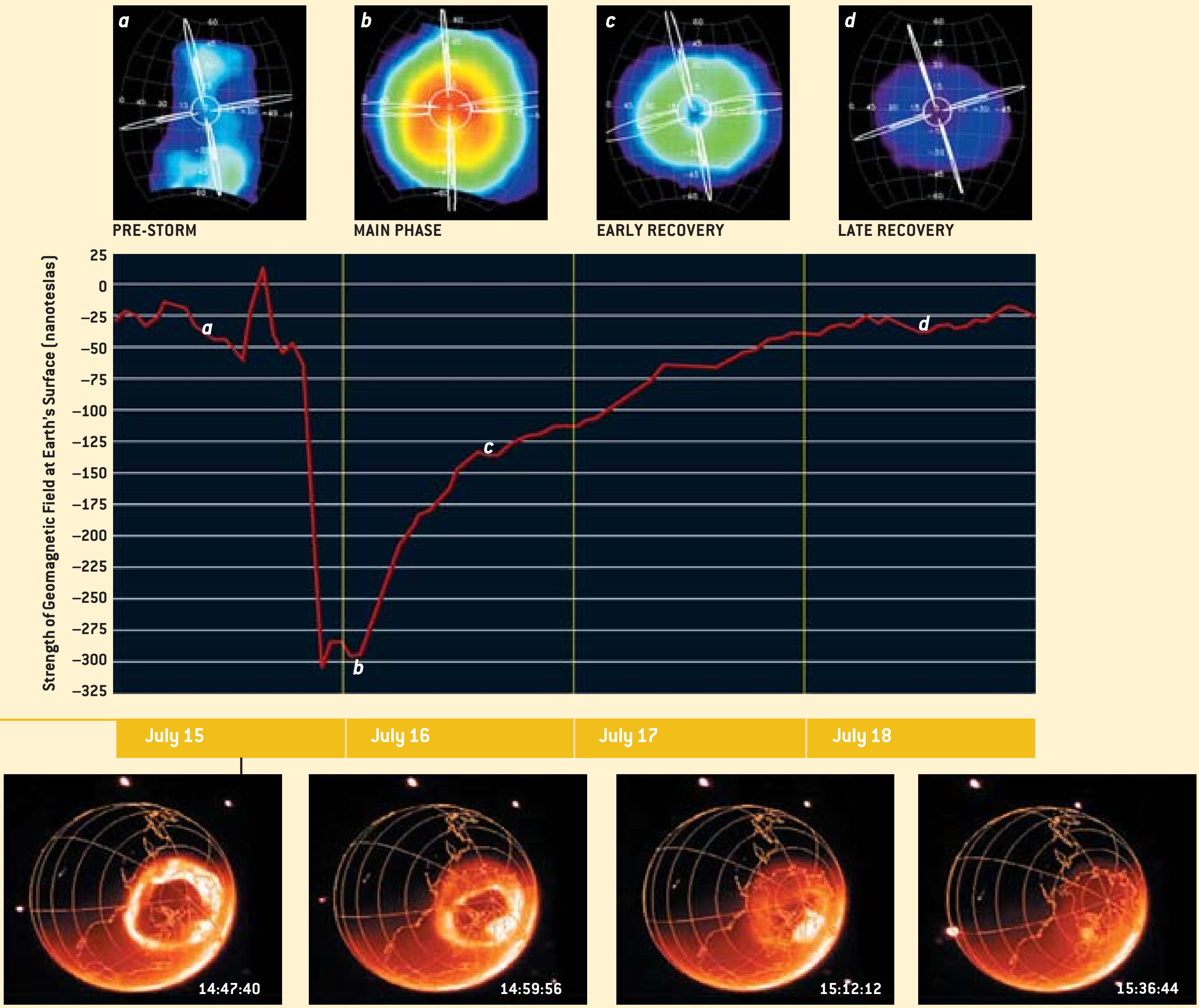
NASA/ESA, SOHO (*solar images*); NASA/IMAGE FUV TEAM (*auroral images*); NASA/IMAGE HENA TEAM (*ring current images*); WORLD DATA CENTER FOR GEOMAGNETISM, KYOTO, JAPAN (*graph*) number and energies of the charged particles in the ring current. An extended period of southward IMF can also lead to a rapid succession of substorms, each of which injects more particles toward Earth. The resulting growth in strength of the ring current is the classic hallmark of a full-fledged geomagnetic storm.

Here Comes the Sun

THE ORIENTATION OF THE IMF turns southward quite frequently, so magnetospheric substorms are fairly common: on average, they happen a few times every day and last for one to three hours. But major geomagnetic storms such as the Bastille Day event are much rarer. Although they can occur anytime during the 11-year solar cycle, they are most frequent in the solar maximum period.

Until the early 1990s, it was widely believed that solar flares triggered geomagnetic storms. Space and solar physicists, however, had been assembling evidence that pointed strongly to another culprit, and in 1993 John T. Gosling of Los Alamos National Laboratory wove the various threads of evidence together in an article in the *Journal of Geophysical Research* that challenged the "solar flare myth." Gosling set forth a compelling argument for the central role of coronal mass ejections in setting off large geomagnetic storms. Scientists still do not know what causes these violent eruptions in the sun's corona, but the phenomenon most likely involves a reconfiguration of the magnetic field lines there. CMEs are often, but not always, associated with solar flares.

Not all CMEs cause geomagnetic storms. Most of the eruptions are not directed at Earth, and of those that are, only about one in six is "geoeffective"—strong enough to trigger a storm.



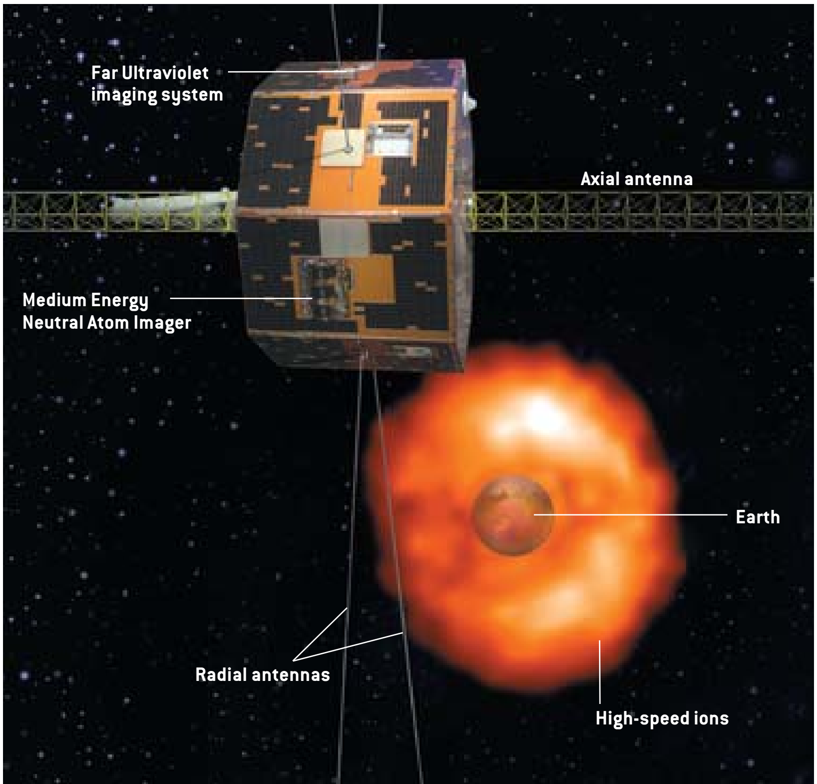


IMAGE SPACECRAFT is shown above a cloud of high-speed ions circling Earth in this illustration. Researchers produced the ion image using data from the satellite's High Energy Neutral Atom Imager (on the side opposite the Medium Energy Neutral Atom Imager). IMAGE's Radio Plasma Imager charts the clouds of charged particles by sending pulses of radio waves from two 10-meter-long axial antennas and four 250-meter-long radial antennas. Although the spacecraft's body is only 2.25 meters wide, the antennas make IMAGE one of the biggest sensors ever flown.

The primary factor is the CME's speed relative to that of the solar wind. Only fast CMEs are geoeffective. Why? When fast CMEs plow through the slower solar wind, they produce interplanetary shock waves, which are responsible for the high-energy particle showers and the severe deformations of Earth's magnetic field. Even more important, a fast-moving CME compresses the solar wind ahead of it, thereby increasing the strength of the magnetic field in the compressed region and in the front part of the CME itself. Moreover, this draping of the field around the CME tends to tilt the IMF more along the north-south direction, which causes a stronger reconnection when the IMF encounters Earth's magnetic field.

A weaker kind of geomagnetic storm occurs during the declining phase of the solar cycle and near the solar minimum period. These disturbances, which tend to recur in phase with the sun's 27-day rotational period, are set off by the interaction between fast solar winds emanating from holes in the corona and slower winds arising from the sun's equatorial streamer belt. Although CMEs are not the primary cause of recurrent geomagnetic storms, they may contribute to their intensity.

With the launch of IMAGE in 2000, researchers finally had the means to obtain global views of the minute-by-minute progress of a large geomagnetic storm. The satellite travels in an elliptical polar orbit, with its altitude varying from 1,000 to 46,000 kilometers. This orbit allows the craft to observe a large part of the magnetosphere, including the dayside magnetopause, the inner reaches of the magnetotail and the polar cusp regions, which are the main entryways for the particles from the solar wind.

THE AUTHOR

JAMES L. BURCH is vice president of the Space Science and Engineering Division of the Southwest Research Institute in San Antonio, Tex., and principal investigator for the IMAGE mission. Burch earned his Ph.D. in space science from Rice University in 1968. His main research interests are auroral processes, magnetic reconnection and magnetospheric imaging. He is a Fellow of the American Geophysical Union and former chairman of the Committee on Solar and Space Physics of the National Research Council.

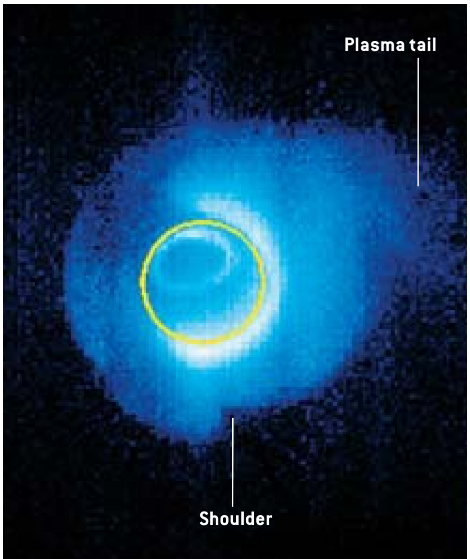
The Perfect Solar Storm

IMAGE'S INSTRUMENTS are designed to observe the magnetosphere's plasmas, but they do so in different ways. The craft contains three Energetic Neutral Atom (ENA) imagers that indirectly measure ion flows. When a fast-moving ion (such as an oxygen ion) collides with one of the neutral hydrogen atoms in the magnetosphere, it sometimes strips away the hydrogen atom's lone electron and becomes an energetic neutral atom. Because this atom no longer carries a charge, it does not have to move along the geomagnetic field lines. Instead it travels in a straight path from where it was created. The ENA imagers record the number and energies of the neutral atoms coming from a particular region, and researchers can deduce from those data the mass, speed, direction and density of the ions in that region.

The satellite also carries several instruments that monitor emissions in the ultraviolet part of the spectrum. The Extreme Ultraviolet (EUV) imager measures the density of singly ionized helium atoms in the plasmasphere—a doughnut-shaped region of the inner magnetosphere containing low-energy plasma—by detecting the solar ultraviolet light that they absorb and then reradiate. The Far Ultraviolet (FUV) imaging system has two instruments for viewing auroras—the Wideband Imaging Camera and the Spectrographic Imager—as well as the Geocorona Photometers for detecting emissions from neutral hydrogen atoms. Last, the Radio Plasma Imager sends out pulses of radio waves that bounce off clouds of charged particles. It works like a state trooper's radar gun: the returning radio signals convey information about the direction, speed and density of the plasma clouds.

During the Bastille Day event in 2000, IMAGE began

NASA/IMAGE HENA TEAM (*ion cloud*); E. MAYFIELD (*spacecraft*) recording the storm's effects less than two minutes after the CME-driven shock wave hit Earth's magnetic field on July 15. The Wideband Imaging Camera sent back stunning photographs of the aurora borealis triggered by the compression of the field [*see bottom illustrations on pages 46 and 47*]. A movie created from the images shows a sudden dramatic brightening of a ring above the Arctic region—the auroral oval—with brilliant emissions racing like brushfires toward the North Pole. The aurora quieted less than an hour after the storm began but flared up again when a second shock hit at about 17:00 UT. Powerful substorms followed, as energy stored in the magnetotail was explosively released into the upper atmosphere. Substorms and the attendant auroral displays continued through the rest of July 15 and into the morning of July 16.



PLASMASPHERE appears as a pale blue cloud of ions surrounding Earth (*yellow circle*) in this picture obtained by IMAGE during a moderate geomagnetic storm on May 24, 2000. The results confirmed the existence of the hypothesized plasma tail and revealed an unexpected structure that was dubbed the "shoulder."

During the storm's main phase, which began four hours after its start, the magnetic field strength on Earth's surface fell precipitously, dropping 300 nanoteslas below its normal value. This phenomenon, the defining feature of geomagnetic storms, resulted from the rapid growth of the ring current. IMAGE's Energetic Neutral Atom imagers produced vivid pictures of this flow of ions and electrons around Earth as it reached its peak on July 16 and then began to diminish [*see top illustrations on page 47*]. Once the transfer of energy from the solar wind abates, the flow of plasma into the inner magnetosphere slows, and ions are lost from the ring current more rapidly than new ones arrive. As the current weakens, the magnetic field on Earth's surface rebounds. The return to pre-storm levels usually takes one to a few days but may require more than a month in the case of major tempests.

Geomagnetic storms also change the shape of the plasmasphere. The enhanced flow of plasma from the magnetotail toward Earth erodes the plasmasphere by sweeping its charged particles toward the dayside magnetopause. When a storm subsides, the plasmasphere is refilled by ion outflow from the upper atmosphere. Scientists had hypothesized from modeling studies that the eroded material from the plasmasphere would form a long tail extending to the dayside magnetopause and that from there, it would become lost in the solar wind. Global images of the plasmasphere from IMAGE's EUV instrument have confirmed this decades-old hypothesis [*see illustration above*]. At the same time, the images have revealed structures in the plasmasphere that raise new questions about its dynamic response to magnetospheric disturbances. Perhaps these questions will be answered: NASA has extended IMAGE's mission through October 2009.

On the Horizon

ALTHOUGH IMAGE has opened a new window on the magnetosphere, our view of space weather is still imperfect. Unlike terrestrial clouds, the clouds of plasma seen by IMAGE are completely transparent: nothing is hidden from sight, but depth perception is lacking. Thus, there will always be the need for satellites that make local measurements of the plasmas, as well as the fields and currents that govern their motion.

The next step for space weather observation will involve clusters of satellites acting like hurricane-hunter planes—they will go where the action is. The European Space Agency began conducting the first such mission, called Cluster II, in the summer of 2000. (A predecessor mission, Cluster I, was destroyed in a rocket explosion just after liftoff in 1996.) Cluster II consists of four closely grouped identical spacecraft designed to probe turbulent plasma phenomena in the magnetosphere and nearby solar wind. NASA is also planning a cluster mission for launch in 2010. The Magnetospheric Multiscale mission will study reconnection, charged particle acceleration, and turbulence at the dayside magnetopause and at specific locations in the magnetotail where substorms are triggered.

The space agencies are considering even more ambitious missions involving constellations of satellites: dozens of tiny spacecraft that will monitor large regions of space, just as the global weather networks now monitor conditions on Earth. The first constellations would most likely observe the inner magnetosphere and the dayside magnetopause, with each cake-size craft recording the basic characteristics of the plasmas and magnetic fields.

Earth's magnetosphere is at once protective and dangerous. Its strong magnetic field shields humans from penetrating radiation that would otherwise be lethal. But the field is not strong enough to ward off the most powerful shock waves from the sun. Like the tornado belt or the tropical cyclone zone, the magnetosphere is a place of sudden storms. And that's why storm watchers such as the IMAGE satellite are so important. SA

MORE TO EXPLORE

From the Sun: Auroras, Magnetic Storms, Solar Flares, Cosmic Rays. Edited by S. T. Suess and B. T. Tsurutani. American Geophysical Union, 1998.

The 23rd Cycle: Learning to Live with a Stormy Star. Sten Odenwald. Columbia University Press, 2001.

More pictures and data from the IMAGE mission are available at **http://image.gsfc.nasa.gov/**

General information on space weather can be found at **www.spaceweather.com/**

NASA/IMAGE EUV TEAM

IMPACT CRASH

CRASH SHOCK

THIS IS NOT A SIGHT you would ever want to see. If a white dwarf star hit the sun, it would trigger a calamitous series of events—despite the fact that the dwarf is barely a hundredth the sun's diameter. As the dwarf approached, it would suck matter toward it and distort the sun into a pear shape. Thankfully, such a collision is unlikely. But similar events occur regularly in denser parts of the galaxy, such as globular star clusters.

SMASH INTO

WHEN STARS

RAM

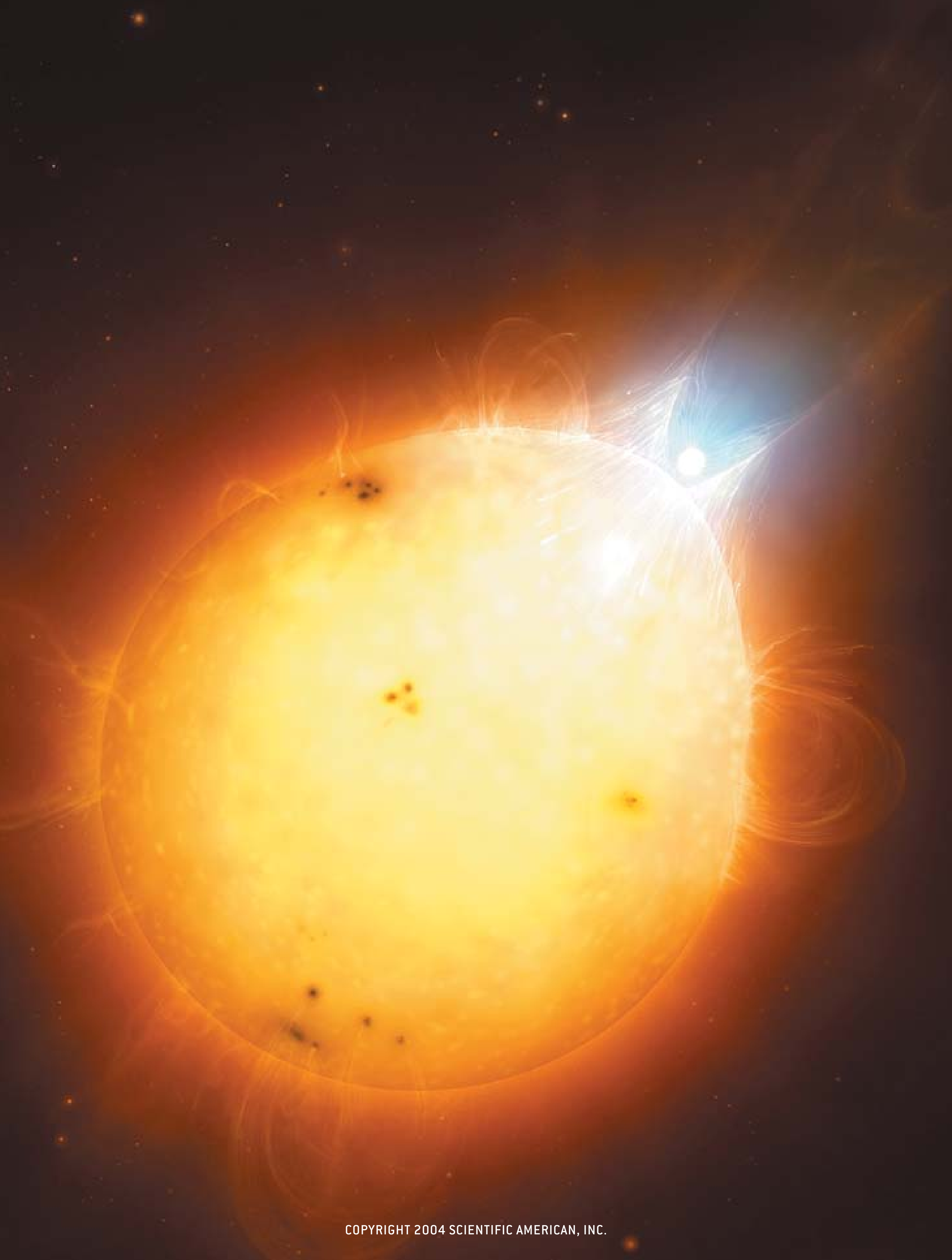
COLLIDE

VIOLENT BLOW

SMACK JOLT

WHEN TWO STARS SMASH INTO EACH OTHER, IT CAN BE A VERY PRETTY SIGHT (AS LONG AS YOU'RE NOT TOO CLOSE BY). THESE OCCURRENCES WERE ONCE CONSIDERED IMPOSSIBLE, BUT THEY HAVE TURNED OUT TO BE COMMON IN CERTAIN GALACTIC NEIGHBORHOODS

BY MICHAEL SHARA



Of all the ways for life on Earth to end, the collision of the sun and another star might well be the most dramatic. If the incoming projectile were a white dwarf—a superdense star that packs the mass of the sun into a body a hundredth the size—the residents of Earth would be treated to quite a fireworks show. The white dwarf would penetrate the sun at hypersonic speed, over 600 kilometers a second, setting up a massive shock wave that would compress and heat the entire sun above thermonuclear ignition temperatures.

It would take only an hour for the white dwarf to smash through, but the damage would be irreversible. The superheated sun would release as much fusion energy in that hour as it normally does in 100 million years. The buildup of pressure would force gas outward at speeds far above escape velocity. Within a few hours the sun would have blown itself apart. Meanwhile the agent of this catastrophe, the white dwarf, would continue blithely on its way—not that we would be around to care about the injustice of it all.

For much of the 20th century, the notion that stellar collisions might be worth studying seemed ludicrous to astronomers. The distances between stars in the neighborhood of the sun are just too vast for them to bump into one another. Other calamities will befall the sun (and Earth) in the distant future, but a collision with a nearby star is not likely to be one of them. In fact, simple calculations carried out early in the 20th century by British astrophysicist James Jeans suggested that not a single one of the 100 billion stars in the disk of our galaxy has ever run into another star.

But that does not mean collisions are uncommon. Jeans's assumptions and conclusion apply to the environs of the sun but not to other, more exotic parts of the Milky Way. Dense star clusters are a veritable demolition derby. Within these tight knots of stars, observers in recent years have discovered bodies that are forbidden by the principles of ordinary stellar evolution—but that are naturally explained as smashed-up stars. Collisions can modify the long-term evolution of entire clusters, and the most violent ones can be seen halfway across the universe.

The tranquil night sky masks a universe in which a THOUSAND PAIRS OF STARS COLLIDE EVERY HOUR.

Overview/*Stellar Collisions*

- This article represents one of those cases in which the textbooks need to be revised. The conventional wisdom that stars can never hit each other is wrong. Collisions can occur in star clusters, especially globular clusters, where the density of stars is high and where gravitational interactions heighten the odds of impact.
- The leading observational evidence for collisions is twofold. Globular clusters contain stars called blue stragglers that are best explained as the outcome of collisions. And globulars contain an anomalously high number of x-ray sources—again the likely product of collisions.

A Star-Eat-Star World

THE 1963 DISCOVERY of quasars was what inspired skeptical astronomers to take stellar collisions seriously. Many quasars radiate as much power as 100 trillion suns. Because some brighten or dim significantly in less than a day, their energy-producing regions must be no larger than the distance light can travel in a day—about the size of our solar system. If you could somehow pack millions of stars into such a small volume, astronomers asked, would stars crash? And could this jostling liberate those huge energies?

By 1970 it became clear that the answer to the second question was no. Nor could stellar slam dancing explain the narrow jets that emanate from the central powerhouses of many quasars. The blame fell instead on supermassive black holes. (Ironically, some astronomers have recently proposed that stellar collisions could help feed material into these holes.)

Just as extragalactic astronomers were giving up on stellar collisions, their galactic colleagues adopted them with a vengeance. The Uhuru satellite, launched in 1970 to survey the sky for x-ray-emitting objects, discovered about 100 bright sources in the Milky Way. Fully 10 percent were in the densest type of star cluster, globular clusters. Yet such clusters make up only 0.01 percent of the Milky Way's stars. For some reason, they contain a wildly disproportionate number of x-ray sources.

To express the mystery in a different way, consider what produces such x-ray sources. Each is thought to be a pair of stars, one of which has died and collapsed into a neutron star or a black hole. The ex-star cannibalizes its partner and in doing so heats the gas to such high temperatures that it releases x-rays. Such morbid couplings are rare. The simultaneous evolution of two newborn stars in a binary system succeeds in producing a luminous x-ray binary just once in a billion tries.

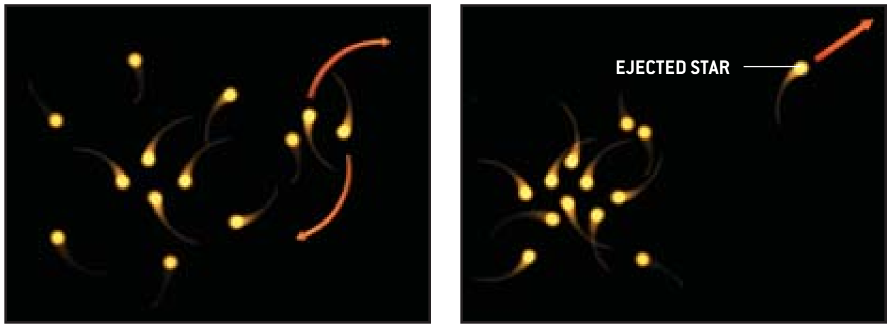
What is it about globular clusters of stars that overcomes these odds? It dawned on astronomers that the crowded conditions in globulars could be the deciding factor. A million stars are crammed into a volume a few dozen light-years across; an equivalent volume near the sun would accommodate only a hundred stars. Like bees in a swarm, these stars move in ever changing orbits. Lower-mass stars tend to be ejected from the cluster as they pick up energy during close encounters with more massive single and double stars, a process referred to as evaporation because it resembles the escape of molecules from the surface of a liquid. The remaining stars, having lost energy, concentrate closer to the cluster center. Giv-

MARK A. GARLICK (*preceding page*)

PROCESSES THAT MAKE COLLISIONS MORE LIKELY

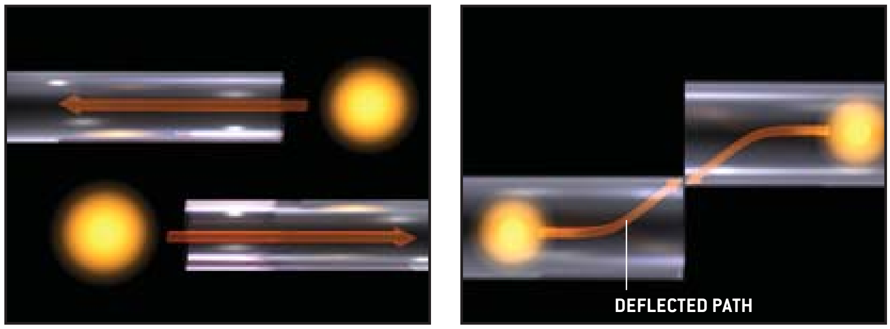
EVAPORATION

STARS IN A GLOBULAR CLUSTER zip around like bees in a swarm. Occasionally three or four come close to one another. Their close encounter redistributes energy and can fling one of the stars out of the cluster altogether. The remaining cluster members huddle together more tightly. If enough stars are ejected, the ones left behind begin to collide. This process typically occurs over billions of years.



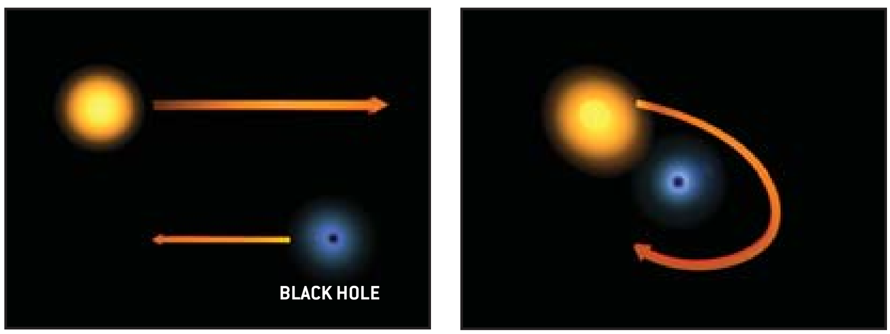
GRAVITATIONAL FOCUSING

IN THE COSMIC SCHEME of things, stars are small targets for impacts. Each sweeps out a very narrow region of space, and at first glance it appears that two such regions are unlikely to overlap. But gravity makes stars into larger targets by deflecting the paths of any approaching objects. In effect, each star actually sweeps out a region many times its own size, greatly increasing the probability of overlap and collision.



TIDAL CAPTURE

BLACK HOLE or neutron star makes an even smaller target than a normal star. But it can exert powerful tidal forces that bend a passing star out of shape. The distortion dissipates energy and can cause the two bodies to go into orbit. A collision between the two is then just a matter of time, as successive close passages rob ever more orbital energy.



ALFRED T. KAMAJIAN

en enough time, the tightly packed stars will begin to collide.

Even in a globular cluster, the average distance between stars is much larger than the stars themselves. But Jack G. Hills and Carol A. Day, both then at the University of Michigan at Ann Arbor, showed in 1975 that the probability of impact is not a simple matter of a star's physical cross section. Because the stars in a globular cluster move at a lackadaisical (by cosmic standards) 10 to 20 kilometers a second, gravity has plenty of time to act during close encounters. Without gravity, two stars can hit only if they are aimed directly at each other; with gravity, each star pulls on the other, deflecting its path. The stars are transformed from ballistic missiles with a preset flight path into guided missiles that home in on their target. A collision becomes up to 10,000 times more likely. In fact, half the stars in the central regions of some globular clusters have probably undergone one or more collisions over the past 13 billion years.

Around the same time, Andrew C. Fabian, James E. Pringle and Martin J. Rees of the University of Cambridge suggested that a grazing collision or a very near miss could cause two isolated stars to pair up. Normally a close encounter of two celestial bodies is symmetrical: they approach, gather speed, swing past each other and, unless they make contact, fly apart. But if one is a neutron star or a black hole, its intense gravity can contort the other, sapping some of its kinetic energy and preventing it from escaping, a process known as tidal capture. The neutron star or black hole proceeds to feast on its ensnared prey, spewing x-rays.

If the close encounter involves not two but three stars, it is even more likely to produce an x-ray binary. The dynamics of three bodies is notoriously complex and sometimes chaotic; the stars usually redistribute their energy in such a way that the two most massive ones pair up and the third gets flung away. The typical situation involves a loner neutron star that comes a little too close to an ordinary binary pair. One of the ordinary stars in the binary is cast off, and the neutron star takes its place, producing an x-ray source. The bottom line is that three-body dynamics and tidal capture lead to a thousandfold increase in the rate at which x-ray sources form in globular clusters, neatly solving the puzzle raised by Uhuru.

Crash Scene

WHAT HAPPENS WHEN TWO STARS smack into each other? As in a collision involving two vehicles, the outcome depends on several factors: the speed of the colliding objects, their internal structures and the impact parameter (which specifies whether the collision is head-on or a sideswipe). Some incidents are fender benders, some are total wrecks and some fall in between. Higher-velocity and head-on collisions are the best at converting kinetic energy into heat and pressure, making for a total wreck.

Although astronomers rely on supercomputers to study collisions in detail, a few simple principles govern the overall effect. Most important is the density contrast. A higher-density star will suffer much less damage than a tenuous one, just as a cannonball is barely marked as it blows a watermelon to shreds. A head-on collision between a sunlike star and a vastly denser star, such as a white dwarf, was first studied in the 1970s and 1980s by me and my colleagues Giora Shaviv and Oded Regev, both then at Tel Aviv University and now at the Technion-Israel Institute of Technology in Haifa. Whereas the sunlike star is annihilated, the white dwarf, being 10 million times as dense, gets away with only a mild warming of its outermost layers. Except for an anomalously high surface abundance of nitrogen, the white dwarf should appear unchanged.

The dwarf is less able to cover its tracks during a grazing collision, as first modeled by me, Regev and Mario Livio of the Space Telescope Science Institute. The disrupted sunlike star could form a massive disk in orbit around the dwarf. No such disks have yet been shown to exist, but it is possible that astronomers might be mistaking them for mass-transferring binary stars in star clusters.

When the colliding stars are of the same type, density and size, a very different sequence of events occurs. The case of two sunlike stars was first simulated in the early 1970s by Alastair G. W. Cameron, then at Yeshiva University, and Frederick G. P. Seidl, then at the NASA Goddard Institute for Space Studies. As the initially spherical stars increasingly overlap, they compress and distort each other into half-moon shapes. Temperatures and densities never climb high enough to ignite disruptive thermonuclear burning. As a small percent of the total mass squirts out perpendicular to the direction of stellar motion, the rest mixes together. Within an hour, the two stars have fused into one.

It is much more likely that two stars will collide somewhat off-axis than exactly head-on; it is also more likely that they will have slightly different rather than identical masses. This general case has been studied in detail by Willy Benz of the University of Bern in Switzerland, Frederic A. Rasio of Northwestern University, James C. Lombardi of Vassar College, Joshua E. Barnes of the University of Hawaii at Manoa and their collaborators. It is a beautiful mating dance that ends in the perpetual union of the two stars.

The object that results is fundamentally different from an isolated star such as our sun. An isolated star has no way of replenishing its initial allotment of fuel; its life span is preordained. The more massive the star is, the hotter it is and the faster it burns itself out. Given a star's color, which indicates its temperature, computer models of energy production can predict its life span with high precision. But a coalesced star does not follow the same rules. Mixing of the layers of gas during the collision can add fresh hydrogen fuel to the core, with a rejuvenating effect rather like tossing twigs on a dying campfire. Moreover, the object, being more massive than its progenitors, will be hotter, bluer and brighter. Observers who look at the star and use its color and luminosity to deduce its age will be wrong.

If a white dwarf hit the sun, Earth would fly off AND WANDER LIFELESSLY AROUND THE GALAXY.

For instance, the sun has a total life span of 10 billion years, whereas a star twice its mass is 10 times brighter and lasts only 800 million years. Therefore, if two sunlike stars merge halfway through their lives, they will form a single hot star that is five billion years old at the moment of its creation but looks as though it must be younger than 800 million years. The lifetime remaining to this massive fused star depends on how much hydrogen fuel was thrown to its center by the collision. Usually this lifetime will be much shorter than that of each of its parents. Even in death the star distinguishes itself. When it dies (by swelling to become a red giant, a planetary nebula and finally a white dwarf), it will be much hotter than other, older white dwarfs of similar mass.

Got the Blues

IN A GLOBULAR CLUSTER, massive merged stars will stand out conspicuously. All the members of a globular are born at roughly the same time; their temperature and brightness evolve in lockstep. But a coalesced star is out of sync. It looks preternaturally young, surviving when others of equal brightness and color have passed on. The presence of such stars in the cores of

THE AUTHOR

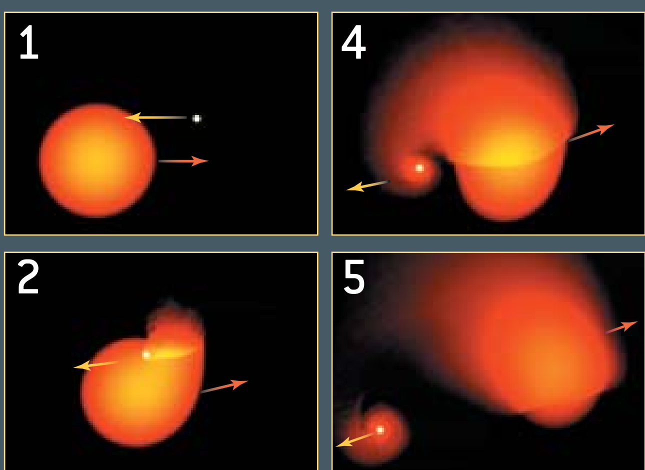
MICHAEL SHARA wanted to be an astronomer from age seven. His earliest interest came from observing binary stars with surplus World War II binoculars. Today he is curator and chair of the department of astrophysics at the American Museum of Natural History in New York City. Before joining the museum, he put in 17 years at the Space Telescope Science Institute, where he oversaw the peer-review committees for the Hubble Space Telescope. Shara's research interests include stellar collisions, novae and supernovae, and the populations of stars that inhabit star clusters and galaxies. Nowadays he observes with Hubble and ground-based instruments.

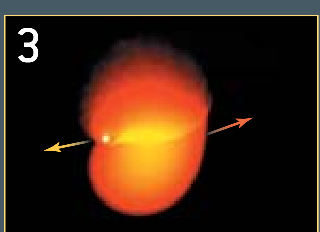
Having an Impact

	SUPERGIANT	RED GIANT	MAIN SEQUENCE	BROWN DWARF	WHITE DWARF	NEUTRON STAR	BLACK HOLE
BLACK HOLE	black hole + disk + white dwarf	black hole + disk + white dwarf	black hole + disk	black hole + disk	black hole + disk	black hole + disk	black hole
NEUTRON STAR	neutron star or black hole + disk + white dwarf	neutron star or black hole + disk + white dwarf	neutron star or black hole + disk	neutron star or black hole + disk	neutron star or black hole + disk	neutron star or black hole + disk	
WHITE DWARF	white dwarf + white dwarf	white dwarf + white dwarf	white dwarf	white dwarf or neutron star	neutron star or white dwarf		
BROWN DWARF	brown dwarf + white dwarf	brown dwarf + white dwarf	main sequence	main sequence or brown dwarf			
MAIN SEQUENCE	main sequence + white dwarf	main sequence + white dwarf	main sequence				
RED GIANT	white dwarf + white dwarf	white dwarf + white dwarf					
SUPERGIANT	white dwarf + white dwarf						

STARS COME IN seven basic types, with black holes having the greatest density and supergiants the least. Our sun is a main-sequence star. This table lists the outcomes of the 28 different pairings. In many cases, a collision can have more than one possible outcome, depending on impact speed, angle and other parameters. The results here assume deeply penetrating collisions at modest speeds. Two such collisions (*yellow*) are shown below.

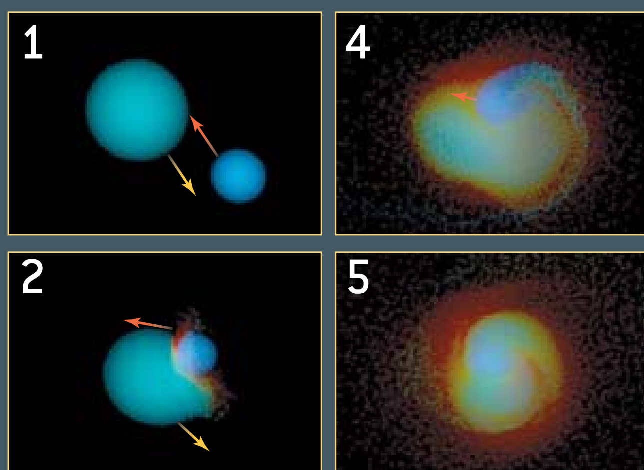
WHITE DWARF HITS RED GIANT

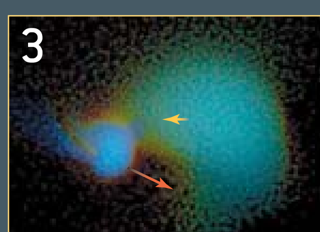




WHITE DWARF STAR takes a month to penetrate the bloated red giant. It escapes unscathed and spirits away some of the giant's gas. The giant, however, falls apart, although its core remains intact and becomes another white dwarf. (The full movie is available at www.ukaff.ac.uk/movies/collisions.mov)

MAIN SEQUENCE HITS MAIN SEQUENCE





ORDINARY STARS of unequal mass strike off-center. The less massive one is also denser, so it stays intact for longer. For an hour, it burrows into the larger star. A single, rapidly spinning star results. Some mass is lost to deep space. (The movie is at www.ifa.hawaii.edu/faculty/barnes/research_stellar_collisions)

MATTHEW BATE *University of Exeter* (*bottom left*); JOSHUA BARNES *University of Hawaii* (*bottom right*)

IN THE AFTERMATH of the collision between the sun and a white dwarf, the sun explodes as a giant thermonuclear bomb, leaving a gaseous nebula. A small percent of the sun's mass collects in a disk around the white dwarf, which continues on its way. Earth survives, but the oceans and atmosphere boil away. No longer held by the gravity of a central star, the planets all fly off into interstellar space and wander lifelessly around the galaxy.

dense star clusters is one of the most compelling predictions of stellar-collision theory.

As it happens, Allan R. Sandage of the Carnegie Institution of Washington discovered in the early 1950s that globular clusters contain anomalously hot and bright stars called blue stragglers. Researchers have since advanced a dozen theories of their origin. But it is only in the past decade that the Hubble Space Telescope has provided a strong link with collisions.

implicated in the enormous energy releases associated with gamma-ray bursts.

Collisions are already proving crucial to understanding globulars and other celestial bodies. Computer simulations suggest that the evolution of clusters is controlled largely by tightly bound binary systems, which exchange energy and angular momentum with the cluster as a whole. Clusters can dissolve altogether as near-collisions fling stars out one by one. Piet Hut

Astronomers have recently detected DISKS OF GAS ORBITING BLUE STRAGGLERS— remnants of the stars' violent births.

In 1991 Francesco Paresce, George Meylan and I, all then at the Space Telescope Science Institute, found that the very center of the globular cluster 47 Tucanae is crammed with blue stragglers, exactly where collision theory predicted they should exist in greatest number. Six years later David Zurek of the Space Telescope Science Institute, Rex A. Saffer of Villanova University and I carried out the first direct measurement of the mass of a blue straggler in a globular cluster. It has approximately twice the mass of the most massive ordinary stars in the same cluster—as expected if stellar coalescence is responsible. Saffer and his colleagues have found another blue straggler to be three times as massive as any ordinary star in its cluster. Astronomers know of no way other than a collisional merger to manufacture such a heavy object in this environment.

We are now measuring the masses and spins of dozens of blue stragglers. Orsola De Marco of the American Museum of Natural History in New York City and her colleagues have recently detected disks of gas orbiting several blue stragglers—perhaps remnants of their violent births. Meanwhile observers are also looking for the other predicted effects of collisions. For instance, S. George Djorgovski of the California Institute of Technology and his co-workers have noted a decided lack of red giant stars near the cores of globular clusters. Red giants have cross sections thousands of times as large as the sun's, so they are unusually big targets. Their dearth is naturally explained by collisions, which would strip away their outer layers and transform the stars into a different breed.

To be sure, all this evidence is circumstantial. Definitive proof is harder to come by. The average time between collisions in the 150 globular clusters of the Milky Way is about 10,000 years; in the rest of our galaxy it is billions of years. Only if we are extraordinarily lucky will a direct collision occur close enough—say, within a few million light-years—to permit today's astronomers to witness it with present technology. The first real-time detection of a stellar collision may come from the gravitational-wave observatories that are capturing data. Close encounters between stellar-mass objects should lead to distortions in the spacetime continuum. The signal is especially strong for colliding black holes or neutron stars. Such events have been

of the Institute for Advanced Study in Princeton, N.J., and Alison Sills of McMaster University in Ontario have argued that stellar dynamics and stellar evolution regulate each other by means of subtle feedback loops.

The fates of planets whose parent stars undergo close encounters is another addition to the topic of stellar collisions. Numerical simulations by Jarrod R. Hurley of the American Museum of Natural History show that the planets often fare badly: cannibalized by their parent star or one of their planetary siblings, set adrift within the star cluster, or even ejected from the cluster and doomed to tramp through interstellar space. Recent Hubble observations by Ron Gilliland of the Space Telescope Science Institute suggest that stars in a nearby globular cluster do indeed lack Jupiter-size planets, although the cause of this deficiency is not yet known for sure.

Despite the outstanding questions, the progress in this field has been astonishing. The very idea of stellar collisions was once absurd; today it is central to many areas of astrophysics. The apparent tranquillity of the night sky masks a universe of almost unimaginable power and destruction, in which a thousand pairs of stars collide somewhere every hour. And the best is surely yet to come. New technologies may soon allow direct and routine detection of these events. We will watch as some stars die violently, while others are reborn, phoenixlike, during collisions. SA

MORE TO EXPLORE

The First Direct Measurement of the Mass of a Blue Straggler in the Core of a Globular Cluster: BSS 19 in 47 Tucanae. Michael M. Shara, Rex A. Saffer and Mario Livio in *Astrophysical Journal Letters,* Vol. 489, No. 1, Part 2, pages L59–L62; November 1, 1997.

Star Cluster Ecology III: Runaway Collisions in Young Compact Star Clusters. Simon Portegies Zwart, Junichiro Makino, Stephen L. W. McMillan and Piet Hut in *Astronomy and Astrophysics,* Vol. 348, No. 1, pages 117–126; 1999. **arxiv.org/abs/astro-ph/9812006**

Evolution of Stellar Collision Products in Globular Clusters–II: Off-Axis Collision. Alison Sills, Joshua A. Faber, James C. Lombardi, Jr., Frederic A. Rasio and Aaron Warren in *Astrophysical Journal,* Vol. 548, No. 1, Part 1, pages 323–334; February 10, 2001. **astro-ph/0008254**

The Promiscuous Nature of Stars in Clusters. Jarrod R. Hurley and Michael M. Shara in *Astrophysical Journal,* Vol. 570, No. 1, Part 1, pages 184–189; May 1, 2002. **astro-ph/0201217**

MARK A. GARLICK

X-RAY BINARIES

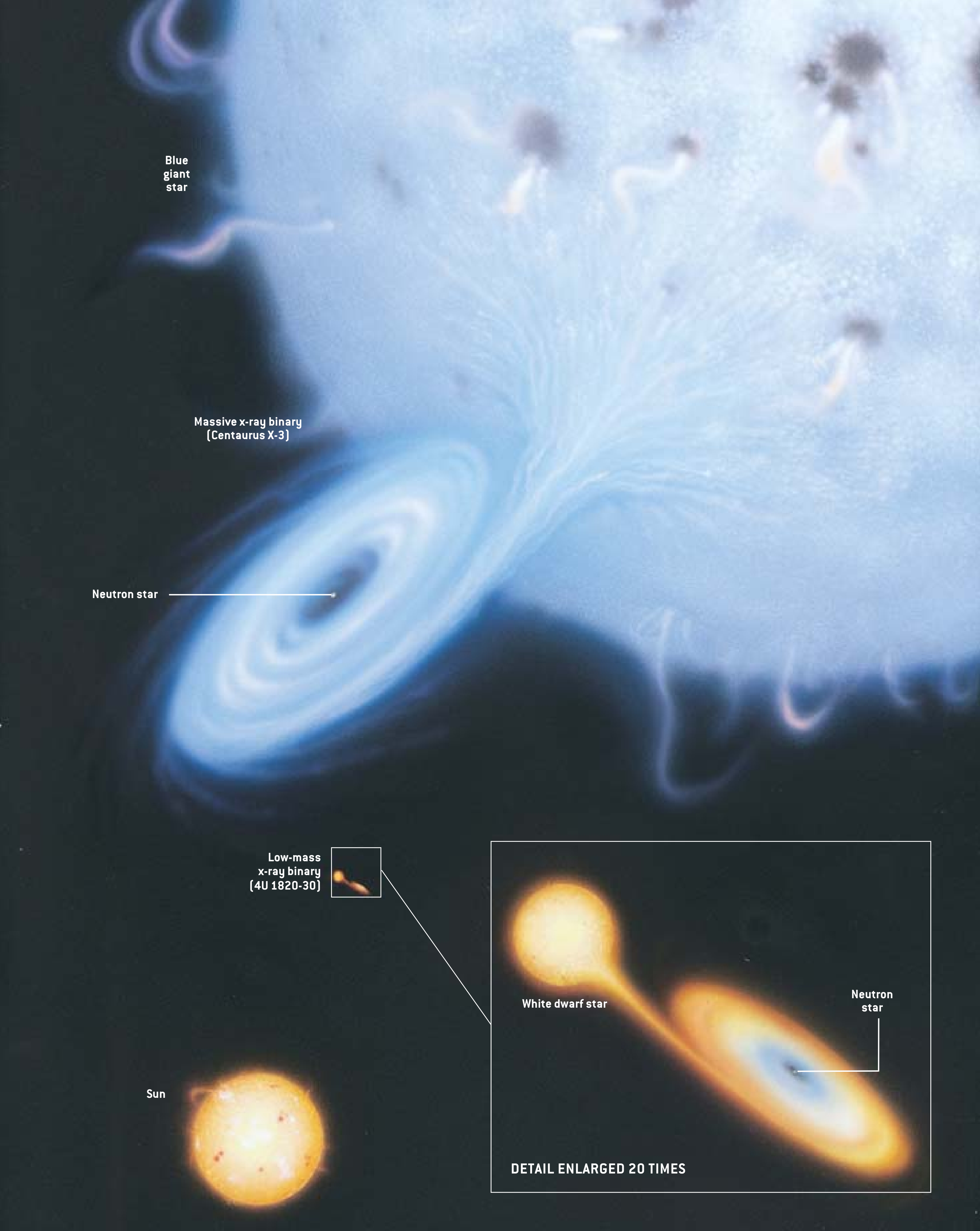
In these systems, ultradense neutron stars feed on their more sedate companions. Such stellar cannibalism produces brilliant outpourings of x-rays and drastically alters the evolution of both stars

By Edward P. J. van den Heuvel and Jan van Paradijs

All the shimmering stars that pierce the night sky shine because of the same fundamental process: nuclear fusion. When two or more atomic nuclei collide and fuse into one, they release virtually unimaginable amounts of energy. The fusion of one gram of hydrogen, for example, liberates as much energy as the combustion of 20,000 liters of gasoline. In stars such as the sun, fusion reactions burn brilliantly for billions of years.

X-RAY BINARIES make up two very different classes of double-star systems. In both cases, a neutron star lies at the heart of the x-ray source. Most young x-ray binaries, such as Centaurus X-3 (*top*), contain a bright blue star having 10 to 40 times the mass of the sun. Low-mass x-ray binaries usually contain far older, sunlike stars; in the tiny system 4U 1820-30 (*bottom*), both stars must be compact objects, presumably a neutron star and a larger but less massive white dwarf.

ALFRED T. KAMAJIAN


Blue giant star
Massive x-ray binary (Centaurus X-3)
Neutron star
Low-mass x-ray binary (4U 1820-30)
White dwarf star
Neutron star
DETAIL ENLARGED 20 TIMES
Sun

They are not the only source of stellar energy, however. In 1971 astronomers recognized a class of bizarre, x-ray-emitting stars, known as x-ray binaries, whose intense emissions require an energy source far more efficient than even fusion.

Theorists have deduced that these objects consist of a normal star orbiting a collapsed stellar corpse, usually a neutron star. Neutron stars are so dense that the entire mass of the star is squeezed into what is essentially a single atomic nucleus 20 kilometers across. The stars in these binaries lie so close together that gas can flow from the normal star to the neutron star. That captured material forms a rapidly swirling disk whose inner edge, just above the neutron star's surface, races around at nearly the speed of light. Friction within the disk eventually causes the gas to fall inward, or accrete, onto the neutron star. In the process, violent collisions between particles heat the gas to temperatures of 10 million to 100 million kelvins. Under such incredibly hot conditions, the gas emits torrents of energetic x-rays. Pound for pound, accretion unleashes 15 to 60 times as much energy as does hydrogen fusion.

Astronomers now recognize that accretion powers a rich diversity of astrophysical objects. These range from infant stars to quasars, objects about the size of the solar system that outshine entire galaxies, most likely as a result of gas spiraling into a supermassive black hole. X-ray binaries serve as ideal showcases for learning in detail how the accretion process works. They are bright and relatively nearby, residing well within our galaxy.

The study of x-ray binaries also provides a glimpse into the life cycle of some of the most exotic and dynamic stellar systems in the sky. In these stellar duos, one or both members spend some time feeding off their partner. That transfer of material stunningly alters both stars' development. One star may pay for its gluttony by prematurely ending its life in a spectacular supernova explosion. On the other hand, placid, elderly neutron stars may receive an infusion of rotational energy that causes them to become a prominent source of rapidly pulsed radio waves.

One star may pay for its gluttony IN A PREMATURE SUPERNOVA EXPLOSION, but another may receive an energy infusion.

Rocket to the Pulsars

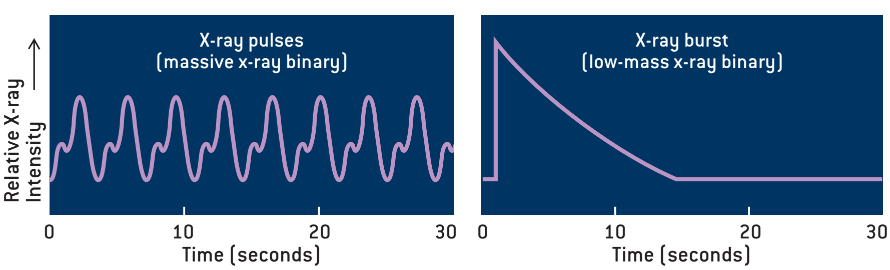
DESPITE THEIR PROMINENCE in the x-ray sky, x-ray binaries escaped the notice of researchers until the dawn of the space age in the 1960s. Celestial x-rays are absorbed high in the upper atmosphere, precluding their detection from the ground. The advent of space technology opened up an entirely new field of investigation by making it possible to loft telescopes above the obscuring layers of Earth's turbulent atmosphere.

In 1962 Riccardo Giacconi and his associates at American Science and Engineering in Cambridge, Mass., placed an x-ray detector on board a rocket and discovered the first known celestial x-ray source, Scorpius X-1. (In 2002 Giacconi was awarded the Nobel Prize in Physics for this feat.) The name Scorpius X-1 indicates that it is the brightest x-ray-emitting object in the constellation Scorpius. Scorpius X-1 shines about 1,000 times brighter in x-rays than in visible light. The identity of the object emitting this radiation was a total mystery.

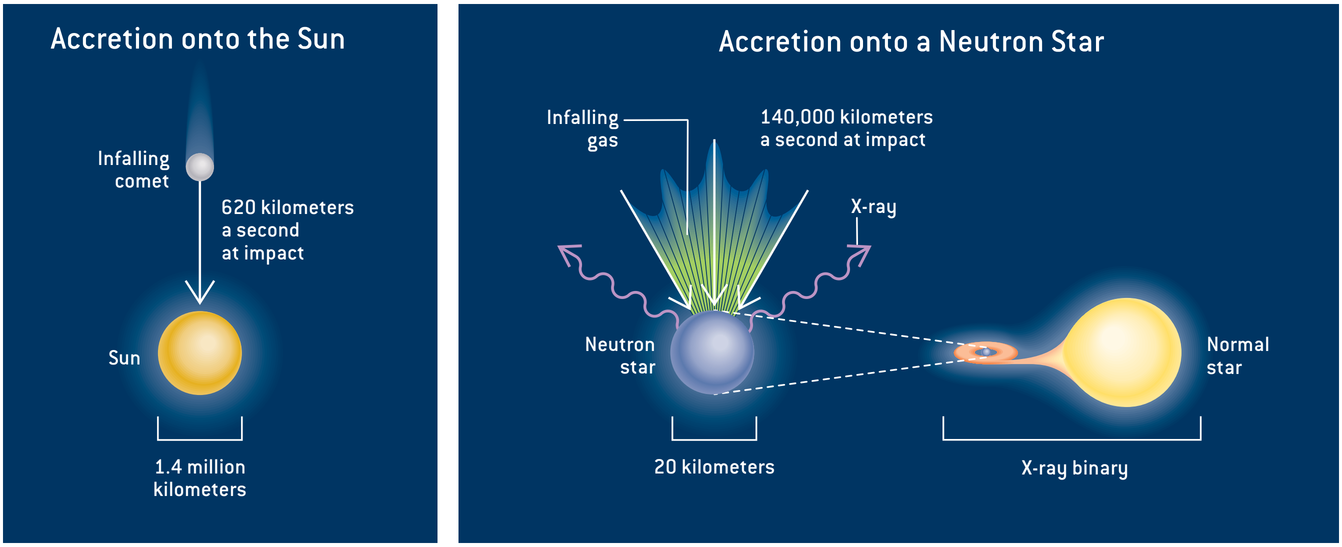
In the following years, x-ray detectors placed on rockets and very high altitude balloons revealed a few dozen similar "x-ray stars." Astronomers truly began to understand these objects only after 1970, when the National Aeronautics and Space Administration launched Uhuru, the first x-ray satellite, which was designed and built by a team led by Giacconi. Suddenly, astronomers could study the x-ray sky around the clock. Within its first few months of service, Uhuru revealed two intriguing x-ray sources, Centaurus X-3 and Hercules X-1. Both objects vary in brightness in a rapid, extremely regular manner: once every 4.84 seconds for Centaurus X-3, once every 1.20 seconds for Hercules X-1. These sources turned out to be the first of a whole class of pulsed x-ray stars.

The pulses provided a critical clue to the nature of these objects. In 1967 Antony Hewish and S. Jocelyn Bell of the University of Cambridge, along with several co-workers, discovered pulsars, a class of stars that emit regular blips of radio emission. After some initial puzzlement, theorists realized that radio pulsars are swiftly spinning neutron stars whose powerful magnetic fields

NEUTRON STAR in the young, massive binary Centaurus X-3 emits pulses of x-rays as it rotates (*left*). But in the tiny low-mass binary 4U 1820-30, bursts of x-rays occur erratically, when gas collects on the surface of the old neutron star and undergoes a thermonuclear detonation (*right*).



JARED SCHNEIDMAN



INFALL OF MATTER, or accretion, can be nature's most efficient mechanism for generating energy. The energy liberated depends on surface gravity. Matter falling onto the sun (*left*) attains only a tiny fraction of the velocity of material accreting onto an ultradense neutron star (*right*). Friction converts kinetic energy into thermal energy; infalling gas in an x-ray binary reaches 100 million kelvins, causing it to emit energetic x-rays.

generate a lighthouse beam of radio waves that flashes by the observer once each rotation. The similarly short and constant variations of the newfound x-ray stars hinted that they, too, were associated with neutron stars.

Another noteworthy trait of Centaurus X-3 and Hercules X-1 is that they experience regular eclipses, in which they dip to a small fraction of their normal brightness. These eclipses proved that the objects must be binary stars, presumably a neutron star orbiting a larger but much more sedate stellar companion that occasionally blocks the neutron star from view. Centaurus X-3 has an orbital period of 2.087 days; for Hercules X-1, the period is 1.70 days.

The pieces of the puzzle began to fall into place. The short orbital periods of the pulsating x-ray stars demonstrated that the two stars sit very close to each other. In such proximate quarters the neutron star can steal gas from its companion; the gas settles into a so-called accretion disk around the neutron star. The inner parts of the disk greatly surpass the white-hot temperatures on the surface of the sun (about 6,000 kelvins). As a result, the accretion disk shines mostly in the form of x-rays, radiation thousands of times as energetic as is visible light. Accretion is so efficient that some x-ray binaries emit more than 10,000 times as much energy in x-rays as the sun radiates at all wavelengths.

The x-ray pulsations occur because the neutron star has a strong magnetic field whose axis is inclined with respect to its axis of rotation. Close to the neutron star, the magnetic field directs the infalling, electrically charged gas toward the star's magnetic poles. There the gas crashes onto the surface, giving rise to two columns of hot (100 million kelvins), x-ray-emitting material. As the star rotates, these columns move in and out of view as seen from Earth, explaining the variation in the star's apparent x-ray flux. Several researchers independently arrived at this explanation of pulsating and eclipsing binary x-ray sources; indeed, by 1972, it had already become accepted as the standard model for such objects.

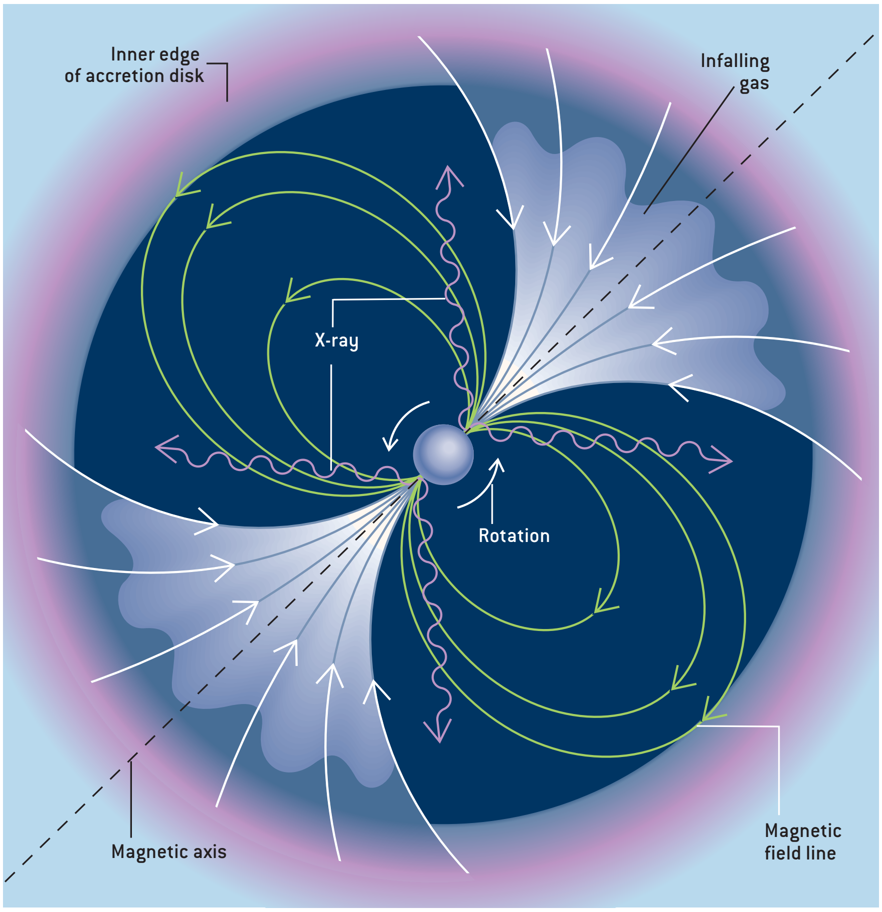
Careful timing of the pulsations of x-ray binaries showed that they are not perfectly regular. Instead the period of pulsation smoothly increases and decreases over an interval equal to the orbital period. This phenomenon results from the motion of the x-ray source around the center of gravity of the binary star system. While the source is moving toward Earth, each pulse travels a shorter distance than the one before and so arrives a minuscule fraction of a second early; while the source is moving away from Earth, each pulse arrives a similar amount late.

The amplitude of this effect reveals the velocity at which the source moves along the line of sight to Earth. Centaurus X-3 swings back and forth at 415 kilometers a second. That velocity implies that the companion star has at least 15 times the mass of the sun, typical of a brilliant, short-lived blue star. Since the early 1970s, astronomers have uncovered about 70 pulsating x-ray binaries. In nearly all cases, the companion stars are luminous blue stars having masses between 10 and 40 times that of the sun.

The bright stars in x-ray binaries show periodic changes in the frequency of dark lines, or absorption lines, in their spectra. These changes, known as Doppler shifts, result from the orbital motion of the visible star around the x-ray source. Radiation from an approaching object appears compressed, or bluer; likewise, radiation from a receding object looks stretched, or redder. The degree of the Doppler shift indicates the star's rate of motion. Because the corresponding velocity of the x-ray source can be deduced from the variations of the pulse period, one can use Newton's law of gravity to derive the mass of the embedded neutron star.

The measured neutron star masses

JARED SCHNEIDMAN



MAGNETIC FIELD of a young neutron star prevents infalling gas from reaching the surface, except at the two magnetic poles. Two hot, x-ray-emitting columns of gas, each about a kilometer across, collect at the poles. The star's rotation axis is inclined with respect to its magnetic axis, so an observer perceives regular pulses of x-rays as the magnetic poles rotate in and out of view.

fall primarily between 1.2 and 1.6 times the mass of the sun, in good agreement with theoretical expectations. Researchers have found, much to their excitement, that several nonpulsating x-ray binaries seem to contain stars having more than about three solar masses. Current theory holds that neutron stars exceeding that mass limit will produce a gravitational field so intense that it collapses without limit. The result is one of nature's most intriguing phenomena: a black hole, an object whose gravity has cut it off from the rest of the universe.

Burst or Pulse, Not Both

AS ASTRONOMERS have found more x-ray binaries, they have come to recognize the existence of two distinct populations: those containing large and luminous blue stars and those containing much older, less massive stars more akin to the sun. The x-ray binaries that include massive blue stars must be very youthful. A star more than 15 times as massive as the sun squanders its supply of hydrogen fuel in less than 10 million years, a blink of the eye compared with the 13-billion-year age of the Milky Way. Hence, the double-star systems from which these x-ray binaries evolved must have been born only a few million years ago in interstellar gas clouds. Like these clouds and other young, hot stars, pulsating massive x-ray binaries tend to concentrate in the plane of the Milky Way, but not toward the galactic center.

About half the strong x-ray sources in our galaxy, including Scorpius X-1, belong to a very different stellar population. These x-ray binaries concentrate predominantly in the central lens-shaped bulge of the galaxy and in globular clusters, dense spherical swarms of stars. Such regions harbor mostly older stars, those having ages between about five billion and 13 billion years.

In general, these elderly x-ray binaries do not undergo regular pulsations. The visible-light spectra of the aged x-ray binaries also appear utterly unlike those of normal stars. Instead they grow steadily brighter toward the blue end of the spectrum; some of their radiation emerges at distinct wavelengths, or colors. Theoretical models indicate that such a spectrum would be produced by an inflowing disk of gas heated by intense x-rays streaming from inner parts of the disk, just above the neutron star's surface.

Emission from the disk almost completely drowns out the light from the companion star. That disparity implies that the companion must be fairly faint, which in turn indicates that its mass is no greater than that of the sun. These double-star systems are therefore known as low-mass x-ray binaries. Solar-mass stars remain stable for at least 10 billion years, consistent with the age of the stellar population in which low-mass x-ray binaries reside.

Low-mass x-ray sources undergo occasional extreme flare-ups, or x-ray bursts, which have yielded a great deal of information about these systems. Within a few seconds of the beginning of a burst, the object's x-ray brightness increases by a factor of 10 or more, peaks for a few seconds to a few minutes and then decays to the original level in about a minute. X-ray bursts recur irregularly every few hours or so.

Researchers have deduced that the x-ray bursts result from runaway nuclear fusion reactions in the gas accreted onto the surface of a neutron star. Between bursts, new matter flowing from the companion star replenishes the nuclear fuel. That steady accretion gives rise to the persistent emission of x-rays seen between the bursts. Despite the spectacular nature of the bursts, low-mass x-ray binaries emit more than 90 percent of their total energy during times of quiescence—a testimony to the great efficiency of accretion compared with fusion.

JARED SCHNEIDMAN

X-ray bursts occur only in low-mass binary star systems and x-ray pulses almost solely in high-mass ones; not a single system displays both forms of behavior. The critical factor responsible for this disparity is probably the strength of the neutron star's magnetic field. High-mass x-ray binaries must contain neutron stars having powerful magnetic fields, capable of generating easily detectable pulsations. Neutron stars in low-mass x-ray binaries seem to possess far weaker fields. This explanation is bolstered by theoretical models indicating that a powerful magnetic field would inhibit the nuclear instabilities that produce x-ray bursts.

The disparate characteristics of low-mass and high-mass x-ray binaries underscore the very different ways in which these systems must have formed and evolved. Almost immediately after the discovery of high-mass x-ray binaries in 1971, workers recognized that such objects represent a normal stage in the evolution of close double-star systems in which both objects have more than a few times the mass of the sun. The more massive star quickly consumes its fuel and expands into a bloated red giant, whose outer layers spill over onto the companion star, exposing the red giant's helium-rich center. A few hundred thousand years later, this helium star explodes as a supernova, shedding much of its outer mass; its remnant core collapses into a neutron star. The neutron star attracts gas from its companion and becomes a source of x-rays.

The formation of a low-mass x-ray binary involves a more specialized set of circumstances. Some of these binaries could have started out as a massive star and a stellar lightweight orbiting each other. The small companion star would have too little gravity to capture material from the primary star. When the primary annihilates itself as a supernova, much of the system's mass would escape into interstellar space. In most cases, that loss of mass would disrupt the binary and send the two stars sailing off on separate courses. In the rare instance in which the stars remain bound to each other, they could evolve into a low-mass x-ray binary.

There is also a gentler way. If the primary star initially has less than eight times the mass of the sun, it will not blow up. Instead it will produce a white dwarf, a stellar cinder far denser than a normal star but much less so than a neutron star. In a white dwarf the star's gravity has crushed its constituent atoms into a soup of electrons and nuclei; a white dwarf having the mass of the sun would be about the size of Earth.

As the low-mass star evolves, it will gradually expand; if the two stars are in a close orbit, gas from the low-mass star will accrete onto the surface of the white dwarf. The mass of the white dwarf may eventually exceed a critical value, about 1.4 solar masses, and collapse into a neutron star. This kind of quiet collapse ejects very little material, so the system can remain tightly bound. Later, the stars spiral in closer toward each other, accretion begins and the system becomes a low-mass x-ray binary.

In such binaries, the neutron star's gravity exerts a strong pull on its much larger but less massive companion. The combination of gravitational and centrifugal forces gives rise to a pear-shaped region of stability, called a Roche lobe, surrounding the low-mass star. Any material lying outside the Roche lobe will flow toward the neutron star. The transfer of material causes the distance between the two stars to increase if the mass-losing star is the less massive of the two, as is the case in low-mass x-ray binaries. When the size of the orbit increases, so does the size of the Roche lobe. Once the lobe grows bigger than the companion star, the flow of matter ceases and the neutron star stops emitting x-rays. Evidently, some mechanism keeps feeding gas to the neutron star.

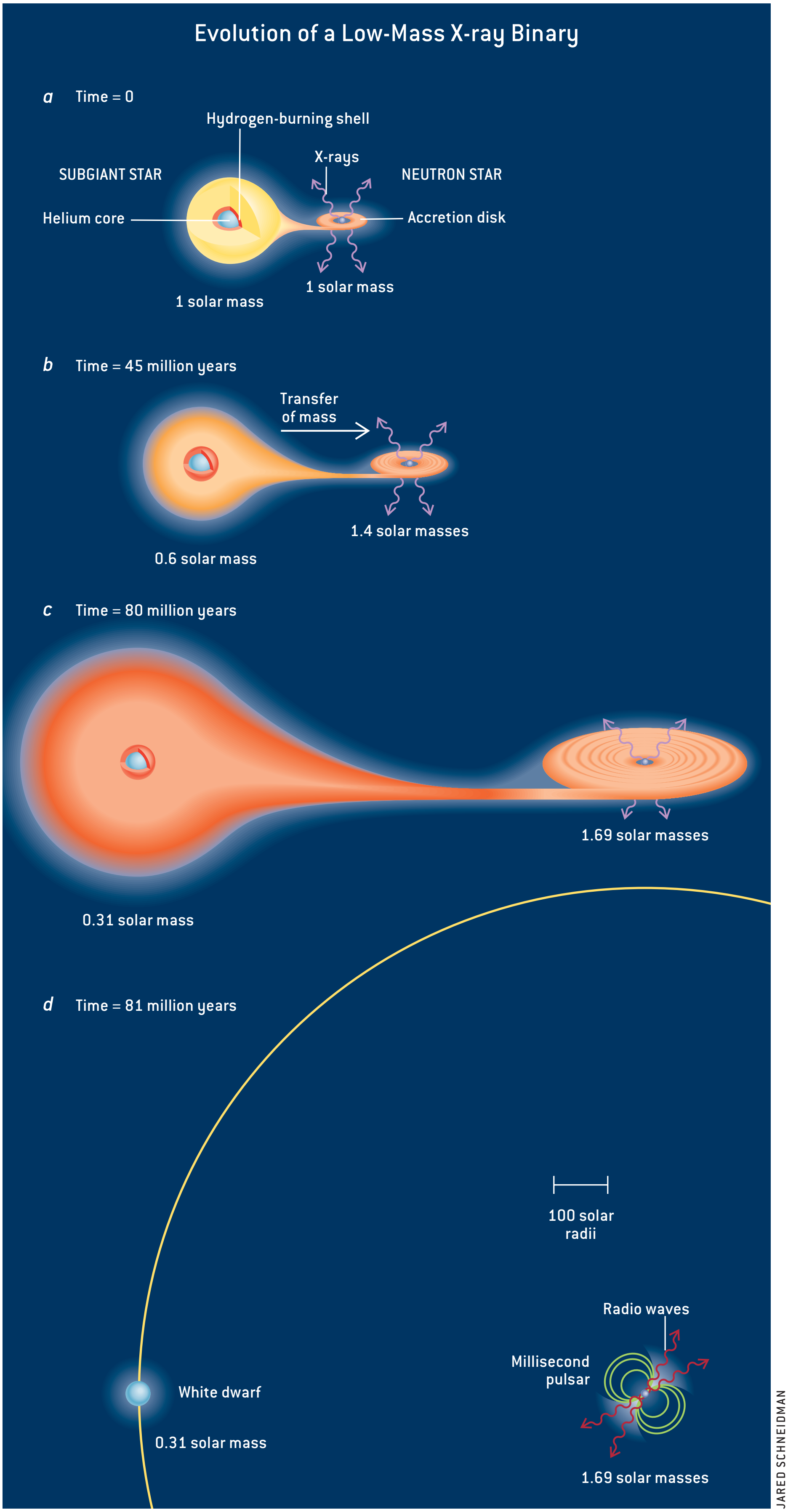
In one class of low-mass x-ray binaries—tightly bound systems whose periods are less than about 10 hours—the flow of gas is maintained by a steady shrinking of the stars' mutual orbit. As the stars orbit, they shed gravitational waves that carry off angular momentum, which causes the stars to draw closer together. That effect negates the tendency of mass transfer to move the stars apart. The stars ultimately settle into a slowly shrinking orbit in which a steady trickle of gas migrates from the companion to the neutron star. In this way, the neutron star accretes about one thousandth of an Earth mass a year, sufficient to account for the observed luminosity of many low-mass x-ray binaries (about 3×10^{30} watts).

Pound for pound, accretion unleashes 15 to 60 times AS MUCH ENERGY AS HYDROGEN FUSION.

The brightest x-ray sources in the central regions of the galaxy emit about 10 times that much energy. These objects constitute a second class of low-mass x-ray binaries that have relatively long orbital periods of approximately one to 10 days. Such leisurely orbits imply that the separation between the two stars, as well as the diameter of the normal companion, must be quite large. Here the flow of matter must result from the swelling of the companion star as a consequence of physical changes in its interior.

Such changes occur in the later evolutionary stages of a sunlike star. Hydrogen fusion produces helium, which accumulates as a dense core; hydrogen fusion takes place in a shell around this core. As the star ages, the hydrogen-burning shell migrates outward, causing the star's outer envelope to expand and cool. That expansion more than compensates for the increasing distance between the stars caused by the transfer of angular momentum. X-ray binaries having a period of about five to 10 days reach equilibrium if they experience a

LOW-MASS X-RAY BINARY initially consists of a neutron star pulling material from its companion (*a*). The low-mass star is an elderly subgiant having a dense, inert helium core. The transfer of mass causes the stars' orbit to widen. At the same time, the low-mass star steadily expands and cools as it evolves (*b*). The neutron star gradually consumes the subgiant's outer envelope (*c*). The exposed helium core (now considered a white dwarf) remains in a circular orbit around the neutron star (*d*). The rotating neutron star is now a millisecond pulsar that emits pulses of radio waves but no x-rays. (This scenario is based on calculations originally made by Paul C. Joss and Saul A. Rappaport of M.I.T.)

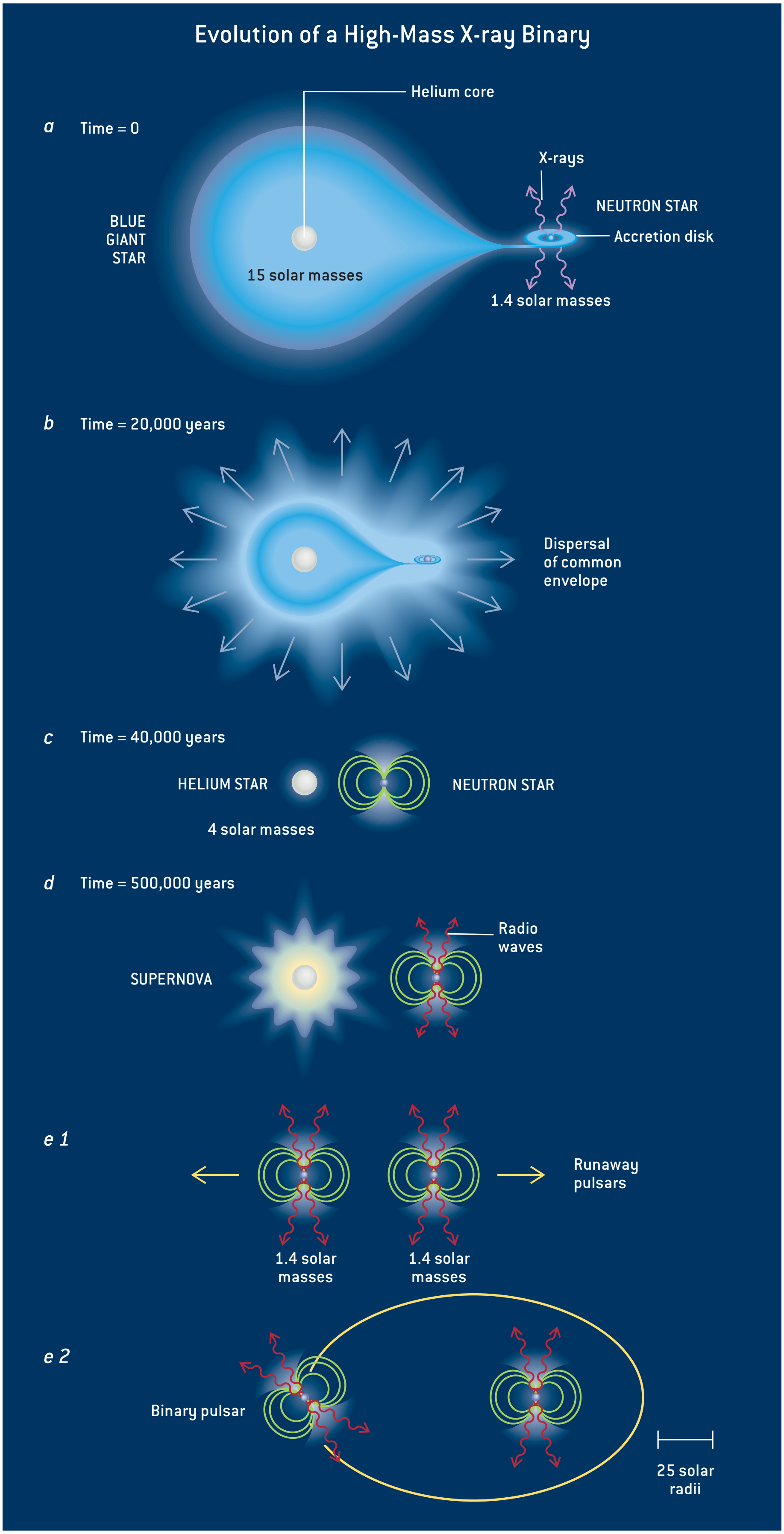


mass transfer rate of about five thousandths of an Earth mass a year, about the rate required to power the bright sources around the galactic center.

In 1982 a group of researchers—Ronald F. Webbink of the University of Illinois, Saul A. Rappaport of the Massachusetts Institute of Technology, G. J. Savonije of the University of Amsterdam and Ronald E. Taam of Northwestern University—investigated the fate of these low-mass x-ray binaries. Their calculations predict that, regardless of their initial traits, these systems always reach the same evolutionary end point. The giant star soon loses its entire hydrogen-rich envelope; its naked helium core remains as a white dwarf containing between 0.25 and 0.45 solar mass. The stars' final orbit is extremely circular because of the tens of millions of years of tidal interaction between the neutron star and its low-mass partner.

After the supply of accreting material dries up, binary star systems no longer emit detectable amounts of x-rays. The last evolutionary stages nonetheless offer a fascinating glimpse at what happens to very old neutron stars. During these later phases, the neutron star's most distinctive emission is in the form of radio waves, not x-rays.

In 1983, while working on the 300-meter radio telescope in Arecibo, Puerto Rico, Valentin Boriakoff of Cornell University, Rosolino Buccheri of the Italian National Research Council in Palermo and Franco Fauci of the University of Palermo discovered the binary radio pulsar PSR 1953+29. Its properties closely



JARED SCHNEIDMAN

MASSIVE X-RAY BINARY contains a bright blue star and an accreting neutron star (*a*). The blue star expands until its outer envelope engulfs both its helium-rich core and the neutron star (*b*). The orbital motion of the two stars inside the envelope heats the envelope and blows it away, leaving behind a helium star and a neutron star (*c*). If the helium star has more than 2.5 solar masses, it explodes as a supernova (*d*) and forms a second pulsing neutron star. The explosion may disrupt the system (*e1*); otherwise, the result is two neutron stars locked in a rapid, eccentric orbit (*e2*). Less massive helium cores do not explode; they end up as white dwarfs in circular orbits about the neutron star.

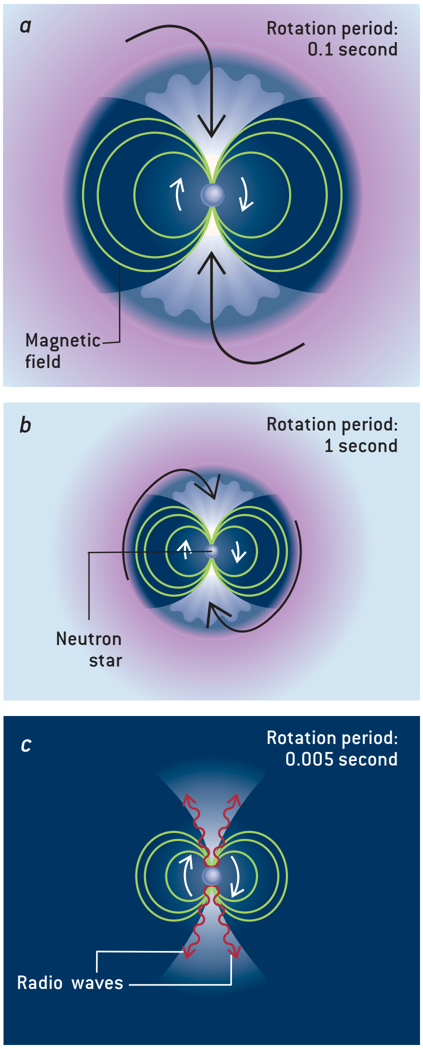
resemble those of the extinct x-ray binaries modeled by Webbink and his colleagues. The pulsar's radio signals displayed no signs of the eclipses or absorption produced by normal stars. The researchers recognized that the pulsar's companion must itself be a compact object. Because of its low mass, it is probably a white dwarf.

Superfast Spinners

ONE OF THE MOST surprising aspects of PSR 1953+29 is its period of radio pulsation: a remarkably swift 6.1 milliseconds, or 160 rotations a second. Half a year before the discovery of PSR 1953+29, Donald C. Backer of the University of California at Berkeley and his co-workers had found another pulsar, PSR 1937+21, which has a period of only 1.6 milliseconds. Astronomers now recognize these objects as the prototypes of a category of rapidly spinning neutron stars known as millisecond pulsars; 70 have been found since.

The inferred history of x-ray binaries made it clear why these pulsars spin so quickly. In low-mass x-ray binaries (and in many massive x-ray binaries as well), orbital motion prevents matter from falling directly onto the neutron star. Instead it goes into orbit about the star, forming an accretion disk. Material from the disk's inner edge falls onto the neutron star. During the later stages of accretion, that infalling material greatly speeds up the star's rotation.

Nearly all binary radio pulsars possess companions that have evolved into white dwarfs or neutron stars. At some



ROTATION OF NEUTRON STAR is strongly influenced by accretion in an x-ray binary. The neutron star's magnetic field defines the inner edge of the surrounding disk of matter, where gas falls onto the star. When the star is young, its field is strong, so the inner edge of the disk is distant and comparatively slow-moving (*a*). As the magnetic field decays, the inner edge of the accretion disk moves inward (*b*). The star now accretes rapidly moving material that causes its rate of rotation to increase. By the time the accretion ceases, the neutron star may be rotating hundreds of times per second (*c*).

stage, those companions were giants that overflowed their Roche lobes and dumped matter onto the neutron stars, increasing the stars' rate of rotation. During that time, the double stars would have appeared as x-ray binaries. After the companion star lost its outer layers and the accretion process ceased, a naked millisecond pulsar remained.

The power of a pulsar's radio emission varies in proportion to the fourth power of the rate of rotation. Millisecond neutron stars can be detected only because they were "spun up" by their companions during the x-ray binary phase. Radio pulsars that acquired their rapid rotation in this way are now called recycled pulsars, a term suggested by V. Radhakrishnan of the Raman Research Institute in Bangalore, India.

Starting in 1987, a number of observers found that globular clusters are incredibly rich hunting grounds for binary and millisecond pulsars. Studies of globular clusters have already revealed more than 60 radio pulsars; 70 percent of these pulsars rotate in less than 10 milliseconds, indicating that they are recycled. This celestial bounty results from the extremely dense nature of globular clusters. In their central regions, these clusters may contain more than 10,000 stars per cubic light-year, a million times the density of stars in the sun's corner of the galaxy. Under such crowded conditions, neutron stars face good odds of passing close to and capturing a stellar companion. Globular clusters harbor 200 to 1,000 times as many x-ray binaries per million stars as does the galaxy as a whole.

The "recycling" model for the origin of millisecond radio pulsars, first proposed in 1982, was elegantly confirmed in 1998 with the discovery of the first millisecond x-ray pulsar in the low-mass x-ray binary SAXJ 1808.4-3658. Rudy Wijnands and Michiel van der Klis of the University of Amsterdam, working with NASA's Rossi X-ray Timing Explorer Satellite, found that this x-ray source is an accreting neutron star rotating 400 times a second. Since then, four more millisecond x-ray pulses in low-mass x-ray binaries have been discovered.

In addition to the binary radio pulsars discussed, astronomers have identified another, rarer class of such objects that have substantially different characteristics. Their orbits often are extremely eccentric, and their companions contain 0.8 to 1.4 times the mass of the sun.

These objects probably arose from high-mass x-ray binaries in the following way. In massive x-ray binaries, accretion causes the two stars to spiral ever closer together (the opposite of the situation for low-mass x-ray binaries). That process, combined with the swelling of the companion star as it evolves, causes the companion to overflow its Roche lobe completely, engulfing the neutron star. Frictional drag quickly sends the neutron star spiraling in toward its companion. At a certain point, the friction generates so much heat that it drives off the gaseous hydrogen envelope. What remains is the neutron star in a close orbit around the stripped core of the companion, which consists of helium and heavier elements.

If the heavy-element core is sufficiently massive, it will later detonate into a supernova and produce a second neutron star. The force of the explosion and the precipitous loss of mass cause the stars' orbit to become elliptical; in many cases, the stars break free entirely to become runaway radio pulsars. If the orbit survives, the neutron stars follow their eccentric courses almost forever; over the ages, their orbits will slowly narrow because of the emission of gravitational waves. One of the most thoroughly studied binary pulsars, PSR 1913+16, consists of two neutron stars that race through a highly elliptical orbit once every seven hours and 45 minutes. This system's extreme properties allow it to serve as a sensitive test bed for many aspects of Einstein's

THE AUTHORS

EDWARD P. J. VAN DEN HEUVEL and *JAN VAN PARADIJS* collaborated on the study of celestial x-ray sources from the late 1970s until van Paradijs's death in 1999. Van den Heuvel received his Ph.D. in mathematical and physical science from the University of Utrecht in the Netherlands. In 1974 he joined the faculty of the University of Amsterdam, where he is now chairman of the astronomy department. He is also a co-founder and the director of the Center for High-Energy Astrophysics, operated jointly by the University of Amsterdam and the University of Utrecht. Van Paradijs earned his Ph.D. in astronomy from the University of Amsterdam and became a professor of astronomy at the university in 1988.

JARED SCHNEIDMAN

theory of relativity, as Joseph Taylor of Princeton University has beautifully demonstrated.

Particularly noteworthy is Taylor's discovery that the very precisely measured rate of shrinking of this system's orbit is exactly as expected from the emission of gravitational waves as predicted by Einstein's theory. For this finding, Taylor and his former student Russell Hulse were awarded the 1993 Nobel Prize in Physics.

Studies of binary pulsars have overturned a long-standing idea about how neutron stars change over the eons. Statistical analyses of pulsars had led most astronomers to conclude that a neutron star's magnetic field decays without any outside assistance and in due time vanishes completely. The existence of recycled pulsars proved, however, that some magnetic field persists even in extremely old systems. Moreover, the companion stars in binary pulsars offer a way to determine just how old those stars are.

Three of the millisecond pulsars have observable white dwarf companions, which serve as natural chronometers. A white dwarf steadily radiates away the heat left behind from its days as the core of a red giant star. Over the eons, white dwarfs grow progressively cooler and redder; the color of a white dwarf therefore betrays its age.

In 1986, Shrinivas R. Kulkarni of the California Institute of Technology measured the color of the white dwarf companion to PSR 0655+64 and concluded it must be at least 500 million years old. Using similar reasoning, three sets of researchers—led by J. F. Bell of the Mount Stromlo and Siding Spring Observatories; John Danziger of the European Southern Observatory; and Charles D. Bailyn of Yale University—determined that the white dwarf in the binary pulsar system PSR J0437-4715 is about two billion years old. The pulsars in these systems must be considerably older because they would have formed long before their companions evolved into white dwarfs, yet they retain substantial magnetic fields—or they could not be detected.

Magnetic No More

WORK BY Frank Verbunt, Ralph A.M.J. Wijers and Hugo Burm of the Center for High-Energy Astrophysics in the Netherlands further demonstrates the persistence of neutron-star magnetic fields. The researchers studied three anomalous, low-mass x-ray binaries that also are x-ray pulsars, indicating that they each contain a strongly magnetized neutron star. No matter how a neutron star originates, it always loses at least a few tenths of a solar mass in the form of neutrinos. When this happens, the binary system widens, shutting off the flow of gas. Accretion cannot occur until the binary system has shrunk through the emission of gravitational radiation or until the companion star has begun to evolve into a giant.

Both these mechanisms require considerable time to take effect. This knowledge allowed Verbunt to set lower limits to the ages of the accreting neutron stars in low-mass x-ray binary pulsars. In the case of Hercules X-1, the strongly magnetized neutron star is at least 500 million years old. Neutron stars' magnetic fields evidently do not spontaneously decay, at least not on such timescales.

And yet the magnetic fields of almost all binary radio pulsars are 100 to 10,000 times weaker than those of normal, youthful radio pulsars, regardless of whether the binary pulsar descended from a high-mass or a low-mass x-ray binary. The weakness of their fields seems to be attributable to some factor that all binary pulsars have in common. The most obvious common factor is accretion. In 1986 Taam and one of us (van den Heuvel) proposed, on observational grounds, a link between field decay and accretion. Theorists then advanced several models to explain the details.

LONG AFTER THEIR ERA OF FLAMBOYANT emission, x-ray binaries become the most STEADY ENTITIES IN THE COSMOS.

One model holds that newly accreted layers on the surface of a neutron star form an electrically conductive layer that allows only a small fraction of the star's magnetic field to reach the outside. Another possibility, proposed by G. Shrinivasan of the Raman Research Institute, is that it is the gradual slowing of a neutron star's rotation that causes its magnetic field to dissipate. Such deceleration occurs before and during the early stages of accretion. Once the magnetic field has weakened below a critical threshhold, the action of accretion reverses the spin-down trend, but that infusion of rotational energy cannot restore the field to its original strength.

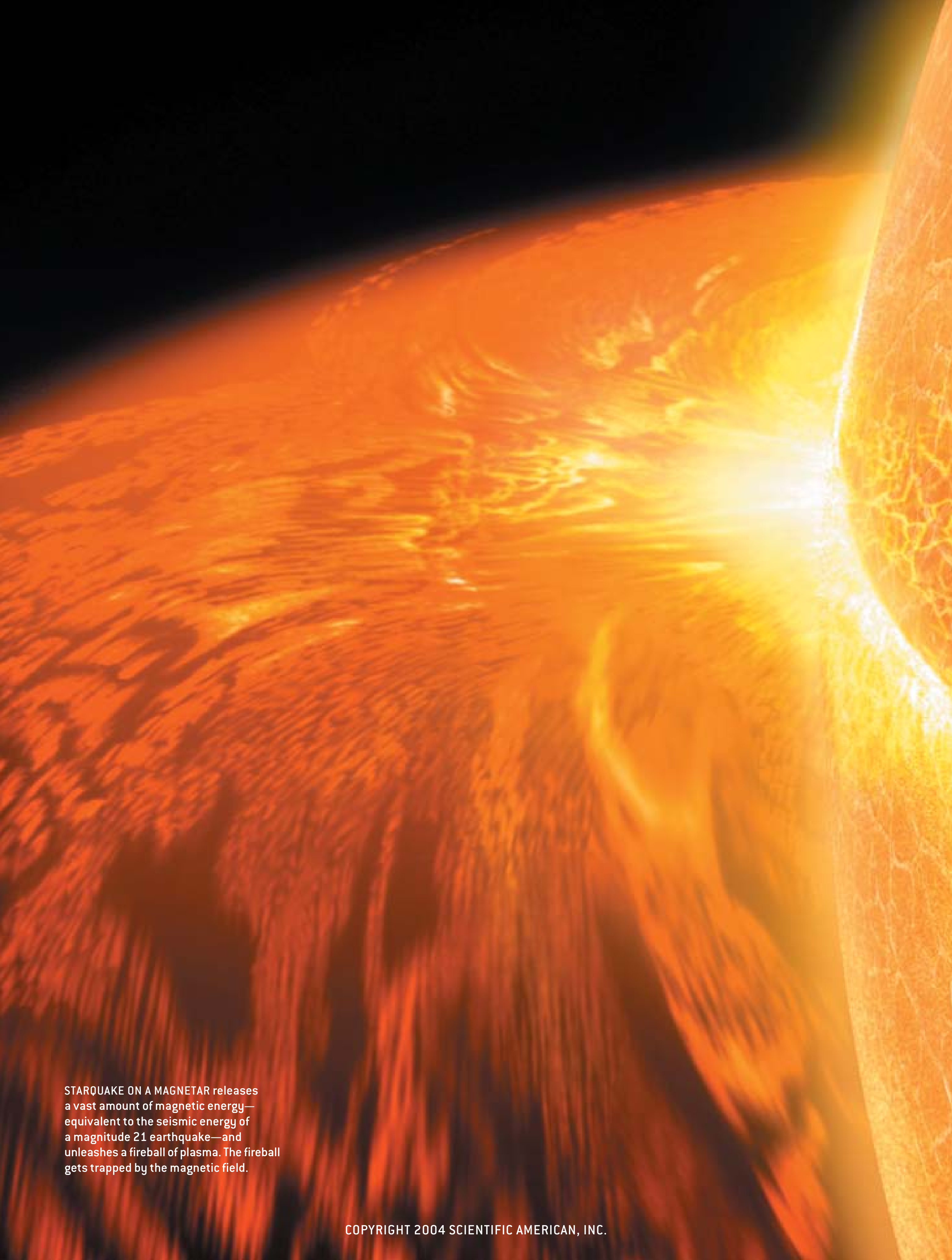
In any case, there is every indication that millisecond radio pulsars will retain their fields and continue to pulse untold billions of years into the future. Thus it happens that long after their era of flamboyant x-ray emission, x-ray binaries settle down to become some of the most steady, unchanging entities in the cosmos. SA

MORE TO EXPLORE

A New Class of Pulsars. Donald C. Backer and Shrinivas R. Kulkarni in *Physics Today,* Vol. 43, No. 3, pages 26–35; March 1990.

Pulsars Today. Francis Graham Smith in *Sky and Telescope,* Vol. 80, No. 3, pages 240–244; September 1990.

X-Ray Binaries. Edited by W.H.G. Lewin, J. A. van Paradijs and E.P.J. van den Heuvel. Cambridge University Press, 1995.



STARQUAKE ON A MAGNETAR releases a vast amount of magnetic energy—equivalent to the seismic energy of a magnitude 21 earthquake—and unleashes a fireball of plasma. The fireball gets trapped by the magnetic field.

MAGNETARS

Some stars are magnetized so intensely that they emit huge bursts of magnetic energy and alter the very nature of the quantum vacuum

By Chryssa Kouveliotou, Robert C. Duncan and Christopher Thompson

On March 5, 1979, several months after dropping probes into the toxic atmosphere of Venus, two Soviet spacecraft, Venera 11 and 12, were drifting through the inner solar system on an elliptical orbit. It had been an uneventful cruise. The radiation readings onboard both craft hovered around a nominal 100 counts per second. But at 10:51 A.M. EST, a pulse of gamma radiation hit them. Within a fraction of a millisecond, the radiation level shot above 200,000 counts per second and quickly went off scale.

Eleven seconds later gamma rays swamped the NASA space probe Helios 2, also orbiting the sun. A plane wave front of high-energy radiation was evidently sweeping through the solar system. It soon reached Venus and saturated the Pioneer Venus Orbiter's detector. Within seconds the gamma rays reached Earth. They flooded detectors on three U.S. Department of Defense Vela satellites, the Soviet Prognoz 7 satellite and the Einstein Observatory. Finally, on its way out of the solar system, the wave also blitzed the International Sun-Earth Explorer.

The pulse of highly energetic, or "hard," gamma rays was 100 times as intense as any previous burst of gamma rays detected from beyond the solar system, and it lasted just two tenths of a second. At the time, nobody noticed; life continued calmly beneath our planet's protective atmosphere. Fortunately, all 10 spacecraft survived the trauma without permanent damage.

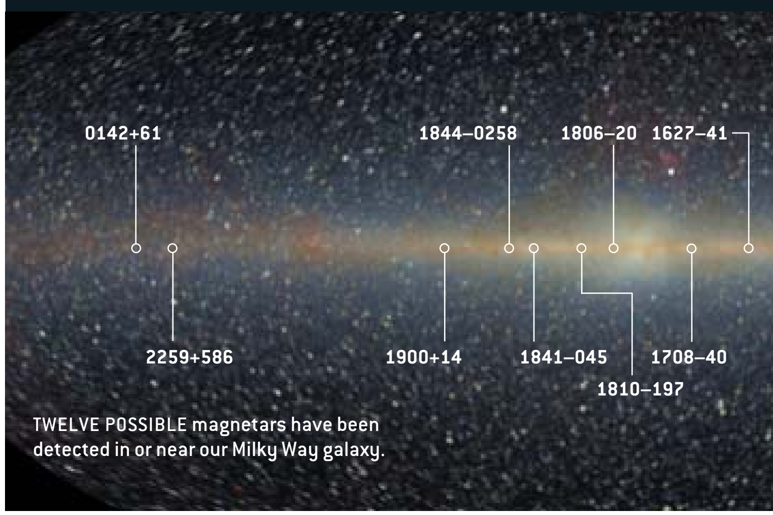
The hard pulse was followed by a fainter glow of lower-energy, or "soft," gamma rays, as well as x-rays, which steadily faded over the subsequent three minutes. As it faded away, the signal oscillated gently, with a period of eight seconds. Fourteen and a half hours later, at 1:17 A.M. on March 6, another, fainter burst of x-rays came from the same spot on the sky. Over the ensuing four years, Evgeny P. Mazets of the Ioffe Institute in St. Petersburg, Russia, and his collaborators detected 16 bursts coming from the same direction. They varied in intensity, but all were fainter and shorter than the March 5 pulse.

Astronomers had never seen anything like this. For want of a better idea, they initially listed these bursts in catalogues alongside the better-known gamma-ray bursts (GRBs), even though they clearly differed in several ways. In the mid-1980s Kevin C. Hurley of the University of California at Berkeley realized that similar outbursts were coming from two other areas of the sky. Evidently these sources were all repeating—unlike GRBs, which are one-shot events [see "The Brightest Explosions in the Universe," on page 92]. At a July 1986 meeting in Toulouse, France, astronomers agreed on the approximate locations of the three sources and dubbed them "soft gamma repeaters" (SGRs). The alphabet soup of astronomy had gained a new ingredient.

Another seven years passed before two of us (Duncan and Thompson) devised an explanation for these strange objects, and only in 1998 did one of us (Kouveliotou) and her team find compelling evidence for that explanation. Recent observations connect our theory to yet another class of celestial enigmas, known as anomalous x-ray pulsars (AXPs). These developments have led to a breakthrough in our understanding of one of the most exotic members of the celestial bestiary, the neutron star.

Neutron stars are the densest material objects known, packing slightly more than the sun's mass inside a ball just 20 kilometers across. Based on the study of SGRs, it seems that some neutron stars have magnetic fields so intense that they radically alter the material within them and the state of the quantum vacuum surrounding them, leading to physical effects observed nowhere else in the universe.

MAGNETAR CANDIDATES



TWELVE POSSIBLE magnetars have been detected in or near our Milky Way galaxy.

Not Supposed to Do That

BECAUSE THE MARCH 1979 BURST was so bright, theorists at the time reckoned that its source was in our galactic neighborhood, hundreds of light-years from Earth at most. If that had been true, the intensity of the x-rays and gamma rays would have been just below the theoretical maximum steady luminosity that can be emitted by a star. That maximum, first derived in 1926 by English astrophysicist Arthur Eddington, is set by the force of radiation flowing through the hot outer layers of a star. If the radiation is any more intense, it will overpower gravity, blow away ionized matter and destabilize the star. Emission below the Eddington limit would have been fairly straightforward to explain. For example, various theorists proposed that the outburst was triggered by the impact of a chunk of matter, such as an asteroid or a comet, onto a nearby neutron star.

But observations soon confounded that hypothesis. Each spacecraft had recorded the time of arrival of the hard initial pulse. These data allowed astronomers, led by Thomas Lytton Cline of the NASA Goddard Space Flight Center, to triangulate the burst source. The researchers found that the position coincided with the Large Magellanic Cloud, a small galaxy about 170,000 light-years away. More specifically, the event's position matched that of a young supernova remnant, the glow-

Overview/*Ultramagnetic Stars*

- Astronomers have seen a handful of stars that put out flares of gamma and x-radiation, which can be millions of times as bright as any other repeating outburst known. The enormous energies and pulsing signals implicate the second most extreme type of body in the universe (after the black hole): the neutron star.
- These neutron stars have the strongest magnetic fields ever measured—hence their name, magnetars. Magnetic instabilities analogous to earthquakes can account for the flares.
- Magnetars remain active for only about 10,000 years, implying that millions of them are drifting undetected through our galaxy.

DON DIXON (*preceding pages*); E. L. WRIGHT *University of California, Los Angeles,* THE COBE PROJECT, DIRBE AND NASA (*above*) ing remains of a star that exploded 5,000 years ago. Unless this overlap was pure coincidence, it put the source 1,000 times as far away as theorists had thought—and thus made it a million times brighter than the Eddington limit. In 0.2 second the March 1979 event released as much energy as the sun radiates in roughly 10,000 years, and it concentrated that energy in gamma rays rather than spreading it across the electromagnetic spectrum.

No ordinary star could account for such energy, so the source was almost certainly something out of the ordinary—either a black hole or a neutron star. The former was ruled out by the eight-second modulation: a black hole is a featureless object, lacking the structure needed to produce regular pulses. The association with the supernova remnant further strengthened the case for a neutron star. Neutron stars are widely believed to form when the core of a massive but otherwise ordinary star exhausts its nuclear fuel and abruptly collapses under its own weight, thereby triggering a supernova explosion.

Identifying the source as a neutron star did not solve the puzzle; on the contrary, it merely heightened the mystery. Astronomers knew several examples of neutron stars that lie within supernova remnants. These stars were radio pulsars, objects that are observed to blink on and off in radio waves. Yet the March 1979 burster, with an apparent rotation period of eight seconds, was spinning much more slowly than any radio pulsar then known. Even when not bursting, the object emitted a steady glow of x-rays with more radiant power than could be supplied by the rotation of a neutron star. Oddly, the star was significantly displaced from the center of the supernova remnant. If it was born at the center, as is likely, then it must have recoiled with a velocity of about 1,000 kilometers per second at birth. Such high speed was considered unusual for a neutron star.

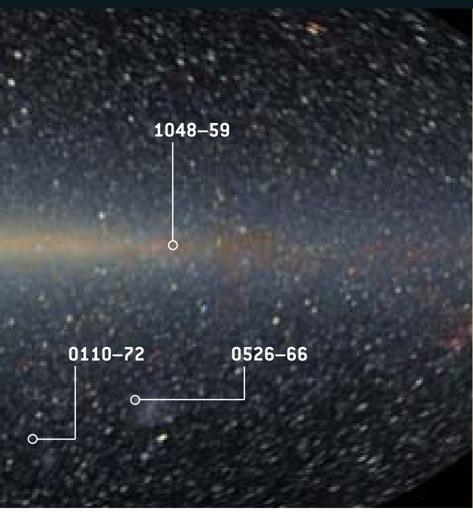
Finally, the outbursts themselves seemed inexplicable. X-ray flashes had previously been detected from some neutron stars, but they never exceeded the Eddington limit by very much. Astronomers ascribed them to thermonuclear fusion of hydrogen or helium or to the sudden accretion of matter onto the star. But the brightness of the SGR bursts was unprecedented, so a new physical mechanism seemed to be required.

Spin Forever Down

THE FINAL BURST FROM the March 1979 source was detected in May 1983; none has been seen since. Two other SGRs, both within our Milky Way galaxy, went off in 1979 and have remained active, emitting hundreds of bursts in subsequent years. A fourth SGR was located in 1998. Three of these four objects have possible, but unproved, associations with young supernova remnants. Two also lie near very dense clusters of massive young stars, intimating that SGRs tend to form from such stars. A fifth candidate SGR has gone off only twice; its precise location is still unknown.

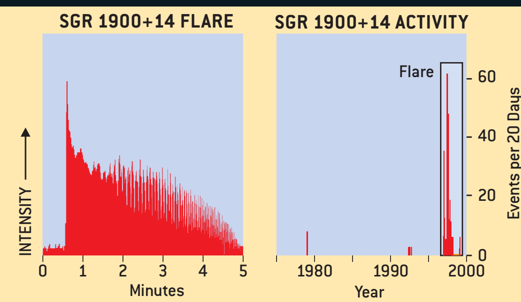
As Los Alamos National Laboratory scientists Baolian L. Cheng, Richard I. Epstein, Robert A. Guyer and A. Cody Young pointed out in 1996, SGR bursts are statistically similar to earthquakes. The energies have very similar mathematical distributions, with less energetic events being more common. Our then graduate student Ersin Gögüs of the University of Alabama at Huntsville verified this behavior for many bursts from various sources. This and other statistical properties are a hallmark of self-organized criticality, whereby a composite system attains a critical state in which a small perturbation can trigger a chain reaction. Such behavior occurs in systems as diverse as avalanches on sandpiles and magnetic flares on the sun.

But why would a neutron star behave like this? The solution emerged from an entirely separate line of work, on radio pulsars. Pulsars are widely thought to be rapidly rotating, magnetized neutron stars. The magnetic field, which is supported by electric currents flowing deep inside the star, rotates with the star. Beams of radio waves shine outward from the star's magnetic poles and sweep through space as it rotates, like lighthouse beacons—hence the observed pulsing. The pulsar also blows out a wind of charged particles and low-frequency elec-



NAME	YEAR OF DISCOVERY	ROTATION PERIOD (seconds)
SGR 0526–66	1979	8.0
SGR 1900+14	1979	5.16
SGR 1806–20	1979	7.47
SGR 1801–23*	1997	?
SGR 1627–41	1998	?
AXP 1E 2259+586	1981	6.98
AXP 1E 1048–59†	1985	6.45
AXP 4U 0142+61	1993	8.69
AXP 1RXS 1708–40†	1997	11.0
AXP 1E 1841–045	1997	11.8
AXP AXJ1844–0258	1998	6.97
AXP CXJ0110–7211†	2002	5.44
AXP XTE J1810–197	2003	5.54

*Not on map; location not known precisely †Abbreviated name



GIANT X-RAY FLARE in August 1998 confirmed the existence of magnetars. It started with a spike of radiation lasting less than a second (*left*). Then came an extended train of pulses with a period of 5.16 seconds. This event was the most powerful outburst to come from the object, designated SGR 1900+14, since its discovery in 1979 (*right*).

SOURCES FOR TABLE: CHRYSSA KOUVELIOTOU, ROBERT C. DUNCAN AND CHRISTOPHER THOMPSON tromagnetic waves, which carry away energy and angular momentum, causing its rate of spin to decrease gradually.

Perhaps the most famous pulsar lies within the Crab Nebula, the remnant of a supernova explosion that was observed in 1054. The pulsar rotates once every 33 milliseconds and is currently slowing at a rate of about 1.3 millisecond every century. Extrapolating backward, it was born rotating once every 20 milliseconds. Astronomers expect it to continue to spin down, eventually reaching a point when its rotation will be too slow to power the radio pulses. The spin-down rate has been measured for almost every radio pulsar, and theory indicates that it depends on the strength of the star's magnetic field. From this, most young radio pulsars are inferred to have magnetic fields between 10^{12} and 10^{13} gauss. For comparison, a refrigerator magnet has a strength of about 100 gauss.

The Ultimate Convection Oven

THIS PICTURE LEAVES a basic question unanswered: Where did the magnetic field come from in the first place? The traditional assumption was: it is as it is, because it was as it was. That is, most astronomers supposed that the magnetic field is a relic of the time before the star went supernova. All stars have weak magnetic fields, and those fields can be amplified simply by the act of compression. According to Maxwell's equations of electromagnetism, as a magnetized object shrinks by a factor of two, its magnetic field strengthens by a factor of four. The core of a massive star collapses by a factor of 10^5 from its birth through neutron star formation, so its magnetic field should become 10^{10} times stronger.

If the core magnetic field started with sufficient strength, this compression could explain pulsar magnetism. Unfortunately, the magnetic field deep inside a star cannot be measured, so this simple hypothesis cannot be tested. There are also good reasons to believe that compression is only part of the story.

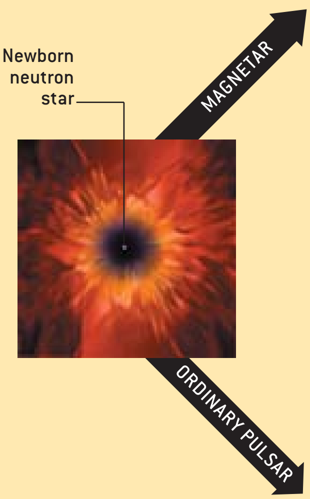
Within a star, gas can circulate by convection. Warm parcels of ionized gas rise, and cool ones sink. Because ionized gas conducts electricity well, any magnetic field lines threading the gas are dragged with it as it moves. The field can thus be reworked and sometimes amplified. This phenomenon, known as dynamo action, is thought to generate the magnetic fields of stars and planets. A dynamo might operate during each phase of the life of a massive star, as long as the turbulent core is rotating rapidly enough. Moreover, during a brief period after the core of the star turns into a neutron star, convection is especially violent.

This was first shown in computer simulations in 1986 by Adam Burrows of the University of Arizona and James M. Lattimer of Stony Brook University. They found that temperatures in a newborn neutron star exceed 30 billion kelvins.

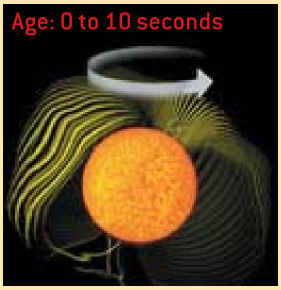
TWO TYPES OF NEUTRON STARS

1 Most neutron stars are thought to begin as massive but otherwise ordinary stars, between eight and 20 times as heavy as the sun.

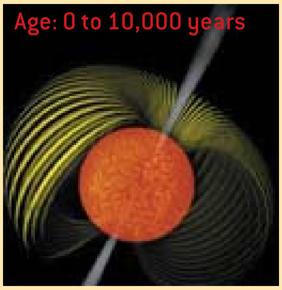
2 Massive stars die in a type II supernova explosion, as the stellar core implodes into a dense ball of subatomic particles.



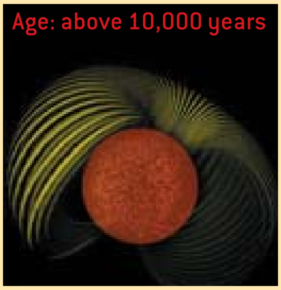
3 A If the newborn neutron star spins fast enough, it generates an intense magnetic field. Field lines inside the star get twisted.



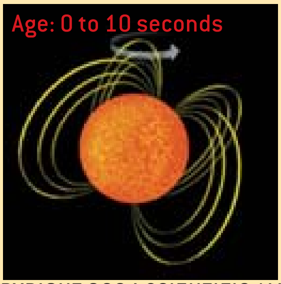
4 A The magnetar settles into neat layers, with twisted field lines inside and smooth lines outside. It might emit a narrow radio beam.



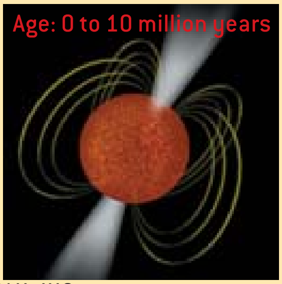
5 A The old magnetar has cooled off, and much of its magnetism has decayed away. It emits very little energy.



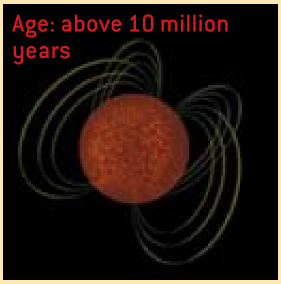
3 B If the newborn neutron star spins slowly, its magnetic field, though strong by normal standards, does not reach magnetar levels.



4 B The mature pulsar is cooler than a magnetar of equal age. It emits a broad radio beam, which radio telescopes can readily detect.



5 B The old pulsar has cooled off and no longer emits a radio beam.

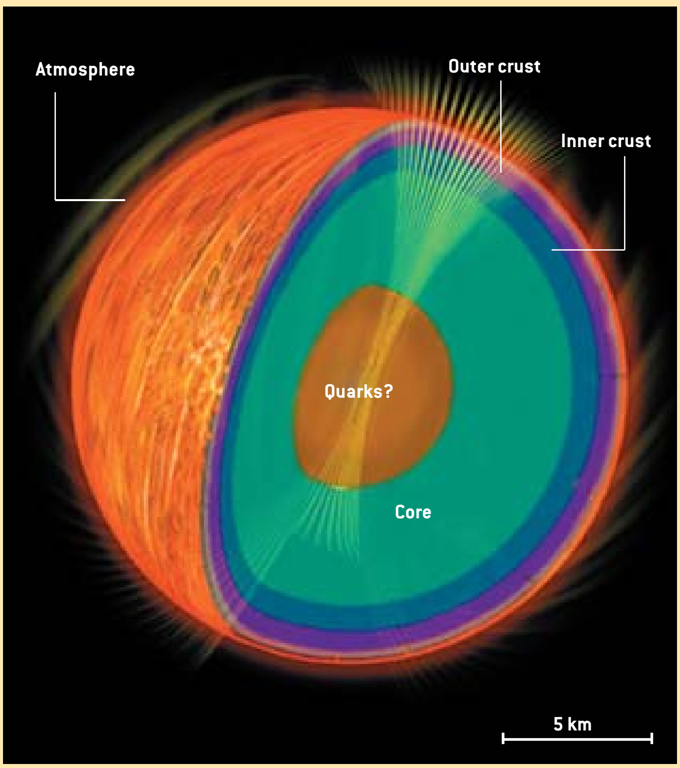


Hot nuclear fluid circulates in 10 milliseconds or less, carrying enormous kinetic energy. After about 10 seconds, the convection ceases.

Not long after Burrows and Lattimer conducted their first simulations, Duncan and Thompson, then at Princeton University, estimated what this furious convection means for neutron star magnetism. The sun, which undergoes a sedate version of the same process, can be used as a reference point. As solar fluid circulates, it drags along magnetic field lines and gives up about 10 percent of its kinetic energy to the field. If the moving fluid in a newborn neutron star also transfers a tenth of its kinetic energy to the magnetic field, then the field would grow stronger than 10^{15} gauss, which is more than 1,000 times as strong as the fields of most radio pulsars.

Whether the dynamo operates globally (rather than in limited regions) would depend on whether the star's rate of rotation was comparable to its rate of convection. Deep inside the sun, these two rates are similar, and the magnetic field is able to organize itself on large scales. By analogy, a neutron star born rotating as fast as or faster than the convective period of 10 milliseconds could develop a widespread, ultrastrong magnetic field. In 1992 we named these hypothetical neutron stars "magnetars."

An upper limit to neutron star magnetism is about 10^{17} gauss; beyond this limit, the fluid inside the star would tend to mix and the field would dissipate. No known objects in the universe can generate and maintain fields stronger than this level. One ramification of our calculations is that radio pulsars are neutron stars in which the large-scale dynamo has *failed* to operate. In the case of the Crab pulsar, the newborn neutron star rotated once every 20 milliseconds, much slower than the rate of convection, so the dynamo never got going.



DON DIXON

STRUCTURE OF A NEUTRON STAR can be inferred from theories of nuclear matter. Starquakes can occur in the crust, a lattice of atomic nuclei immersed in a sea of electrons. The core consists mainly of neutrons and perhaps a lump of quarks. An atmosphere of hot plasma might extend all of a few centimeters.

Crinkle Twinkle Little Magnetar

ALTHOUGH WE DID NOT develop the magnetar concept to explain SGRs, its implications soon became apparent to us. The magnetic field should act as a strong brake on a magnetar's rotation. Within 5,000 years a field of 10^{15} gauss would slow the spin rate to once every eight seconds—neatly explaining the oscillations observed during the March 1979 outburst.

As the field evolves, it changes shape, driving electric currents along the field lines outside the star. These currents, in turn, generate x-rays. Meanwhile, as the magnetic field moves through the solid crust of a magnetar, it bends and stretches the crust. This process heats the interior of the star and occasionally breaks the crust in a powerful "starquake." The accompanying release of magnetic energy creates a dense cloud of electrons and positrons, as well as a sudden burst of soft gamma rays—accounting for the fainter bursts that give SGRs their name.

More infrequently, the magnetic field becomes unstable and undergoes a large-scale rearrangement. Similar (but smaller) upheavals sometimes happen on the sun, leading to solar flares. A magnetar easily has enough energy to power a giant flare such as the March 1979 event. Theory indicates that the first half-second of that tremendous outburst came from an expanding fireball. In 1995 we suggested that part of the fireball was trapped by the magnetic field lines and held close to the star. This trapped fireball gradually shrank and then evaporated, emitting x-rays all the while. Based on the amount of energy released, we calculated the strength of the magnetic field needed to confine the enormous fireball pressure: greater than 10^{14} gauss, which agrees with the field strength inferred from the spin-down rate.

A separate estimate of the field had been given in 1992 by Bohdan Paczynski of Princeton. He noted that x-rays can slip

THE AUTHORS

CHRYSSA KOUVELIOTOU, ROBERT C. DUNCAN and *CHRISTOPHER THOMPSON* have studied magnetars for a collective 40 years and have collaborated for the past six. Kouveliotou, an observer, works at the National Space Science and Technology Center in Huntsville, Ala. Besides soft-gamma repeaters, her pets include gamma-ray bursts, x-ray binaries and her cat, Felix; her interests range from jazz to archaeology to linguistics. Duncan and Thompson are theorists, the former at the University of Texas at Austin, the latter at the Canadian Institute for Theoretical Astrophysics in Toronto. Duncan has studied supernovae, quark matter and intergalactic gas clouds. In his younger days he ran a 2:19 marathon in the 1980 U.S. Olympic trials. Thompson has worked on topics from cosmic strings to giant impacts in the early solar system. He, too, is an avid runner as well as a backpacker.

through a cloud of electrons more easily if the charged particles are immersed in a very intense magnetic field. For the x-rays during the burst to have been so bright, the magnetic field must have been stronger than 10^{14} gauss.

What makes the theory so tricky is that the fields are stronger than the quantum electrodynamic threshold of 4×10^{13} gauss. In such strong fields, bizarre things happen. X-ray photons readily split in two or merge together. The vacuum itself is polarized, becoming strongly birefringent, like a calcite crystal. Atoms are deformed into long cylinders thinner than the quantum-relativistic wavelength of an electron [*see box on opposite page*]. All these strange phenomena have observable effects on magnetars. Because this physics was so exotic, the theory attracted few researchers at the time.

Zapped Again

AS THESE THEORETICAL developments were slowly unfolding, observers were still struggling to see the objects that were the sources of the bursts. The first opportunity came when NASA's orbiting Compton Gamma Ray Observatory recorded a burst of gamma rays late one evening in October 1993. This was the break Kouveliotou had been looking for when she joined the Compton team in Huntsville. The instrument that registered the burst could determine its position only to within a fairly broad swath of sky. Kouveliotou turned for help to the Japanese ASCA satellite. Toshio Murakami of the Institute of Space and Astronautical Science in Japan and his collaborators soon found an x-ray source from the same swath of sky. The source held steady, then gave off another burst—proving beyond all doubt that it was an SGR. The same object had first been seen in 1979 and, based on its approximate celestial coordinates, was identified as SGR 1806–20. Now its position was fixed much more precisely, and it could be monitored across the electromagnetic spectrum.

The next leap forward came in 1995, when NASA launched the Rossi X-ray Timing Explorer (RXTE), a satellite designed to be highly sensitive to variations in x-ray intensity. Using this instrument, Kouveliotou found that the emission from SGR 1806–20 was oscillating with a period of 7.47 seconds—amazingly close to the 8.0-second periodicity observed in the March 1979 burst (from SGR 0526–66). Over the course of five years, the SGR slowed by nearly two parts in 1,000. Although the slowdown may seem small, it is faster than that of any radio pulsar known, and it implies a magnetic field approaching 10^{15} gauss.

More thorough tests of the magnetar model would require a second giant flare. Luckily, the heavens soon complied. In the early morning of August 27, 1998, some 19 years after the giant flare that began SGR astronomy was observed, an even more intense wave of gamma rays and x-rays reached Earth from the depths of space. It drove detectors on seven scientific spacecraft to their maximum or off scale. One interplanetary probe, NASA's Near Earth Asteroid Rendezvous mission, was forced into a protective shutdown mode. The gamma rays hit Earth on its nightside, with the source in the zenith over the mid-Pacific Ocean.

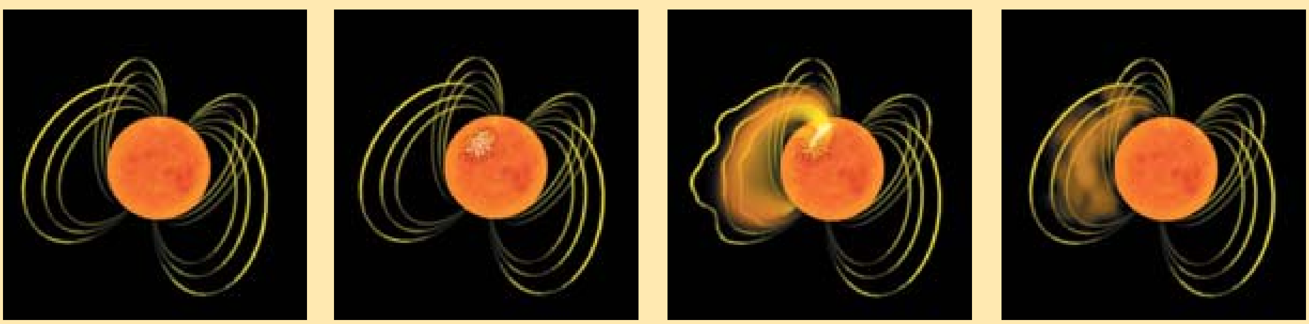
Fortuitously, in those early-morning hours electrical engineer Umran S. Inan and his colleagues from Stanford University were gathering data on the propagation of very low frequency radio waves around Earth. At 3:22 A.M. PDT, they noticed an abrupt change in the ionized upper atmosphere. The inner edge of the ionosphere plunged down from 85 to 60 kilometers for five minutes. It was astonishing. This effect on our planet was caused by a neutron star far across the galaxy, 20,000 light-years away.

Another Magneto Marvel

THE AUGUST 27 FLARE was almost a carbon copy of the March 1979 event. Intrinsically, it was only one tenth as powerful, but because the source was closer to Earth it remains the most intense burst of gamma rays from beyond our solar sys-

HOW MAGNETAR BURSTS HAPPEN

The magnetic field of the star is so strong that the rigid crust sometimes breaks and crumbles, releasing a huge surge of energy.



1 Most of the time the magnetar is quiet. But magnetic stresses are slowly building up.

2 At some point the solid crust is stressed beyond its limit. It fractures, probably into many small pieces.

3 This "starquake" creates a surging electric current, which decays and leaves behind a hot fireball.

4 The fireball cools by releasing x-rays from its surface. It evaporates in minutes or less.

DON DIXON

tem ever detected. The last few hundred seconds of the flare showed conspicuous pulsations, with a 5.16-second period. Kouveliotou and her team measured the spin-down rate of the star with RXTE; sure enough, it was slowing down at a rate comparable to that of SGR 1806–20, implying a similarly strong magnetic field. Another SGR was placed into the magnetar hall of fame.

The precise localizations of SGRs in x-rays have allowed them to be studied using radio and infrared telescopes (though not in visible light, which is blocked by interstellar dust). This work has been pioneered by many astronomers, notably Dale Frail of the National Radio Astronomy Observatory and Shrinivas R. Kulkarni of the California Institute of Technology. Other observations have shown that all four confirmed SGRs continue to release energy, albeit faintly, even between outbursts. "Faintly" is a relative term: this x-ray glow represents 10 to 100 times as much power as the sun radiates in visible light.

By now one can say that magnetar magnetic fields are better measured than pulsar magnetic fields. In isolated pulsars, almost the only evidence for magnetic fields as strong as 10^{12} gauss comes from their measured spin-down. In contrast, the combination of rapid spin-down and bright x-ray flares provides several arguments for 10^{14}- to 10^{15}-gauss fields in magnetars. Alaa Ibrahim of the NASA Goddard Space Flight Center reported yet another line of evidence for strong magnetic fields in magnetars: x-ray spectral lines that seem to be generated by protons gyrating in a 10^{15}-gauss field.

One intriguing question is whether magnetars are related to cosmic phenomena besides SGRs. The shortest-duration gamma-ray bursts, for example, have yet to be convincingly explained, and at least a handful of them could be flares from magnetars in other galaxies. If seen from a great distance, even a giant flare would be near the limit of telescope sensitivity. Only the brief, hard, intense pulse of gamma rays at the onset of the flare would be detected, so telescopes would register it as a GRB.

In the mid-1990s Thompson and Duncan suggested that magnetars might also explain anomalous x-ray pulsars, a class of objects that resemble SGRs in many ways. The one difficulty was that AXPs had not been observed to burst. But, Victoria M. Kaspi and Fotis P. Gavriil of McGill University and Peter M. Woods of the National Space and Technology Center in Huntsville detected bursts from two of the seven known AXPs. One of these objects is associated with a young supernova remnant in the constellation Cassiopeia.

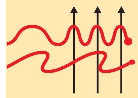
Another AXP in Cassiopeia is the first magnetar candidate to have been detected in visible light. Ferdi Hulleman and Marten van Kerkwijk of Utrecht University in the Netherlands, working with Kulkarni, spotted it. Though exceedingly faint, the AXP fades in and out with the x-ray period of a neutron star. These observations lend support to the idea that it is indeed a magnetar. The main alternative—that AXPs are ordinary neutron stars surrounded by disks of matter—predicts too much visible and infrared emission with too little pulsation.

In view of these discoveries, and the apparent silence of the Large Magellanic Cloud burster for nearly 20 years, it appears that magnetars can change their clothes. They can remain quiescent for years, even decades, before undergoing sudden periods of extreme activity. Some astronomers argue that AXPs are younger on average than SGRs, but this is still a matter of debate. If both SGRs and AXPs are magnetars, then magnetars plausibly constitute a substantial fraction of all neutron stars.

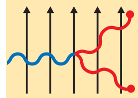
The story of magnetars is a sobering reminder of how much we have yet to understand about our universe. Thus far we have discerned at most a dozen magnetars among the countless stars. They reveal themselves for a split second, in light that only the most sophisticated telescopes can detect. Within 10,000 years, their magnetic fields freeze and they stop emitting bright x-rays. So those dozen magnetars betray the presence of more than a million, and perhaps as many as 100 million, other objects—old magnetars that long ago went dark. Dim and dead, these strange worlds wander through interstellar space. What other phenomena, so rare and fleeting that we have not recognized them, lurk out there? SA

EXTREME MAGNETISM

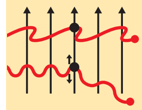
Magnetar fields wreak havoc with radiation and matter.



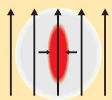
VACUUM BIREFRINGENCE
Polarized light waves (*orange*) change speed and hence wavelength when they enter a very strong magnetic field (*black lines*).



PHOTON SPLITTING
In a related effect, x-rays freely split in two or merge together. This process is important in fields stronger than 10^{14} gauss.



SCATTERING SUPPRESSION
A light wave can glide past an electron (*black circle*) with little hindrance if the field prevents the electron from vibrating with the wave.



DISTORTION OF ATOMS
Fields above 10^9 gauss squeeze electron orbitals into cigar shapes. In a 10^{14}-gauss field, a hydrogen atom becomes 200 times narrower.

BRYAN CHRISTIE DESIGN

MORE TO EXPLORE

Formation of Very Strongly Magnetized Neutron Stars: Implications for Gamma-Ray Bursts. Robert C. Duncan and Christopher Thompson in *Astronomical Journal,* Vol. 392, No. 1, pages L9–L13; June 10, 1992. Available at **http://adsabs.harvard.edu/abs/1992Apj...392L...9D**

An X-ray Pulsar with a Superstrong Magnetic Field in the Soft Gamma-Ray Repeater SGR1806–20. C. Kouveliotou, S. Dieters, T. Strohmayer, J. Von Paradijs, G. J. Fishman, C. A. Meegan, K. Hurley, J. Kommers, I. Smith, D. Frail and T. Murakami in *Nature,* Vol. 393, pages 235–237; May 21, 1998.

The Life of a Neutron Star. Joshua N. Winn in *Sky & Telescope,* Vol. 98, No. 1, pages 30–38; July 1999.

Physics in Ultra-strong Magnetic Fields. Robert C. Duncan. Fifth Huntsville Gamma-Ray Burst Symposium, February 23, 2002. Available at **arxiv.org/abs/astro-ph/0002442**

Flash! The Hunt for the Biggest Explosions in the Universe. Govert Schilling. Cambridge University Press, 2002.

More information can be found at Robert C. Duncan's Web site: **solomon.as.utexas.edu/magnetar.html**

Supersoft X-ray Stars

Several years ago astronomers came across a new type of star that spews out unusually low energy x-rays. These so-called supersoft sources are now thought to be white dwarf stars that cannibalize their stellar companions and then, in many cases, explode

By Peter Kahabka, Edward P. J. van den Heuvel and Saul A. Rappaport

DAVID AND GOLIATH STARS form a symbiotic binary system: a white dwarf and a red giant star in mutual orbit. The dwarf, with its intense gravity, is slurping off the outer layers of the giant. The pilfered gas goes into an accretion disk around the dwarf and eventually settles onto its surface, whereupon it can ignite nuclear fusion and generate a large quantity of low-energy x-rays.

and Supernovae

Since the 1930s astronomers have known that ordinary stars shine because of nuclear fusion deep in their interior. In the core of the sun, for example, 600 million tons of hydrogen fuse into helium every second. This process releases energy in the form of x-rays and gamma rays, which slowly wend their way outward through the thick layers of gas. By the time the radiation reaches the surface of the star, it has degraded into visible light.

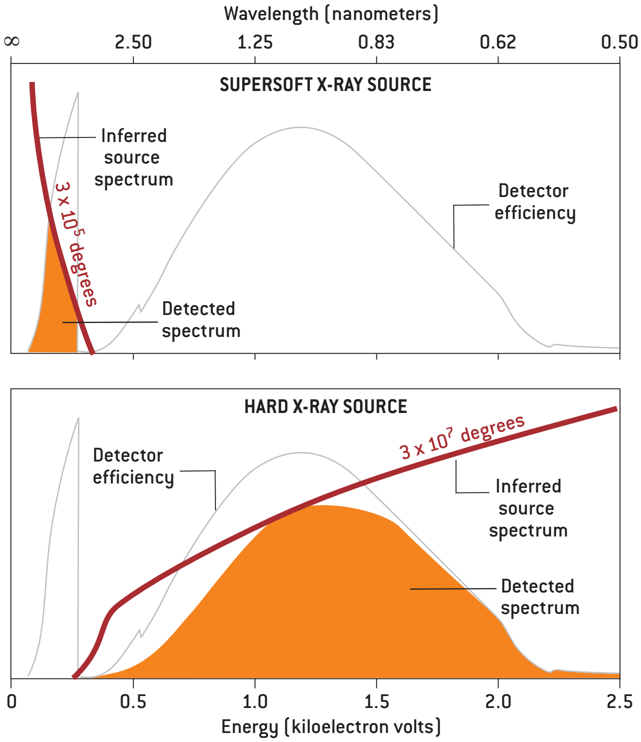
Recently, however, researchers have discovered a new class of stars in which the nuclear fusion takes place not in the deep interior but in the outer layers just below the surface. These stars appear to be white dwarfs—dense, burned-out stars that have exhausted their nuclear fuel—in orbit around ordinary stars. The dwarfs steal hydrogen gas from their companions, accumulate it on their surface and resume fusion. The result is a torrent of x-rays with a distinctive "soft" range of wavelengths; such stars are known as luminous supersoft x-ray sources. As the dwarfs gain weight, they eventually grow unstable, at which point they can collapse into an even denser neutron star or explode.

The disruption of white dwarfs has long been conjectured as the cause of one sort of supernova explosion, called type Ia. With the discovery of the supersoft sources, observers have identified for the first time a class of star system that can detonate in this way. Type Ia supernovae have become important as bright "standard candles" for measuring distances to faraway galaxies and thereby the pace of cosmic expansion. Much of the lingering uncertainty in estimates of the age and the expansion rate of the universe is connected to astronomers' ignorance of what gives rise to these supernovae. Supersoft sources may be one of the long-sought missing links.

The story of the supersoft sources began with the launch of the German x-ray satellite ROSAT in 1990. This orbiting observatory carried out the first complete survey of the sky in soft x-rays, a form of electromagnetic radiation that straddles ultraviolet light and the better-known "hard" x-rays. Soft x-rays have wavelengths that are one thousandth to one fiftieth those of visible light—which means that the energy of their photons (the unit x-ray astronomers prefer to think in) is between about 0.09 and 2.5 kiloelectron volts (keV). Hard x-rays have energies up to a few hundred keV. With the exception of the National Aeronautics and Space Administration's orbiting Einstein Observatory, which covered the energy range from 0.2 to 4.0 keV, previous satellites had concentrated on the hard x-rays.

Almost immediately the ROSAT team, led by Joachim Trümper of the Max Planck Institute for Extraterrestrial Physics near Munich, noticed some peculiar objects during observations of the Large Magellanic Cloud, a small satellite galaxy of the Milky Way. The objects emitted x-rays at a prodigious rate—some 5,000 to 20,000 times the total energy output of our sun—but had an unexpectedly soft spectrum. Bright x-ray sources generally have hard spectra, with peak energies in the range of 1 to 20 keV, which are produced by gas at temperatures of 10 million to 100 million kelvins. These hard x-ray sources represent neutron stars and black holes in the process of devouring their companion stars [see "X-ray Binaries," on page 58]. But the soft spectra of the new stars—with photon energies a hundredth of those in other bright x-ray sources—implied temperatures of only a few hundred thousand kelvins. On an x-ray color picture, the objects appear red, whereas classical, hard x-ray sources look blue [*see illustration at bottom left of next page*].

The reason the supersoft sources had not been recognized before as a separate class of star is that the earlier x-ray detectors were less sensitive to low energies. In fact, after the ROSAT findings, researchers went back through their archives and realized that two of the sources had been discovered 10 years earlier by Knox S. Long and his colleagues at the Columbia University



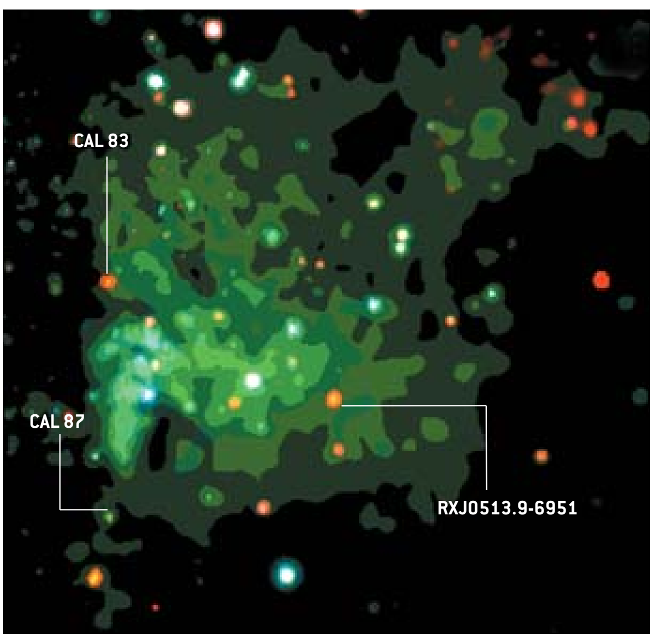
SOFT AND HARD x-ray sources are distinguished by their spectra, as measured by the ROSAT orbiting observatory. A typical supersoft source (*top*) emits x-rays with a fairly low energy, indicative of a comparatively cool temperature of 300,000 degrees Celsius. A typical hard x-ray source (*bottom*) is 100 times hotter and therefore emits higher-energy x-rays. In both cases, the intrinsic spectrum of the source (*red curves*) is distorted by the response of the ROSAT detector (*gray curves*) and by interstellar gas absorption.

Astrophysics Laboratory (CAL), using the Einstein Observatory. These sources, named CAL 83 and CAL 87, had not been classified as distinct from other strong sources in the Large Magellanic Cloud, although the Columbia team did remark that their spectra were unusually soft.

Back of the Envelope

AT THE TIME, Anne P. Cowley and her co-workers at Arizona State University surmised that CAL 83 and 87 were accreting black holes, which often have softer spectra than neutron stars do. This suggestion seemed to receive support in the 1980s, when faint stars were found at the locations of both sources. The stars' brightnesses oscillated, a telltale sign of a binary star system, in which two stars are in mutual orbit. In 1988 an international observing effort led by Alan P. Smale of University College London found that the brightness of CAL 83 fluctuated with a period of just over one day. A similar project led by Tim Naylor, now at the University of Exeter in England, obtained a period of 11 hours for CAL 87. These visible companion stars are the fuel for the hypothesized black holes. Assuming they have not yet been decimated, the various measurements indicated that they weighed 1.2 to 2.5 times as much as the sun.

But the ROSAT observations suddenly made this explanation very unlikely. The sources were much cooler than any known black hole system. Moreover, their brightness and temperature revealed their size. According to basic physics, each unit area of a star radiates an amount of energy proportional to the fourth power of its temperature. By dividing this energy into the total emission of the star, astronomers can easily calculate its surface area and, assuming it to be spherical,



X-RAY COLOR IMAGE (*left*) shows how a nearby mini galaxy, the Large Magellanic Cloud, might appear to someone with x-ray vision. A red color denotes lower-energy (or, equivalently, longer-wavelength) radiation; blue means higher energy (shorter wavelength). Supersoft sources stand out as red or orange dots; hard x-ray sources look blue. The supersoft star CAL 87 seems green because a cloud of hydrogen alters its true color. (Some red dots are actually sunlike stars in the foreground.) The view is rather different from an ordinary photograph of the area (*right*).

ALFRED T. KAMAJIAN (*preceding pages*); BRYAN CHRISTIE; SOURCE: PETER KAHABKA, EDWARD P. J. VAN DEN HEUVEL AND SAUL A. RAPPAPORT (*graphs*); WOLFGANG PIETSCH *ROSAT Team* (*bottom left*); SVEN KOHLE *University of Bonn* (*bottom right*) its diameter. It turns out that CAL 83, CAL 87 and the other Magellanic Cloud sources each have a diameter of 10,000 to 20,000 kilometers (16,000 to 32,000 miles)—the size of a white dwarf star. They are therefore 500 to 1,000 times as large as a neutron star or the "horizon" at the edge of a stellar-mass black hole. When Trümper first described the supersoft sources at a conference at the Santa Barbara Institute for Theoretical Physics in January 1991, several audience members quickly made this calculation on the proverbial back of the envelope.

Some conference participants, among them Jonathan E. Grindlay of Harvard University, suggested that the sources were white dwarfs that gave off x-rays as gas crashed onto their surface—much as hard x-ray sources result from the accretion of matter onto a neutron star or into a black hole. Others, including Trümper, his colleagues Jochen Greiner and Günther Hasinger, and, independently, Nikolaos D. Kylafis and Kiriaki M. Xilouris of the University of Crete, proposed that the sources were neutron stars that had somehow built up a gaseous blanket some 10,000 kilometers thick. In either case, the ultimate source of the energy would be gravitational. Gravity would pull material toward the dwarf or neutron star, and the energy of motion would be converted to heat and radiation during collisions onto the stellar surface or within the gas.

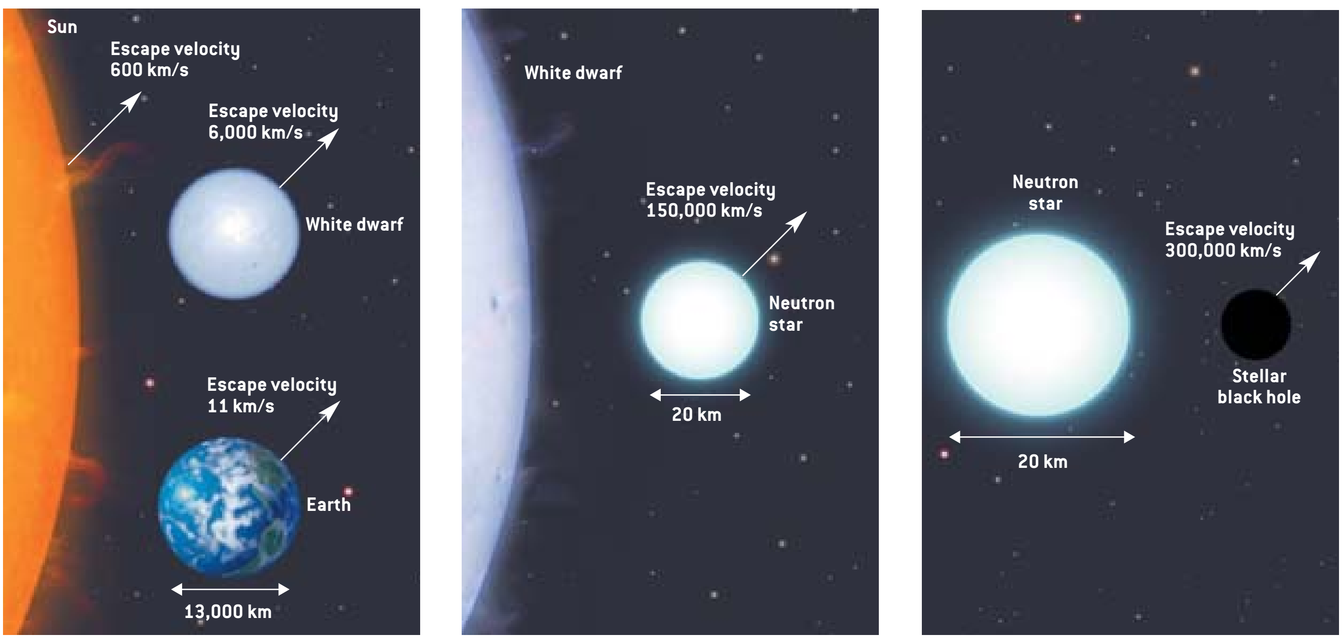
Both models seemed worth detailed study, and two of us (van den Heuvel and Rappaport), collaborating with Dipankar Bhattacharya of the Raman Research Institute in Bangalore, India, were lucky enough to be able to start such studies immediately. The conference was part of a half-year workshop at Santa Barbara, where several dozen scientists from different countries had the time to work together on problems related to neutron stars.

It soon became clear that neither model worked. The supersoft sources emit about the same power as the brightest accreting neutron stars in binaries. Yet gas collisions onto neutron stars are 500 to 1,000 times as forceful as the same process on white dwarfs, because the effect of gravity at the surface of a neutron star is that much greater. (For bodies of the same mass, the available gravitational energy is inversely proportional to the radius of the body.) Thus, for a dwarf to match the output of a neutron star, it would need to sweep up material at 500 to 1,000 times the rate. In such a frenetic accretion flow—equivalent to several Earth masses a year—the incoming material would be so dense that it would totally absorb any x-rays.

Neutron stars with gaseous blankets also ran into trouble. Huge envelopes of gas (huge, that is, with respect to the 10-kilometer radius of the neutron star) would be unstable; they would either collapse or be blown away in a matter of seconds or minutes. Yet CAL 83 and CAL 87 had been shining for at least a decade. Indeed, the ionized interstellar gas nebula surrounding CAL 83 took many tens of thousands of years to create.

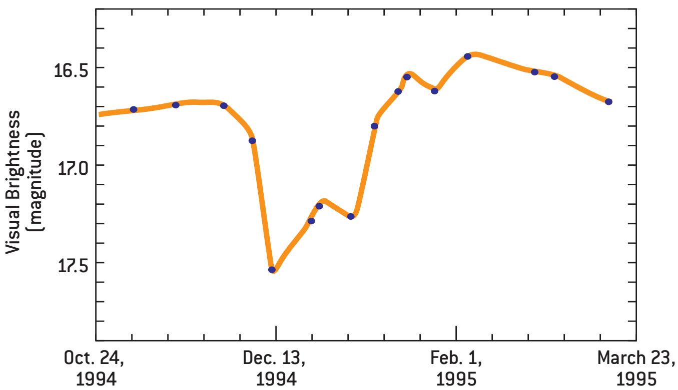
Nuclear Power

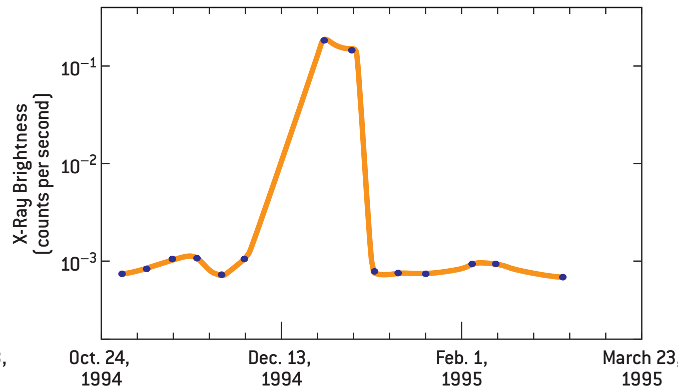
AFTER WEEKS OF DISCUSSING and evaluating models, none of which worked, astrophysicists realized the crucial difference between accretion of material onto neutron stars or black holes and accretion onto white dwarfs. The former generates much more energy than nuclear fusion of the same amount of hydrogen, whereas the latter produces much less en-



ALFRED T. KAMAJIAN

COMPACT STARS have colossal escape velocities. A typical white dwarf (*left*) packs the mass of the sun into the volume of a planet. To break free of its gravity, an object must travel at some 6,000 kilometers per second. This is also about the speed that a body doing the reverse trip—falling onto the dwarf from afar—would have on impact. Denser stars, such as neutron stars with the same mass (*center*), have an even mightier grip. The densest possible star, a black hole, is defined by a "horizon" from which the escape velocity equals the speed of light (*right*).





ON/OFF EMISSION of supersoft star RXJ0513.9-6951 is a sign that it is poised between two different modes of behavior. When it shines brightly in visible light (*left graph*), its x-ray output (*right graph*) is low, and vice versa. (The lower x-ray counts are upper limits.) The star is at the border between a pure supersoft source (which would emit only x-rays) and a white dwarf surrounded by thick gas (which would emit only visible light). Slight fluctuations in the rate of gas intake switch the star from one behavior to the other.

BRYAN CHRISTIE; SOURCE: KLAUS REINSCH *University of Göttingen (left)* AND STEFAN G. SCHAEIDT *Max Planck Institute for Extraterrestrial Physics (right)*

ergy than fusion. Of the energy inherent in mass (Albert Einstein's famous $E = mc^2$), fusion releases 0.7 percent. Accretion onto a neutron star, however, liberates more than 10 percent; into a black hole, up to 46 percent before the material disappears completely. On the other hand, accretion onto a white dwarf, with its comparatively weak gravity, liberates only about 0.01 percent of the inherent energy.

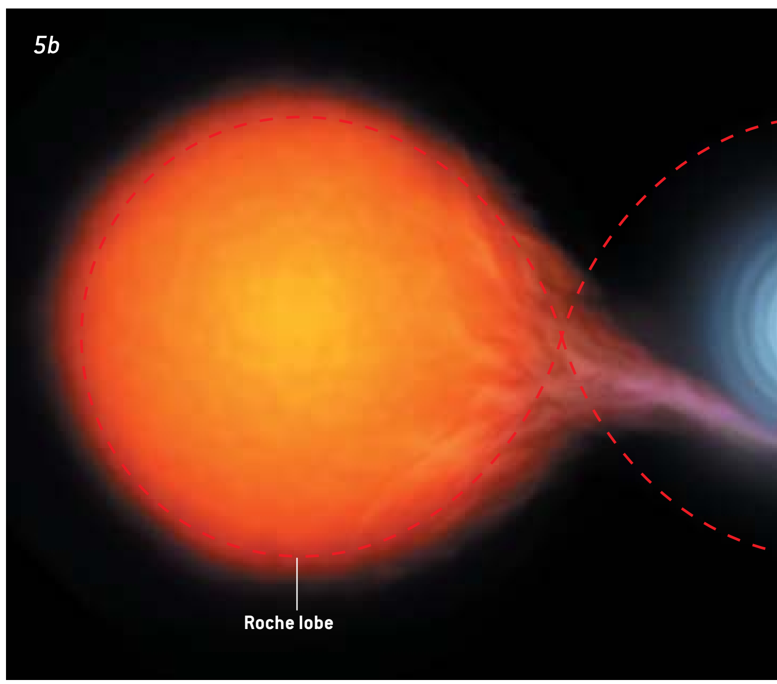
Therefore, on white dwarfs, nuclear fusion is potentially more potent than accretion. If hydrogen accumulated on the surface of a white dwarf and somehow started to "burn" (that is, undergo fusion), only about 0.03 Earth mass would be needed a year to generate the observed soft x-ray luminosity. Because of the lower density of inflowing matter, the x-rays would be able to escape.

Stable nuclear burning of inflowing matter would account for the paradoxical brightness of the supersoft sources. But is it really possible? Here we were lucky. Just when we were discussing this issue, Ken'ichi Nomoto of the University of Tokyo arrived in Santa Barbara. He had already been trying to answer the very same question in order to understand another phenomenon, nova explosions—outbursts much less energetic than supernovae that cause a star suddenly to brighten 10,000-fold but do not destroy it. Novae always occur in close binaries that consist of a white dwarf and a sunlike star. Until the discovery of supersoft sources, they were the only known close binaries.

For over a decade, Nomoto and others had been improving on the pioneering simulations by Bohdan Paczynski and Anna Zytkow, both then at the Nicolaus Copernicus Astronomical Center in Warsaw. According to these analyses, hydrogen that has settled onto the surface of a dwarf can indeed burn. The style of burning depends on the rate of accretion. If it is sufficiently low, below 0.003 Earth mass a year, fusion is spasmodic. The newly acquired hydrogen remains passive, often for thousands of years, until its accumulated mass exceeds a critical value, at which point fusion is abruptly ignited at its base. The ensuing thermonuclear explosion is visible as a nova.

If the accretion rate is slightly higher, fusion is cyclic but not explosive. As the rate increases, the interval between burning cycles becomes shorter and shorter, and above a certain threshold value, stable burning sets in. For white dwarfs of one solar mass, this threshold is about 0.03 Earth mass a year. In the simulations, fusion generates exactly the soft x-ray luminosity observed in the supersoft sources.

LIFE CYCLE of a supersoft star (*sequence 1–6, on opposite page*) begins with an unequal binary star system and ends with a type Ia supernova explosion. The supersoft phase can take one of three forms, depending on the companion star. If it is an ordinary star in a tight orbit, it can overflow its Roche lobe and cede control of its outer layers to the white dwarf (*5a, on opposite page*). If the companion is a red giant star of sufficient size, it also overflows its Roche lobe (*5b, at right*). Finally, if it is a red giant with a smaller size or a wider orbit, it can power a supersoft source with its strong winds (*5c, on opposite page*). Not all supersoft sources blow up, but enough do to account for the observed rate of supernovae.



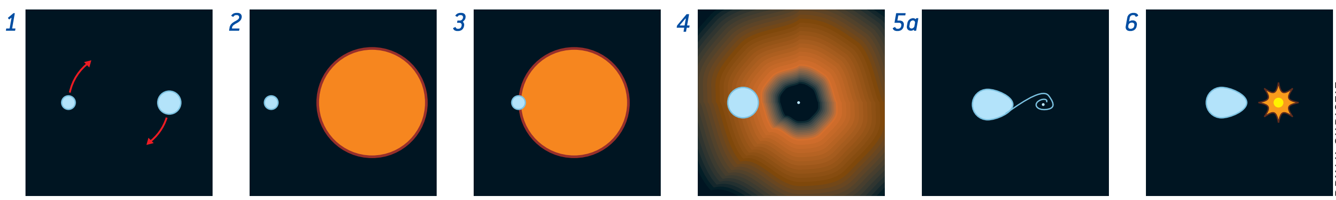
If the rate is still higher, above 0.12 Earth mass a year, the incoming gas does not settle onto the surface but instead forms an extended envelope around the dwarf. Steady burning continues on the surface, but the thick envelope degrades the x-rays into ultraviolet and visible light. Recent calculations have shown that the radiation is so intense that it exerts an outward pressure on gas in the envelope, causing part of it to stream away from the star in a stellar wind.

If the accretion rate hovers around 0.12 Earth mass a year, the system may alternate between x-ray and visible phases. Exactly this type of behavior has been found in the supersoft source known as RXJ0513.9-6951, which was discovered by Stefan G. Schaeidt of the Max Planck institute. It gives off x-rays for weeks at a time, with breaks of several months. This on/off emission puzzled astronomers until 1996, when Karen A. Southwell and her colleagues at the University of Oxford noticed that the visible counterpart to this star fluctuated, too. When the visible star is faint, the x-ray source is bright, and vice versa [*see top illustration on opposite page*]. The system also features two high-speed jets of matter flowing out in opposite directions at an estimated 4,000 to 6,000 kilometers per second. Such jets are common where an accretion disk dumps more material on the star than it can absorb. The excess squirts out in a direction perpendicular to the disk, where there is no inflowing matter to block it. The outflow velocity is expected to be approximately the same as the escape velocity from the surface of the star. In RXJ0513.9-6951 the inferred speed nearly equals the escape velocity from a white dwarf—further confirmation that the supersoft sources are white dwarfs.

Soft-Boiled Star

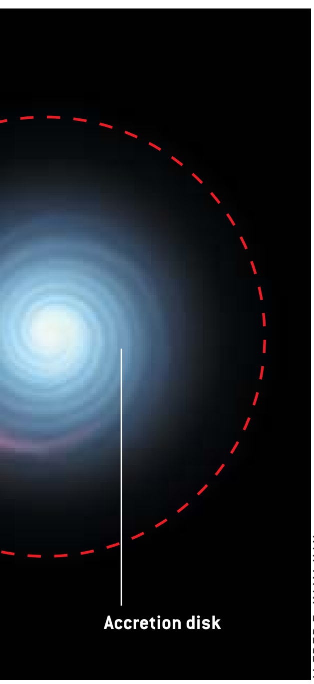
NOT EVERY BINARY SYSTEM can supply material at the rates required to produce a supersoft source. If the companion star is less massive than the white dwarf, as is typically observed in nova-producing systems, the fastest that material can flow in is 0.0003 Earth mass a year. This limit is a consequence of the law of conservation of orbital angular momentum. As the small companion star loses mass, its orbit widens and the flow rate stabilizes.

For the rates to be higher, the donor star must have a mass greater than that of the dwarf. Then the conservation of angular momentum causes the orbit to shrink as a result of the mass transfer. The stars come so close that they begin a gravi-

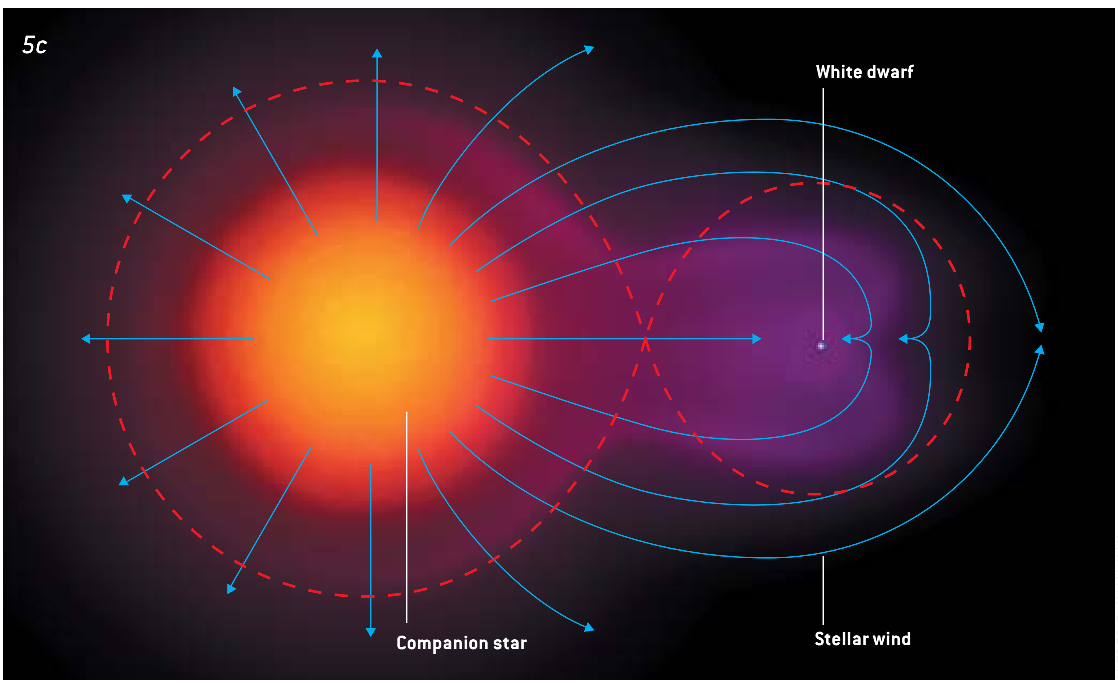


BRYAN CHRISTIE

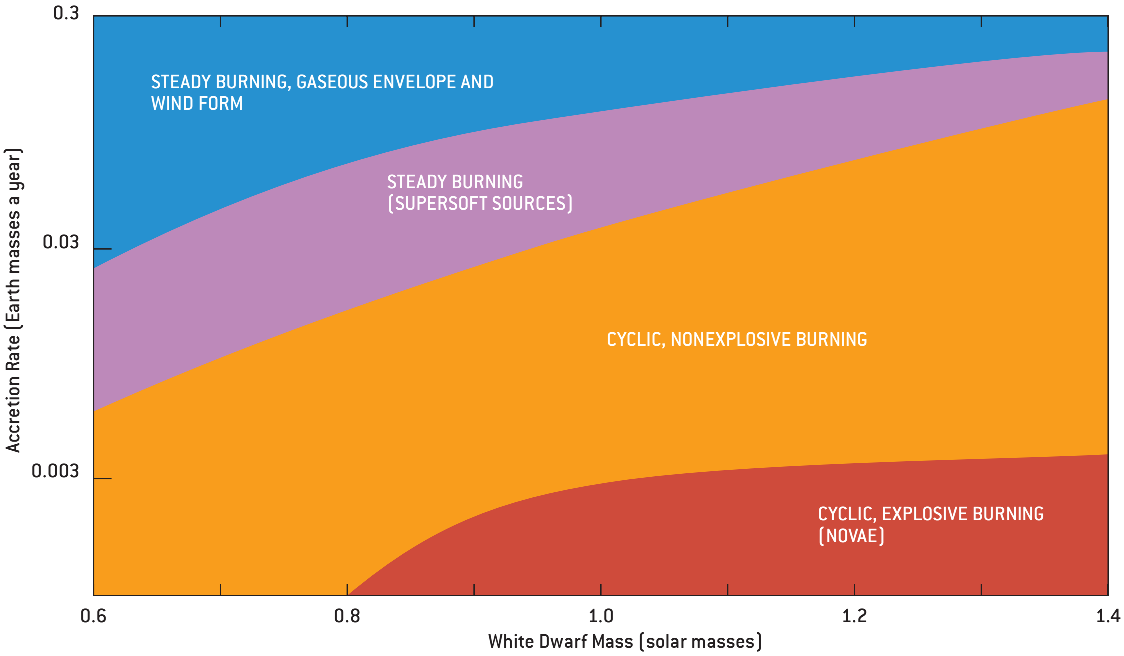
Pair of ordinary stars burn hydrogen in their cores | One exhausts fuel in core, becomes red giant | Orbit tightens; giant envelops other star | Giant sheds outer layers, becomes white dwarf | Dwarf steals gas from other star, emits soft x-rays | Dwarf reaches critical mass, explodes



ALFRED T. KAMAJIAN



ALFRED T. KAMAJIAN



STYLE OF NUCLEAR FUSION on the surface of a white dwarf depends on how massive the dwarf is and how fast it is devouring its companion star (*vertical axis*). If the accretion rate is sufficiently low, fusion (which astronomers misleadingly call "burning") occurs in spurts, either gently or explosively. Otherwise it is continuous. As shown above, phenomena once thought to be distinct—such as novae and supersoft sources—are closely related.

tational tug-of-war for control of the outer layers of the donor. Material within a certain volume called the Roche lobe remains under the sway of the donor's gravity, while material beyond it is stripped off by the dwarf. Perversely, the donor abets its own destruction. While it sheds mass at the surface, the amount of energy generated by fusion in the core remains largely unaffected. The continued heating from below exerts pressure on the outer layers to maintain the original shape of the star. This pressure replenishes the material ripped off the dwarf, much as an overflowing pot of soup on a hot burner will continue to pour scalding water onto the stove. The situation stabilizes only when the effects of mass loss are felt by the core itself. For a star originally of two solar masses, the return to equilibrium—and thus the cessation of supersoft emission—takes seven million years after the onset of plundering. By this time the star has shrunk to a fifth of its initial mass and become the lesser star in the system. The average accretion rate onto the dwarf in such a case is about 0.04 Earth mass a year.

Following this reasoning, we predicted in 1991 that many supersoft sources would be white dwarfs in tight orbits (with periods of less than a few days) around a companion star whose original mass was 1.2 to 2.5 solar masses. In fact, CAL 83 and 87 are precisely such systems. Since 1992 orbital periods for four more supersoft sources have been measured; all periods were less than a few days. The explanation may also apply to a class of novalike binary systems, called V Sagittae stars, whose oscillating brightness has perplexed astronomers for a century. In 1998 Joseph Patterson of Columbia and his collaborators and, independently, Joao E. Steiner and Marcos P. Diaz of the National Astrophysical Laboratory in Itajubá, Brazil, demonstrated that the prototype of this class of stars has the appropriate mass and orbital period.

There is one other group of star systems that could give rise to supersoft sources: so-called symbiotic binaries, in which the white dwarf is in a wide orbit about a red giant star. Red giants are willing donors. Bloated by age, they have relatively weak surface gravity and already discharge matter in strong stellar winds. In 1994 one of us (Kahabka), Hasinger and Wolfgang Pietsch of the Max Planck Institute discovered a supersoft symbiotic binary in the Small Magellanic Cloud, another satellite galaxy of the Milky Way. Since then, a further half dozen such sources have been found.

Some supersoft sources are harder to recognize because their accretion rate varies with time. One source in our galaxy alternates between x-ray and visible emission on a cycle of 40 years, as seen on archival photographic plates. A few objects, such as Nova Muscae 1983 and Nova Cygni 1992, combine nova be-

THE AUTHORS

PETER KAHABKA, EDWARD P. J. VAN DEN HEUVEL and *SAUL A. RAPPAPORT* never thought supersoft sources would be explained by white dwarfs. That insight came about during a workshop that van den Heuvel and Rappaport organized on a different topic: neutron stars. Two years later these veteran astronomers met Kahabka, who had discovered many supersoft sources as a member of the ROSAT team. Today Kahabka is research associate at the University of Bonn in Germany. Van den Heuvel is director of the Astronomical Institute at the University of Amsterdam and the 1995 recipient of the Spinoza Award, the highest science award in the Netherlands. An amateur archaeologist, he owns an extensive collection of early Stone Age tools. Rappaport is a physics professor at the Massachusetts Institute of Technology. He was one of the pioneers of x-ray astronomy in the 1970s.

BRYAN CHRISTIE; SOURCE: ICKO IBEN *University of Illinois* havior with supersoft emission, which can be explained by a years-long period of sedate "afterburning" between eruptions.

The Seeds of Supernovae

THE COMPANION MASSES required of supersoft sources with short orbital periods imply that they are relatively young systems (compared with the age of our galaxy). Stars of the inferred mass live at most a few billion years and are always located in or near the youthful central plane of the galaxy. Unfortunately, that location puts them in the region thick with interstellar clouds, which block soft x-rays. For this reason, the observed population is only the tip of the iceberg. Extrapolating from the known number of supersoft sources, we have estimated that the total number in our galaxy at any one time is several thousand. A few new ones are born every 1,000 years, and a few others die.

What happens as they pass away? The fusion of matter received from the companion clearly causes the white dwarf to grow in mass. It could reach the Chandrasekhar limit of about 1.4 solar masses, the maximum mass a white dwarf can have. Beyond this limit, the quantum forces that hold up the dwarf falter. Depending on the initial composition and mass of the dwarf, there are two possible outcomes: collapse to a neutron star or destruction in a nuclear fireball. Dwarfs that either lack carbon or are initially larger than 1.1 solar masses collapse. A number of theorists have analyzed this fate.

White dwarfs that do not meet either of these criteria simply blow up. They may slowly amass helium until they reach the Chandrasekhar limit and explode. Alternatively, the helium layer may reach a critical mass prematurely and ignite itself explosively. In the latter case, shock waves convulse the star and ignite the carbon at its core. And once the carbon burning begins, it becomes a runaway process in the dense, taut material of the dwarf. Within a few seconds the star is converted largely into nickel as well as other elements between silicon and iron. The nickel, dispersed into space, radioactively decays to cobalt and then iron in a few hundred days. Astronomers had already ascribed a kind of explosion to the death of carbon-rich dwarfs—the supernova type Ia. The spectrum of such a supernova lacks any sign of hydrogen or helium, one of the factors that distinguish it from the other types of supernovae (Ib, Ic and II), which all probably result from the implosion and subsequent explosion of massive stars. Type Ia supernovae are thought to be a major source of iron and related elements throughout the universe, including on Earth. Four occur every 1,000 years on average in a galaxy such as the Milky Way.

Before supersoft sources were discovered, astronomers were unsure as to the precise sequence that led to type Ia supernovae. The leading explanations implicated either certain symbiotic stars—in particular, the rare recurrent novae—or mergers of two carbon-rich white dwarfs. But the latter view is now disputed. Although recently double-dwarf systems with the necessary mass and orbital period have been discovered, calculations by Nomoto and his colleague Hadeyuki Saio have shown that such a merger could in many cases be too gentle to produce a thermonuclear explosion and instead would lead to the formation of a neutron star. Supersoft sources and other surface-burning dwarfs seem a good alternative solution. Their death rate roughly matches the observed supernova frequency. The concordance makes the luminous supersoft binary x-ray sources the first firmly identified class of objects that can realistically be expected to end their lives as type Ia supernovae.

This new realization may improve the accuracy of cosmological measurements that rely on these supernovae to determine distance [see "Surveying Space-time with Supernovae," by Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff; SCIENTIFIC AMERICAN, January 1999]. Subtle variations in brightness can make all the difference between conflicting conclusions concerning the origin and fate of the universe. The worry for cosmologists has always been that slight systematic errors—the product, perhaps, of astronomers' incomplete understanding of the stars that go supernova—could mimic real variations. The implications of the supersoft findings for cosmology, however, have yet to be worked out.

When supersoft sources were first detected, nobody expected that the research they provoked would end up uniting so many phenomena into a single coherent theory. Now it is clear that a once bewildering assortment of variable stars, novae and supernovae are all variants on the same basic system: an ordinary star in orbit around a reanimated white dwarf. The universe seems that much more comprehensible. SA

Nuclear burning of inflowing matter WOULD ACCOUNT FOR THE PARADOXICAL BRIGHTNESS of the supersoft sources. But is it really possible?

MORE TO EXPLORE

Luminous Supersoft X-ray Sources. P. Kahabka and E.P.J. van den Heuvel in *Annual Review of Astronomy and Astrophysics,* Vol. 35, pages 69–100. Annual Reviews, 1997.

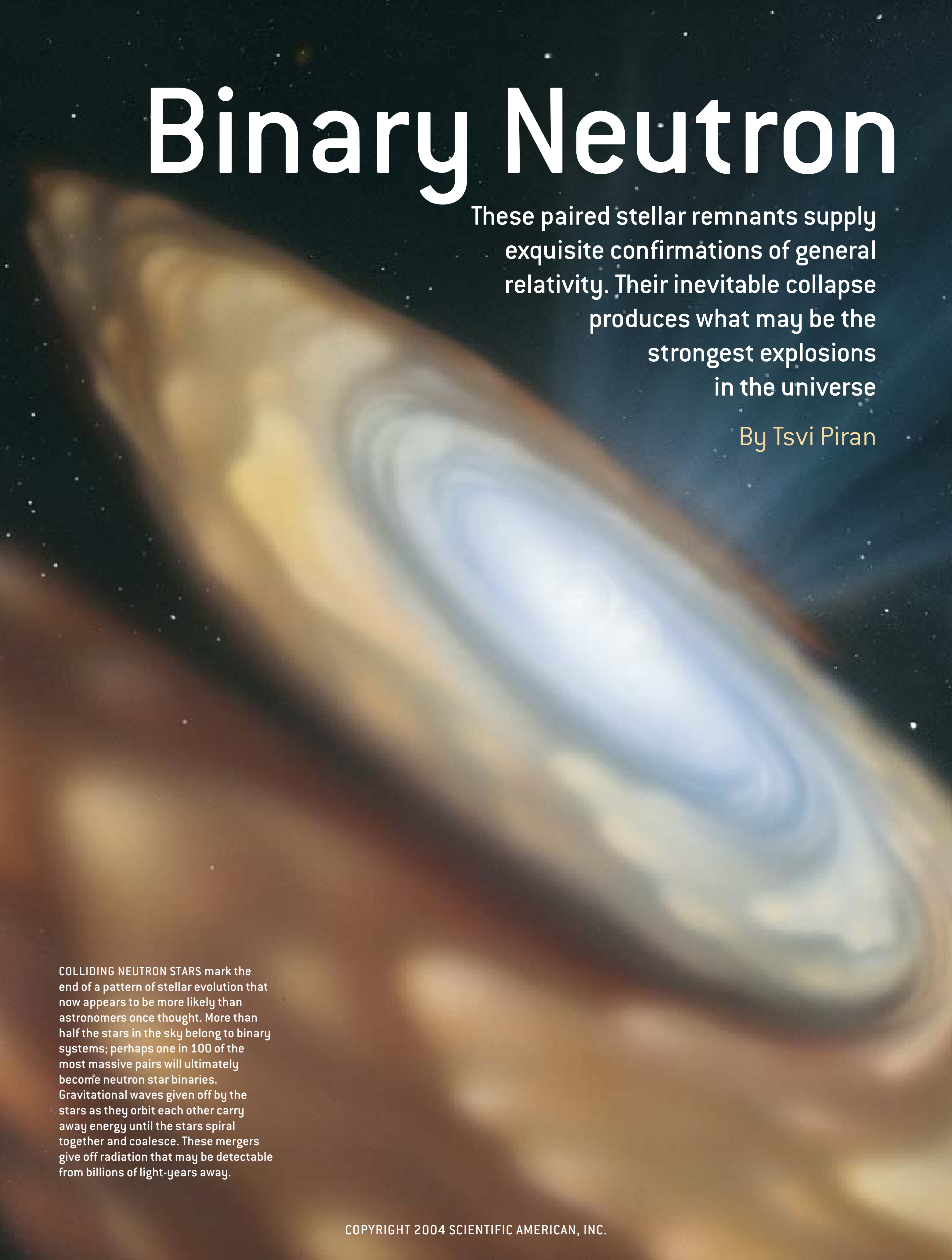
SNeIa: On the Binary Progenitors and Expected Statistics. Pilar Ruiz-Lapuente, Ramon Canal and Andreas Burkert in *Thermonuclear Supernovae.* Edited by Ramon Canal, Pilar Ruiz-Lapuente and Jordi Isern. Kluwer, 1997. Preprint available at **xxx.lanl.gov/abs/astro-ph/9609078**

Type Ia Supernovae: Their Origin and Possible Applications in Cosmology. Ken'ichi Nomoto, Koichi Iwamoto and Nobuhiro Kishimoto in *Science,* Vol. 276, pages 1378–1382; May 30, 1997. Preprint available at **xxx.lanl.gov/abs/astro-ph/9706007**

Binary Neutron

These paired stellar remnants supply exquisite confirmations of general relativity. Their inevitable collapse produces what may be the strongest explosions in the universe

By Tsvi Piran



COLLIDING NEUTRON STARS mark the end of a pattern of stellar evolution that now appears to be more likely than astronomers once thought. More than half the stars in the sky belong to binary systems; perhaps one in 100 of the most massive pairs will ultimately become neutron star binaries. Gravitational waves given off by the stars as they orbit each other carry away energy until the stars spiral together and coalesce. These mergers give off radiation that may be detectable from billions of light-years away.

Stars

ALFRED T. KAMAJIAN

In 1967 Jocelyn Bell and Antony Hewish found the first pulsar. Their radio telescope brought in signals from a source that emitted very regular pulses every 1.34 seconds. After eliminating terrestrial sources and provisionally discarding the notion that these signals might come from extraterrestrial intelligent beings, they were baffled. It was Thomas Gold of Cornell University who realized that the pulses originated from a rotating neutron star, beaming radio waves into space like a lighthouse. Researchers soon tuned in other pulsars.

Even as Bell and Hewish were making their discovery, military satellites orbiting Earth were detecting the signature of even more exotic signals: powerful gamma-ray bursts from outer space. The gamma rays triggered detectors intended to monitor illicit nuclear tests, but it was not until six years later that the observations were made public; even then, another 20 years passed before the bursts' origin was understood. Many people now think gamma-ray bursts are emitted by twin neutron stars in the throes of coalescence.

The discovery of binary neutron stars fell to Russell A. Hulse and Joseph H. Taylor, Jr., then at the University of Massachusetts at Amherst, who began a systematic pulsar survey in 1974. They used the Arecibo radio telescope in Puerto Rico, the largest in the world, and within a few months had found 40 previously unknown pulsars. Among their haul was a strange source by the name of PSR 1913+16 (PSR denotes a pulsar, and the numbers stand for its position in the sky: 19 hours and 13 minutes longitude and a declination of 16 degrees). It emitted approximately 17 pulses per second, but the period of the pulses changed by as much as 80 microseconds from one day to the next. Pulsars are so regular that this small fluctuation stood out clearly.

Hulse and Taylor soon found that the timing of the signals varied in a regular pattern, repeating every seven hours and 45 minutes. This signature was not new; for many years astronomers have noted similar variations in the wavelength of light from binary stars (stars that are orbiting each other). The Doppler effect shortens the wavelength (and increases the frequency) of signals emitted when a source is moving toward Earth and increases wavelength (thus decreasing the frequency) when a source is moving away. Hulse and Taylor concluded that PSR 1913+16 was orbiting a companion star, even though available models of stellar evolution predicted only solitary pulsars.

The surprises did not end there. Analysis of the time delay indicated that the pulsar and its companion were separated by a mere 1.8 million kilometers. At that distance, a normal star (with a radius of roughly 600,000 kilometers) would almost certainly have blocked the pulsar's signal at some point during its orbit. The companion could also not be a white dwarf (radius of about 3,000 kilometers), because tidal interactions would have perturbed the orbit in a way that contradicted the observations. Hulse and Taylor concluded that the companion to PSR 1913+16 must be a neutron star.

This finding earned the two a Nobel Prize in Physics in 1993. Astronomers have since mastered the challenge of understanding how binary neutron stars might exist at all, even as they have employed the signals these strange entities produce to conduct exceedingly fine tests of astrophysical models and of general relativity.

Birth from Death

BY ALL THE ASTROPHYSICAL theories that existed before 1974, binary neutron stars should not have existed. Astronomers believed that the repeated stellar catastrophes needed to create them would disrupt any gravitational binding between two stars.

Neutron stars are the remnants of massive stars, which perish in a supernova explosion after exhausting all their nuclear fuel. The death throes begin when a star of six solar masses or more consumes the hydrogen in its center, expands and becomes a red giant. At this stage, its core is already extremely dense: several solar masses within a radius of several thousand kilometers. An extended envelope more than 100 million kilometers across contains the rest of the mass. In the core, heavier elements such as silicon undergo nuclear fusion to become iron.

When the core reaches a temperature of several billion kelvins, the iron nuclei begin to break apart, absorbing heat from their surroundings and reducing the pressure in the core drastically. Unable to support itself against its own gravitational attraction, the core collapses. As its radius decreases from several thousand kilometers to 15, electrons and protons fuse into neutrons, leaving a very dense star of 1.4 solar masses in a volume no larger than an asteroid.

Meanwhile the energy released in the collapse heats the envelope of the star, which for a few weeks emits more light than an entire galaxy. Observations of old supernovae, such as the Crab Nebula's, whose light reached Earth in A.D. 1054, reveal a neutron star surrounded by a luminous cloud of gas, still moving out into interstellar space.

More than half the stars in the sky belong to binary systems. As a result, it is not surprising that at least a few massive pairs should remain bound together even after one of them undergoes a supernova explosion. The pair then becomes a massive x-ray binary, so named for the emission that the neutron star produces as it strips the outer atmosphere from its companion. Eventually the second star also explodes as a supernova and turns into a neutron star. The envelope ejected by the second supernova contains most of the mass of the binary (since the remaining neutron star contains a mere 1.4 solar masses). The ejection of such a large fraction of the total mass should therefore disrupt the binary and send the two neutron stars flying into space at hundreds of kilometers per second.

As the two stars approach they tear material from each other, THEN COALESCE WITHIN A FRACTION OF A SECOND.

Hulse and Taylor's discovery demonstrated, however, that some binaries survive the second supernova explosion. In retrospect, astronomers realized that the second supernova explosion might be asymmetrical, thereby propelling the newly formed neutron star into a stable orbit rather than out into the void. The second supernova also may be less disruptive if the second star loses its envelope gradually during the massive x-ray binary phase. Since then, the discovery of other neutron star binaries shows that other massive pairs have managed to survive the second supernova.

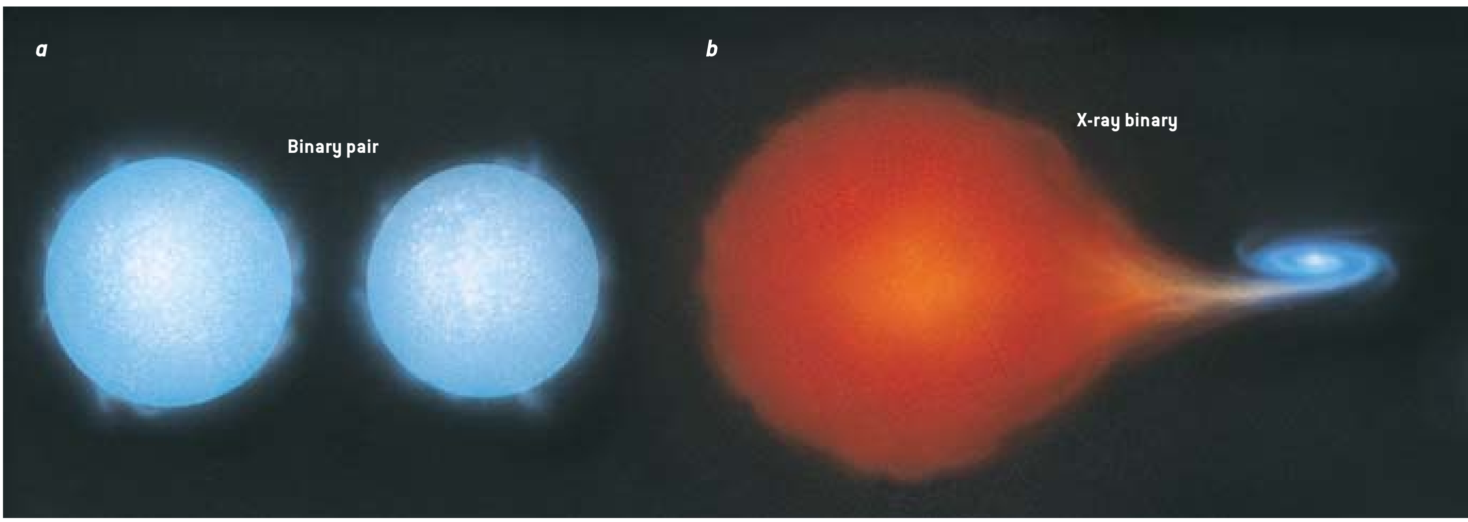
In the early 1990s Ramesh Narayan of Harvard University, Amotz Shemi of Tel Aviv University and I, along with E. Sterl Phinney of the California Institute of Technology, working independently, estimated that about 1 percent of massive x-ray binaries survive to form neutron star binaries. This figure implies that our galaxy contains a population of about 30,000 neutron star binaries. Following a similar line of argument, we also concluded that there should be a comparable number of binaries, yet unobserved, containing a neutron star and a black hole. Such a pair would form when one of the stars in a massive pair formed a supernova remnant containing more than about two solar masses and so collapsed to a singularity instead of a neutron star. Rarer, but still possible in theory, are black hole binaries, which start their lives as a pair of particularly massive stars; they should number about 300 in our galaxy.

Testing General Relativity

PSR 1913+16 has implications that reach far beyond the revision of theories of binary stellar evolution. Hulse and Taylor quickly realized that their discovery had provided an ideal site for testing Einstein's general theory of relativity.

Although this theory is accepted today as the only viable description of gravity, it has had just a few direct tests. Albert Einstein himself computed the precession of Mercury's orbit (the shift of the orbital axes and the point of Mer-

ALFRED T. KAMAJIAN



cury's closest approach to the sun) and showed that observations agreed with his theory. A few other tests followed. In 1964 Irwin I. Shapiro of Harvard pointed out that light signals bent by a gravitational field should be delayed in comparison to those that take a straight path. He measured the delay by bouncing radar signals off other planets in the solar system. Although general relativity passed these tests with flying colors, they were all carried out in the (relativistically) weak gravitational field of the solar system. That fact left open the possibility that general relativity might break down in stronger gravitational fields.

Because a pulsar is effectively a clock orbiting in the strong gravitational field of its companion, relativity makes a range of clear predictions about how the ticks of that clock (the pulses) will appear from Earth. First, the Doppler effect causes a periodic variation in the pulses' arrival time (the pattern that first alerted Taylor and Hulse).

A "second-order" Doppler effect, resulting from time dilation caused by the pulsar's rapid motion, leads to an additional (but much smaller) variation. This second-order effect can be distinguished because it depends on the square of velocity, which varies as the pulsar moves along its elliptical orbit. The second-order Doppler shift combines with the gravitational redshift, a slowing of the pulsar's clock when it is in the stronger gravitational field closer to its companion.

Like Mercury, PSR 1913+16 precesses in its orbit about its companion. The intense gravitational fields, however, mean that the periastron—the nadir of the orbit—rotates by 4.2 degrees a year, compared with Mercury's perihelion shift of a mere 42 arc seconds a century. The measured effects match the predictions of relativistic theory precisely. Remarkably, the precession and other information supplied by the timing of the radio pulses make it possible to calculate the masses of the pulsar and its companion: 1.442 and 1.386 solar masses, respectively, with an uncertainty of 0.003 solar mass. This precision is impressive for objects 15,000 light-years away.

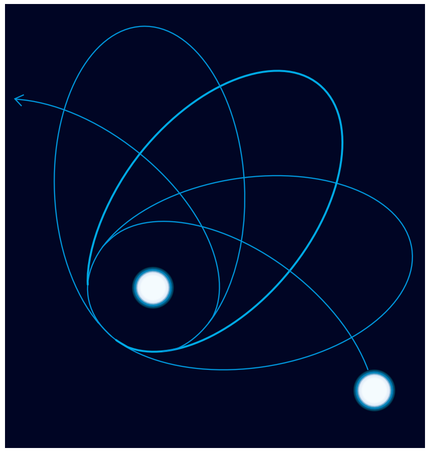
In 1991 Alexander Wolszczan of the Arecibo observatory found another binary pulsar that is almost a twin to PSR 1913+16. Each neutron star weighs between 1.27 and 1.41 solar masses. The Shapiro time delay, which was only marginally measured in PSR 1913+16, stands out clearly in signals from the pulsar that Wolszczan discovered.

Measurements of PSR 1913+16 have also revealed a relativistic effect never seen before. In 1918, several years after the publication of his general theory of relativity, Einstein predicted the existence of gravitational radiation, an analogue to electromagnetic radiation. He said massive particles that move with varying acceleration emit gravitational waves, small ripples in the gravitational field that also propagate at the speed of light.

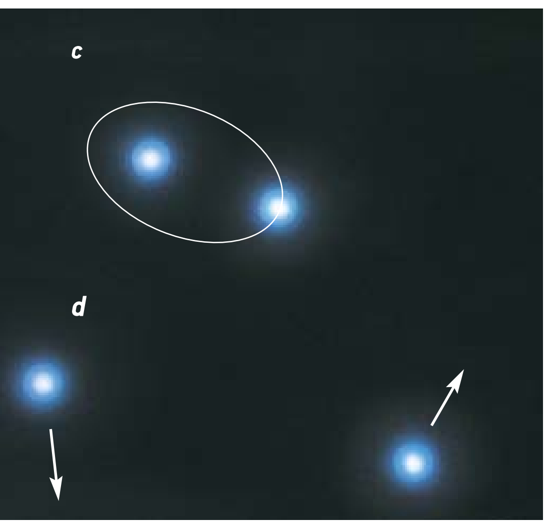
These ripples exert forces on other masses; if two objects are free to move, the distance between them will vary with the frequency of the wave. The size of the oscillation depends on the separation of the two objects and the strength of the waves. In principle, all objects whose acceleration varies emit gravitational radiation. Most objects are so small and move so slowly, however, that their gravitational radiation is utterly insignificant.

Binary pulsars are one of the few exceptions. In 1941, long before the discovery of the binary pulsar, Russian physicists Lev D. Landau and Evgenii M. Lifshitz calculated the effect of this emission on the motion of a binary. Energy conservation requires that the energy carried away by the waves come from somewhere, in this case the orbital energy of the two stars. As a result, the distance between them must decrease.

PSR 1913+16 emits gravitational radiation at a rate of eight quadrillion gigawatts, about a fifth as much energy as the total radiation output of the sun. This luminosity is impressive as far as gravitational radiation sources are concerned but still too weak to be detected directly on Earth. Nevertheless, it has a noticeable effect on the pulsar's orbit. The distance between the two neutron stars decreases by a few meters a year, which suffices to produce a detectable variation in the timing of the radio pulses. By carefully monitoring the pulses from PSR 1913+16 over the years, Taylor and his collaborators have shown that the orbital separation decreases in exact agreement with the predictions of the general theory of relativity.



ORBITAL PRECESSION, the rotation of the major axis of an elliptical orbit, results from relativistic perturbations of the motion of fast-moving bodies in intense gravitational fields. It is usually almost undetectable; Mercury's orbit precesses by less than 0.12 of a degree every century, but that of PSR 1913+16 changes by 4.2 degrees a year.

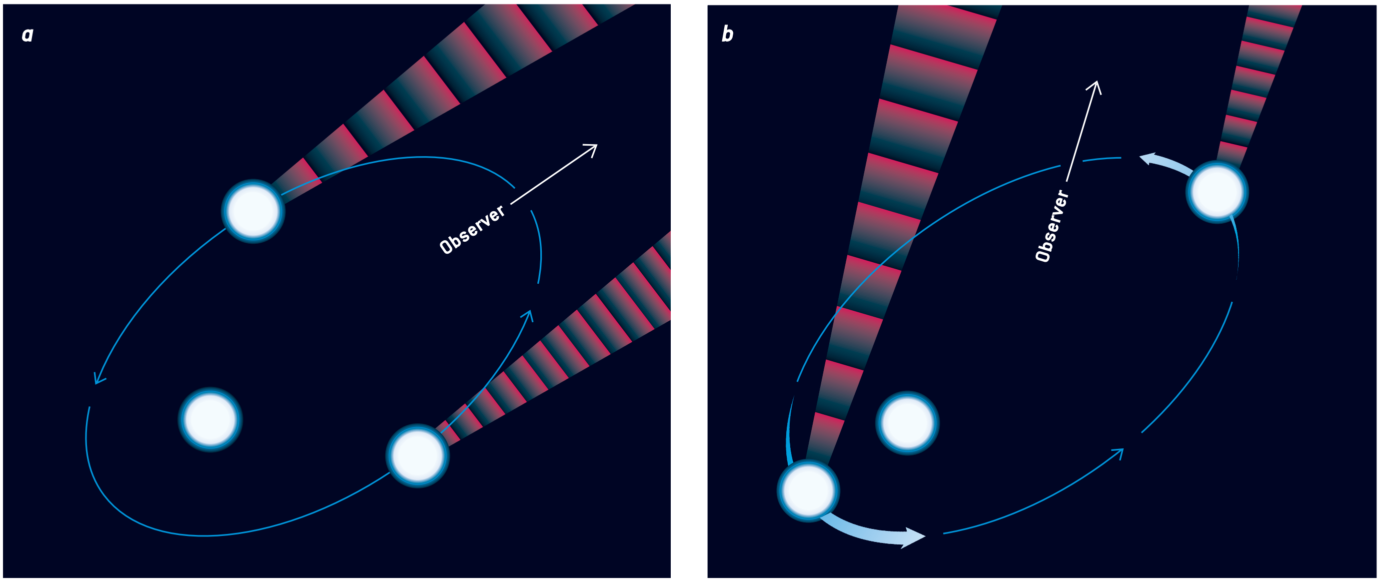


MASSIVE BINARY (*a*) evolves through a sequence of violent events. The heavier star in the pair burns its fuel faster and undergoes a supernova explosion; if the two stars stay bound together, the result is a massive x-ray binary (*b*) in which the neutron star remnant of the first star strips gas from its companion and emits x-radiation. Eventually the second star also exhausts its fuel. In roughly one of 100 cases, the resulting explosion leaves a pair of neutron stars orbiting each other (*c*); in the other 99, the two drift apart (*d*). There are enough binary star systems that a typical galaxy contains thousands of neutron star binaries.

GEORGE RETSECK (*above*)

FIRST-ORDER DOPPLER EFFECT

SECOND-ORDER DOPPLER EFFECT



BINARY PULSAR SIGNALS are affected by relativistic phenomena. (Each illustration shows one of the effects whose combination produces the observed timing of the pulses.) The Doppler effect slows the rate at which pulses reach an observer when the pulsar is moving away from Earth in its orbit and increases the rate when the pulsar is moving toward Earth (*a*). The second-order Doppler effect and the gravitational redshift (*b*) impose

GEORGE RETSECK

The reduction in the distance between the stars can be compared with the other general relativistic effects to arrive at a further confirmation. Just as measurements of the orbital decay produce a mathematical function relating the mass of the pulsar to the mass of its companion, so do the periastron shift and the second-order Doppler effect. All three functions intersect at precisely the same point.

Undetectable Cataclysms

AT PRESENT, THE DISTANCE between PSR 1913+16 and its companion is decreasing only slowly. As the gap shrinks, the gravitational-wave emission will increase, and the orbital decay will accelerate. Eventually the neutron stars will fall toward each other at a significant fraction of the speed of light, collide and merge. The 300 million years until the two coalesce are long on a human scale but rather short on an astronomical one.

Given the number of neutron star binaries in the galaxy, one pair should merge roughly every 300,000 years, a cosmological blink of the eye. Extrapolating this rate to other galaxies implies that throughout the observable universe about one neutron star merger occurs every 20 minutes—frequently enough that astronomers should consider whether they can detect such collisions.

To figure out whether such occurrences are detectable requires a solid understanding of just what happens when two orbiting neutron stars collide. Shortly after the discovery of the first binary pulsar, Paul Clark and Douglas M. Eardley, then both at Yale University, concluded that the final outcome is a black hole. Current estimates of the maximum mass of a neutron star range between 1.4 and 2.0 solar masses. Rotation increases the maximal mass, but most models suggest that even a rapidly rotating neutron star cannot be significantly larger than 2.4 solar masses. Because the two stars together contain about 2.8 solar masses, collapse to a singularity is almost inevitable.

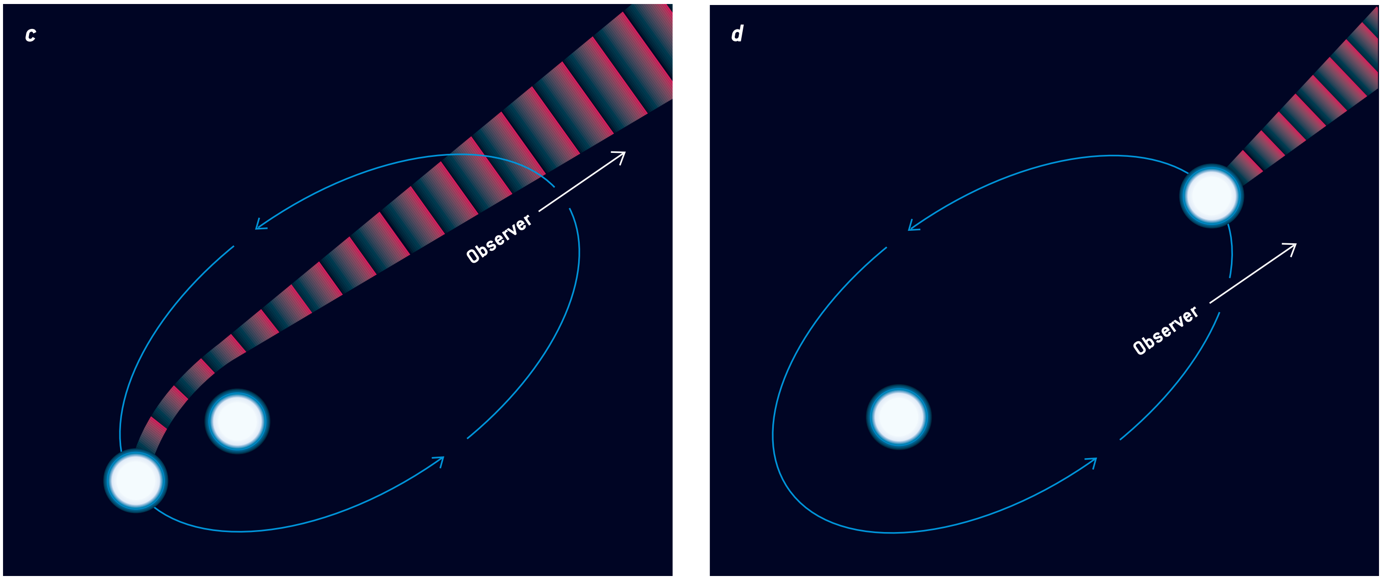
Melvyn B. Davies of Caltech, Willy Benz of the University of Arizona, Friedrich K. Thielemann of the Harvard-Smithsonian Center for Astrophysics and I have simulated the last moments of a neutron star binary in detail. The two objects are very dense and so behave effectively like point masses until they are quite close to each other. Tidal interaction between the stars becomes important only when they approach to within 30 kilometers, about twice the radius of a neutron star. At that stage, they begin to tear material from each other—about two tenths of a solar mass in total. Once the neutron stars touch, within a tiny fraction of a second they coalesce. The matter torn from the stars before the collision forms a disk around the central core and eventually spirals back into it.

What kinds of signals will this sequence of events generate? Clark and Eardley realized that the colliding stars will warm up to several billion kelvins. They figured that most of the thermal energy would be radiated as neutrinos and antineutrinos, much as it is in a supernova. Unfortunately, these weakly interacting, massless particles, which escape from the dense neutron star much more easily than do photons, are almost undetectable. When supernova 1987A exploded, the three detectors on Earth caught 21 neutrinos out of the 5×10^{46} joules of radiation. Although the burst expected from a binary neutron star merger is slightly larger than that of a supernova, the typical event takes place

THE AUTHOR

TSVI PIRAN has studied general relativity and astrophysics for 30 years. He is a professor at the Hebrew University of Jerusalem, where he received his Ph.D. in 1976. He has also worked at the University of Oxford, Harvard University, Kyoto University and Fermilab. Piran, with Steven Weinberg of the University of Texas, established the Jerusalem Winter School for Theoretical Physics.

SHAPIRO TIME DELAY



a similar variation because the pulsar's internal clock slows when it moves more rapidly in its orbit closest to its companion (*longer arrow*). Most subtle is the Shapiro time delay, which occurs as the gravitational field of the companion bends signals passing near it (*c*). The signals travel farther than if they took a straight-line path (*d*) and so arrive later. This effect is undetectable in PSR 1913+16 but is clear in other systems.

much farther away than the mere 150,000 light-years of SN 1987A. To detect one merger a year would require picking up signals one sixteen-millionth the intensity of the 1987 event. Because current neutrino detectors must monitor interactions within thousands of tons of material, it is difficult to imagine the apparatus that would be required. Furthermore, supernovae are 1,000 times more frequent than are neutron star collisions. Even if we detected a neutrino burst from two neutron stars, it is unlikely we would be able to distinguish it among the supernova neutrino bursts.

Before it emits its neutrino burst, the neutron star binary sends out a similarly energetic (but not quite so undetectable) train of gravitational waves. During the 15 minutes before coalescence, the two stars cover the last 700 kilometers between them, and their orbital period shrinks from a fifth of a second to a few milliseconds. The resulting signal is just in the optimal range for terrestrial gravitational-wave detectors.

An international network of such detectors has been built in the U.S. and in Italy. The American Caltech-M.I.T. team has detectors for the Laser Interferometer Gravitational-Wave Observatory (LIGO) near Hanford in Washington State and near Livingston, La. The French-Italian team has its VIRGO facility near Pisa in Italy. The first detectors were designed to be able to detect neutron star mergers up to 70 million light-years away; current estimates suggest that there is only one event per 100 years up to this distance. Researchers have proposed to improve their instruments dramatically over time to be able to detect neutron star mergers as far away as three billion light-years—several hundred a year.

High-Energy Photons

FOR SEVERAL YEARS after the discovery of PSR 1913+16, I kept wondering whether there was a way to estimate what fraction of the coalescing stars' binding energy is emitted as electromagnetic radiation. Even if this fraction is tiny, the binding energy is so large that the resulting radiation would still be enormous. Furthermore, photons are much easier to detect than neutrinos or gravitational waves, and so mergers could be detected even from the most distant parts of the universe.

In 1987 J. Jeremy Goodman of Princeton University, Arnon Dar of the Technion-Israel Institute of Technology and Shmuel Nussinov of Tel Aviv University noticed that about a tenth of a percent of the neutrinos and antineutrinos emitted by a collapsing supernova core collide with one another and annihilate to produce electron-positron pairs and gamma rays. In a supernova the absorption of these gamma rays by the star's envelope plays an important role in the explosion of the outer layers.

In 1989 David Eichler of Ben Gurion University of the Negev, Mario Livio, then at the Technion, David N. Schramm of the University of Chicago and I speculated that a similar fraction of the neutrinos released in a binary neutron star merger would also produce electron-positron pairs and gamma rays. The colliding neutron stars, however, have no envelope surrounding them, and so the gamma rays escape in a short, intense burst.

Gamma-ray bursts might arise from a more complex mechanism. The disk that forms during the neutron star merger falls back onto the central coalesced object within a few seconds, but during that time it, too, can trigger emissions. In 1992 Bohdan Paczynski of Princeton, Narayan of Harvard and I suggested that the rotation of the disk could intensify the neutron-star magnetic fields entangled in the disk's material, causing giant magnetic flares, a scaled-up version of the flares that rise from the surface of the sun. These short-lived magnetic disturbances could generate gamma-ray bursts in the same way that solar flares produce

NASA'S SWIFT SATELLITE, TO LAUNCH THIS FALL, **may determine whether a partner in a neutron star merger** MORE NATURALLY ENDS AS A SUPERNOVA OR A BLACK HOLE.

gamma rays and x-rays. The large variability in the observed bursts implies that both mechanisms may be at work.

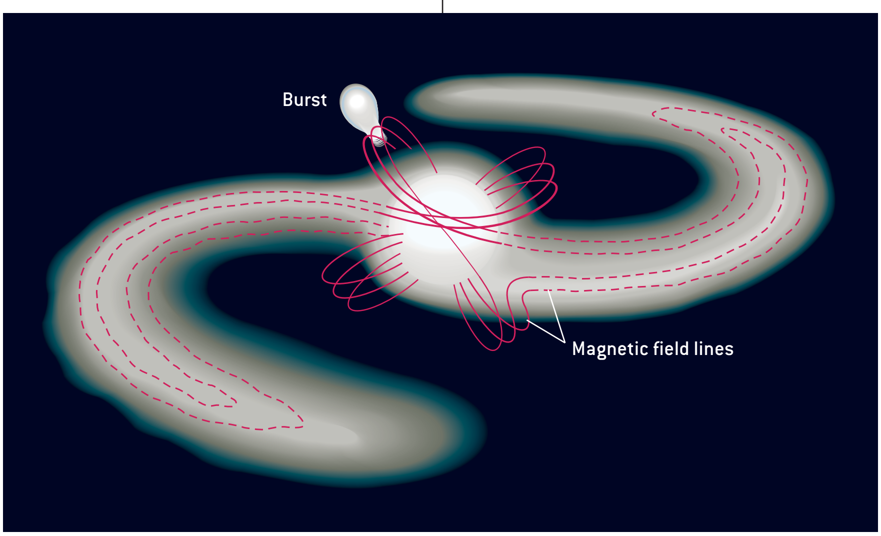
Puzzle Unscrambled

HAD IT NOT BEEN FOR the Limited Test Ban Treaty of 1963, we would not have known about these bursts for decades. No one would have proposed a satellite to look for them, and had such a proposal been made it would surely have been turned down as too speculative. But the U.S. Department of Defense launched a series of satellites known as Vela, which carried omnidirectional x-ray and gamma-ray detectors to verify that no one was testing nuclear warheads in space.

These spacecraft never detected a nuclear explosion, but as soon as the first satellite was launched it began to detect entirely unexpected bursts of high-energy photons in the range of several hundred kiloelectron volts. Bursts lasted between a few dozen milliseconds and about 30 seconds. The lag between the arrival time of the bursts to different satellites indicated that the sources were outside the solar system. Still, the bursts were kept secret for several years, until in 1973 Ray W. Klebesadel, Ian B. Strong and Roy A. Olson of Los Alamos National Laboratory described them in a seminal paper. Theorists proposed more than 100 models in the next 20 years; in the late 1980s a consensus formed that the bursts originated on neutron stars in our own galaxy.

A minority led by Paczynski argued that the bursts originated at cosmological distances. In the spring of 1991 the Compton Gamma Ray Observatory, which was more sensitive than any previous gamma-ray satellite, was launched by NASA. It revealed two unexpected facts. First, the distribution of burst intensities is not homogeneous in the way that it would be if the bursts were nearby. Second, the bursts came from all across the sky rather than being concentrated in the plane of the Milky Way, as they would be if they originated in the galactic disk. Together these facts demonstrate that the bursts do not originate from the disk of our galaxy. A lively debate still prevails over the possibility that the bursts might originate from the distant parts of the invisible halo of our galaxy, but as the Compton observatory collects more data, this hypothesis seems less and less likely. It seems that the minority was right.

In the fall of 1991, I analyzed the distribution of burst intensities, as did Paczynski and his colleague Shude Mao. We concluded that the most distant bursts seen by the Compton observatory came from several billion light-years away. The cosmological origin of gamma-ray bursts was indeed confirmed several years later.



GAMMA-RAY BURSTS may also result from a mechanism similar to the one that powers solar flares. The magnetic field from the coalesced neutron stars is amplified as it winds through the disk of material thrown out during the merger. The field accelerates charged particles until they emit gamma rays.

GEORGE RETSECK

A Rare Laboratory

PSR 1913+16 HAS been a great tool for astrophysics, but it will disappear from our sight by about 2020. According to the general theory of relativity, a pulsar's spin should precess and its direction in space should vary. A significant precession will carry the beam away from us altogether. In 1998 Michael Kramer of the University of California at Berkeley finally determined that the pulsar's beam is pointing toward Earth only 60 years out of every 300. Our distant offspring will be able to see it again around the year 2260.

Several other binary pulsar systems have been discovered, bringing the total known to seven. Most remarkable is J0307-3039, found in December 2003 by Marta Burgay, a Ph.D. student at the University of Bologna in Italy, and an international team of researchers. A few weeks later Andrew Lyne of the University of Manchester in England and the same team reported the detection of the second pulsar in the system, making it the first binary system with two observed pulsars orbiting each other.

This system is even more remarkable than PSR 1913+16. With a separation of only 800,000 kilometers—about twice the distance between Earth and the

LIGO INTERFEROMETERS, such as this one at the Hanford Observatory in Richland, Wash., should be able to detect the gravitational radiation of colliding neutron stars from a distance of billions of light-years. If these signals arrive at the same time as gamma-ray bursts, a decades-old mystery may be solved.

LIGO

moon—the stars orbit each other in just over two hours. With an orbital velocity of 600 kilometers a second, relativistic effects are more pronounced than in any other known binary. The periastron shift, for example, is as large as 16.9 degrees a year (four times faster than PSR 1913+16). Because the line of sight to Earth passes almost within the binary orbital plane, pulses from the pulsars pass close to each other and we can observe the Shapiro time delay. Indeed, the pulses pass so close that we observe an eclipse. So far Einstein's general theory of relativity has superbly passed the numerous tests that J0307-3039 provides. The effect of gravitational-wave emission on the orbital motion has not yet been observed, but researchers expect to measure it within a year. The predicted spin precession period is around 70 years. Not surprisingly, the discoverers called this system "a rare laboratory for relativistic gravity."

We may also soon have more information about gamma-ray bursts. In February 1997 an Italian-Dutch team, headed by Enrico Costa of the National Research Council in Rome, used the BeppoSAX satellite to pinpoint the source of the burst GRB 970228 (each burst is denoted by the date it is observed). The x-ray telescope revealed an x-ray afterglow that persisted for several days. This was followed by optical and radio afterglows lasting months.

The identification of the exact burst position enabled detailed examinations that were impossible before. In April 1998 Titus Galama of the University of Amsterdam and an international team discovered supernova 1998bw within the error box of GRB 980425. Further evidence for association between a supernova signal, emerging within the light curve of the afterglow of the burst GRB 030329, convincingly confirmed this association.

When a massive star consumes its nuclear fuel, its core collapses, usually forming a neutron star. The stellar envelope that surrounds the core explodes as a supernova. But according to the collapsar model, suggested by Stan Woosley of the University of California at Santa Cruz, some collapsing stellar cores form a black hole. A black hole that is fed by an accretion disk produces a burst in a process that is remarkably similar to the one expected in the merger model described earlier. The essential difference is that the black hole is now within the exploding stellar envelope.

Bursts belong to two distinct classes: long (those lasting more than two seconds) and short (those less than two seconds). So far afterglow has been seen only from long bursts, and it is generally accepted that the bursts are associated with the supernova death of a massive star. With no detectable afterglow, however, little is known about the short bursts, and they remain as mysterious as ever.

The collapsar model cannot account for short bursts. On the other hand, the natural outcome of a neutron star merger is a short burst. But can we confirm this? NASA's Swift satellite, scheduled to launch in the fall of 2004, should help. It is designed to overcome current technical difficulties in detecting afterglow from short bursts, but it could turn out that short bursts do not have a detectable afterglow. In that case, hope is not lost. Neutron star mergers emit unique gravitational-radiation signals. With the international network of gravitational detectors operational, we expect that—if the merger model is correct—these detectors, or at least an upgraded LIGO detector, will eventually find a gravitational-wave merger signal that coincides with a gamma-ray burst. If we can detect the unique gravitational-wave signal of spiraling neutron stars in coincidence with a gamma-ray burst, astrophysicists will have opened a new window on the final stages of stellar evolution, one that no visible-light instruments can hope to match. SA

MORE TO EXPLORE

Was Einstein Right? Putting General Relativity to the Test. Clifford M. Will. Basic Books, 1988.

LIGO: The Laser Interferometer Gravitational-Wave Observatory. Alex Abramovici et al. in *Science,* Vol. 256, pages 325–333; April 17, 1992.

Probing the Gamma-Ray Sky. Kevin Hurley in *Sky and Telescope,* Vol. 84, No. 6, pages 631–636; December 1992.

A PICTURE LIKE THIS could not have been drawn with any confidence a decade ago, because no one had yet figured out what causes gamma-ray bursts—flashes of high-energy radiation that light up the sky a couple of times a day. Now astronomers think of them as the ultimate stellar swan song. A black hole, created by the implosion of a giant star, sucks in debris and sprays out some of it. A series of shock waves emits radiation.

The Brightest Explosions in the Universe

Every time a gamma-ray burst goes off, a black hole is born

By Neil Gehrels, Luigi Piro and Peter J. T. Leonard

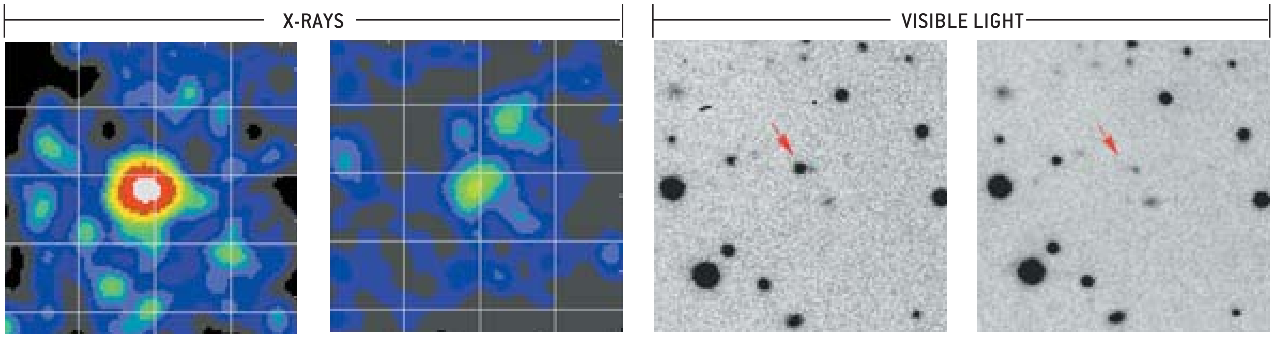
Early in the morning of January 23, 1999, a robotic telescope in New Mexico picked up a faint flash of light in the constellation Corona Borealis. Though just barely visible through binoculars, it turned out to be the most brilliant explosion ever witnessed by humanity. We could see it nine billion light-years away, more than halfway across the observable universe. If the event had instead taken place a few thousand light-years away, it would have been as bright as the midday sun, and it would have dosed Earth with enough radiation to kill off nearly every living thing.

The flash was another of the famous gamma-ray bursts, which in recent decades have been one of astronomy's most intriguing mysteries. The first sighting of a gamma-ray burst (GRB) came on July 2, 1967, from military satellites watching for nuclear tests in space. These cosmic explosions proved to be rather different from the man-made explosions that the satellites

A (VERY) WARM AFTERGLOW

X-RAYS: Eight hours after a burst went off on February 28, 1997, astronomers using the BeppoSAX satellite—including one of the authors (Piro)—saw an x-ray afterglow for the first time. The second image was taken a couple days later, by which time the x-rays had faded by a factor of 20.

VISIBLE LIGHT: A comparably quick reaction by astronomers on La Palma in the Canary Islands allowed the same afterglow to be seen in visible light. Over the next week, the light dimmed to one sixth its original brightness, and as it did so, the surrounding galaxy slowly became apparent.



Eight hours | Three days | 21 hours | Eight days

MARK A. GARLICK (*preceding pages*); ENRICO COSTA *Institute of Space Astrophysics and Cosmic Physics, CNR* AND THE BEPPOSAX TEAM (*left*); PAUL J. GROOT *University of Amsterdam* (*right*)

were designed to detect. For most of the 37 years since then, each new burst had merely heightened the puzzlement. Whenever researchers thought they had the explanation, the evidence sent them back to square one.

The monumental discoveries of the past several years have brought astronomers closer to a definitive answer. Before 1997, most of what we knew about GRBs was based on observations from the Burst and Transient Source Experiment (BATSE) onboard the Compton Gamma Ray Observatory. BATSE revealed that two or three GRBs occur somewhere in the observable universe on a typical day. They outshine everything else in the gamma-ray sky. Although each is unique, the bursts fall into one of two rough categories. Bursts that last less than two seconds are "short," and those that last longer—the majority—are "long." The two categories differ spectroscopically, with short bursts having relatively more high-energy gamma rays than long bursts do. The January 1999 burst emitted gamma rays for a minute and a half.

Arguably the most important result from BATSE concerned the distribution of the bursts. They occur isotropically—that is, they are spread evenly over the entire sky. This finding cast doubt on the prevailing wisdom, which held that bursts came from sources within the Milky Way; if they did, the shape of our galaxy, or Earth's off-center position within it, should have caused them to bunch up in certain areas of the sky. The uniform distribution led most astronomers to conclude that the instruments were picking up some kind of event happening throughout the universe. Unfortunately, gamma rays alone did not provide enough information to settle the question for sure. Researchers would need to detect radiation from the bursts at other wavelengths. Visible light, for example, could reveal the galaxies in which the bursts took place, allowing their distances to be measured. Attempts were made to detect these burst counterparts, but they proved fruitless.

A Burst of Progress

THE FIELD TOOK a leap forward in 1996 with the advent of the x-ray spacecraft BeppoSAX, built and operated by the Italian Space Agency with the participation of the Netherlands Space Agency. BeppoSAX was the first satellite to localize GRBs precisely and to discover their x-ray "afterglows." The afterglow appears when the gamma-ray signal disappears. It persists for days to months, diminishing with time and degrading from x-rays into less potent radiation, including visible light and radio waves. Although BeppoSAX detected afterglows for only long bursts—no counterparts of short bursts have yet been identified—it made follow-up observations possible at last. Given the positional information from BeppoSAX, optical and radio telescopes were able to identify the galaxies in which the GRBs took place. Nearly all lie billions of light-years away, meaning that the bursts must be enormously powerful. Extreme energies, in turn, call for extreme causes, and researchers began to

Overview/*Gamma-Ray Bursts*

- For three decades, the study of gamma-ray bursts was stuck in first gear—astronomers couldn't settle on even a sketchy picture of what sets off these cosmic fireworks.
- Over the past seven years, however, observations have revealed that bursts are the birth throes of black holes. Most of the holes are probably created when a massive star collapses, releasing a pulse of radiation that can be seen billions of light-years away.
- Now the research has shifted into second gear—fleshing out the theory and probing subtle riddles, especially the bursts' incredible diversity.

associate GRBs with the most extreme objects they knew of: black holes.

Among the first GRBs pinpointed by BeppoSAX was GRB970508, so named because it occurred on May 8, 1997. Radio observations of its afterglow provided an essential clue. The glow varied erratically by roughly a factor of two during the first three weeks, after which it stabilized and then began to diminish. The large variations probably had nothing to do with the burst source itself; rather they involved the propagation of the afterglow light through space. Just as Earth's atmosphere causes visible starlight to twinkle, interstellar plasma causes radio waves to scintillate. For this process to be visible, the source must be so small and far away that it appears to us as a mere point. Planets do not twinkle, because, being fairly nearby, they look like disks, not points.

Therefore, if GRB970508 was scintillating at radio wavelengths and then stopped, its source must have grown from a mere point to a discernible disk. "Discernible" in this case means a few light-weeks across. To reach that size, the source must have been expanding at a considerable rate—close to the speed of light.

The BeppoSAX and follow-up observations have transformed astronomers' view of GRBs. The old concept of a sudden release of energy concentrated in a few brief seconds has been discarded. Indeed, even the term "afterglow" is now recognized as misleading: the energy radiated during both phases is comparable. The spectrum of the afterglow is characteristic of electrons moving in a magnetic field at or very close to the speed of light.

GRB990123—the January 1999 burst—was instrumental in demonstrating the immense power of the bursts. If the burst radiated its energy equally in all directions, it must have had a luminosity of a few times 10^{45} watts, which is 10^{19} times as bright as our sun. Although the other well-known type of cosmic cataclysm, a supernova explosion, releases almost as much energy, most of that energy escapes as neutrinos, and the remainder leaks out more gradually than in a GRB. Consequently, the luminosity of a supernova at any given moment is a tiny fraction of that of a GRB. Even quasars, which are famously brilliant, give off only about 10^{40} watts.

If the burst beamed its energy in particular directions rather than in all directions, however, the luminosity estimate would be lower. Evidence for beaming comes from the way the afterglow of GRB990123, among others, dimmed over time. Two days into the burst, the rate of dimming increased suddenly, which would happen naturally if the observed radiation came from a narrow jet of material moving at close to the speed of light. Because of a relativistic effect, the observer sees more and more of the jet as it slows down. At some point, there is no more to be seen, and the apparent brightness begins to fall off more rapidly [*see illustration on next page*]. For GRB990123 and several other bursts, the inferred jet-opening angle is a few degrees. Only if the jet is aimed along our line of sight do we see the burst. This beaming effect reduces the overall energy emitted by the burst approximately in proportion to the square of the jet angle. For example, if the jet subtends 10 degrees, it covers about one 500th of the sky, so the energy requirement goes down by a factor of 500; moreover, for every GRB that is observed, another 499 GRBs go unseen. Even after taking beaming into account, however, the luminosity of GRB990123 was still an impressive 10^{43} watts.

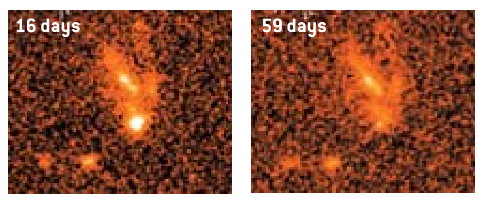
GRB-Supernova Connection

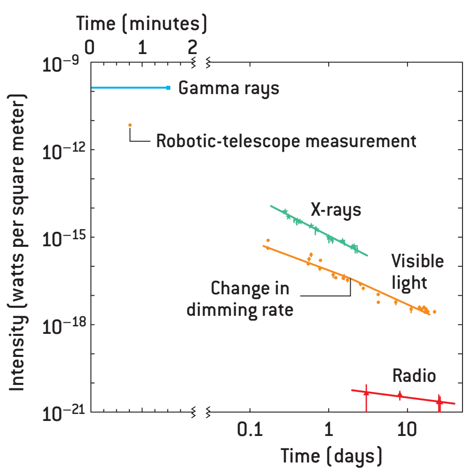
ONE OF THE MOST interesting discoveries has been the connection between GRBs and supernovae. When telescopes went to look at GRB980425, they also found a supernova, designated SN1998bw, that had exploded at about the same time as the burst. The probability of a chance coincidence was one in 10,000. A more firm case is the association of GRB030329 with SN2003dh. This GRB was localized by NASA's second High Energy Transient Explorer satellite (HETE-2), launched in October 2000. The temporally and spatially coincident supernova was discovered via ground-based optical observations, and the broad spectroscopic features of SN2003dh are basically identical to those of SN1998bw.

A link between GRBs and supernovae has also been suggested by the detection of iron in the x-ray spectra of several bursts. Iron atoms are known to be syn-

FADING AWAY

Brightest gamma-ray burst yet recorded went off on January 23, 1999. Telescopes tracked its brightness in gamma rays (*blue in graph*), x-rays (*green*), visible light (*orange*) and radio waves (*red*). At one point, the rate of dimming changed abruptly—a telltale sign that the radiation was coming from narrow jets of high-speed material. About two weeks into the burst, after the visible light had dimmed by a factor of four million, the Hubble Space Telescope took a picture and found a severely distorted galaxy. Such galaxies typically have high rates of star formation. If bursts are the explosions of young stars, they should occur in just such a place.



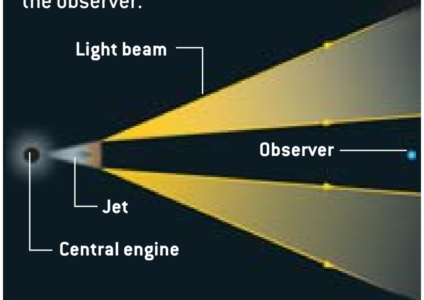


CORNELIA BLIK (*graph*); ANDREW FRUCHTER *Space Telescope Science Institute* AND NASA

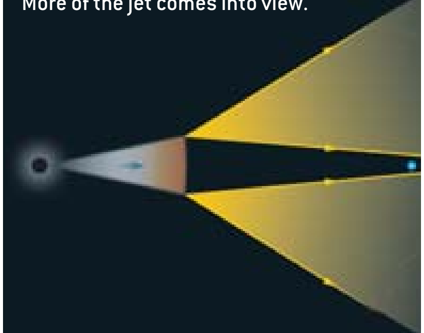
BEAM LINES

Relativity plays tricks on observers' view of jets from gamma-ray bursts.

1 Moving at close to the speed of light, the jet emits light in narrow beams. Some beams bypass the observer.



2 As the jet slows, the beams widen, so fewer of them bypass the observer. More of the jet comes into view.



3 Eventually beams from the edges reach the observer. The entire jet is now visible. Data reveal this transition.

JUAN VELASCO

thesized and dumped into interstellar space by supernovae explosions. If these atoms are stripped of their electrons and later hook up with them again, they give off light at distinctive wavelengths, referred to as emission lines. Early, marginal detections of such lines by BeppoSAX and the Japanese x-ray satellite ASCA in 1997 have been followed by more solid measurements. Notably, NASA's Chandra X-ray Observatory detected iron lines in GRB991216, which yielded a direct distance measurement of the GRB. The figure agreed with the estimated distance of the burst's host galaxy.

Additional observations further support the connection between GRBs and supernovae. An iron-absorption feature appeared in the x-ray spectrum of GRB-990705. In the shell of gas around another burst, GRB011211, the European Space Agency's X-ray Multi-Mirror satellite found evidence of emission lines from silicon, sulfur, argon and other elements commonly released by supernovae.

Although researchers still debate the matter, a growing school of thought holds that the same object can produce, in some cases, both a burst and a supernova. Because GRBs are much rarer than supernovae—every day a couple of GRBs go off somewhere in the universe, as opposed to hundreds of thousands of supernovae—not every supernova can be associated with a burst. But some might be. One version of this idea is that supernovae explosions occasionally squirt out jets of material, leading to a GRB. In most of these cases, astronomers would see either a supernova or a GRB, but not both. If the jets were pointed toward Earth, light from the burst would swamp light from the supernova; if the jets were aimed in another direction, only the supernova would be visible. In some cases, however, the jet would be pointed just slightly away from our line of sight, letting observers see both. This slight misalignment would explain GRB980425.

Whereas this hypothesis supposes that most or all GRBs might be related to supernovae, a slightly different scenario attributes only a subset of GRBs to supernovae. Roughly 90 of the bursts seen by BATSE form a distinct class of their own, defined by ultralow luminosities and long spectral lags, meaning that the high- and low-energy gamma-ray pulses arrive several seconds apart. No one knows why the pulses are out of sync. But whatever the reason, these strange GRBs occur at the same rate as a certain type of supernova, called Type Ib/c, which occurs when the core of a massive star implodes.

Great Balls of Fire

EVEN LEAVING ASIDE the question of how the energy in GRBs might be generated, their sheer brilliance poses a paradox. Rapid brightness variations suggest that the emission originates in a small region: a luminosity of 10^{19} suns comes from a volume the size of one sun. With so much radiation emanating from such a compact space, the photons must be so densely packed that they should interact and prevent one another from escaping. The situation is like a crowd of people who are running for the exit in such a panic that that nobody can get out. But if the gamma rays are unable to escape, how can we be seeing GRBs?

The resolution of this conundrum, developed over the past several years, is that the gammas are not emitted immediately. Instead the initial energy release of the explosion is stored in the kinetic energy of a shell of particles—a fireball—moving at close to the speed of light. The particles include photons as well as electrons and their antimatter counterpart, positrons. This fireball expands to a diameter of 10 billion to 100 billion kilometers, by which point the photon density has dropped enough for the gamma rays to escape unhindered. The fireball then converts some of its kinetic energy into electromagnetic radiation, yielding a GRB.

The initial gamma-ray emission is

THE AUTHORS

NEIL GEHRELS, LUIGI PIRO and PETER J. T. LEONARD bring both observation and theory to the study of gamma-ray bursts. Gehrels and Piro are primarily observers—the lead scientists, respectively, of the Compton Gamma Ray Observatory and the BeppoSAX satellite. Leonard is a theorist, and like most theorists, he used to think it unlikely that the bursts were bright enough to be seen across the vastness of intergalactic space. "I have to admit that the GRBs really had me fooled," he says. Gehrels is head of the Gamma Ray, Cosmic Ray and Gravitational Wave Astrophysics Branch of the Laboratory for High Energy Astrophysics at the NASA Goddard Space Flight Center. Piro is a member of the Institute of Space Astrophysics and Cosmic Physics of the CNR in Rome. Leonard works for Science Systems and Applications, Inc., in support of missions at Goddard.

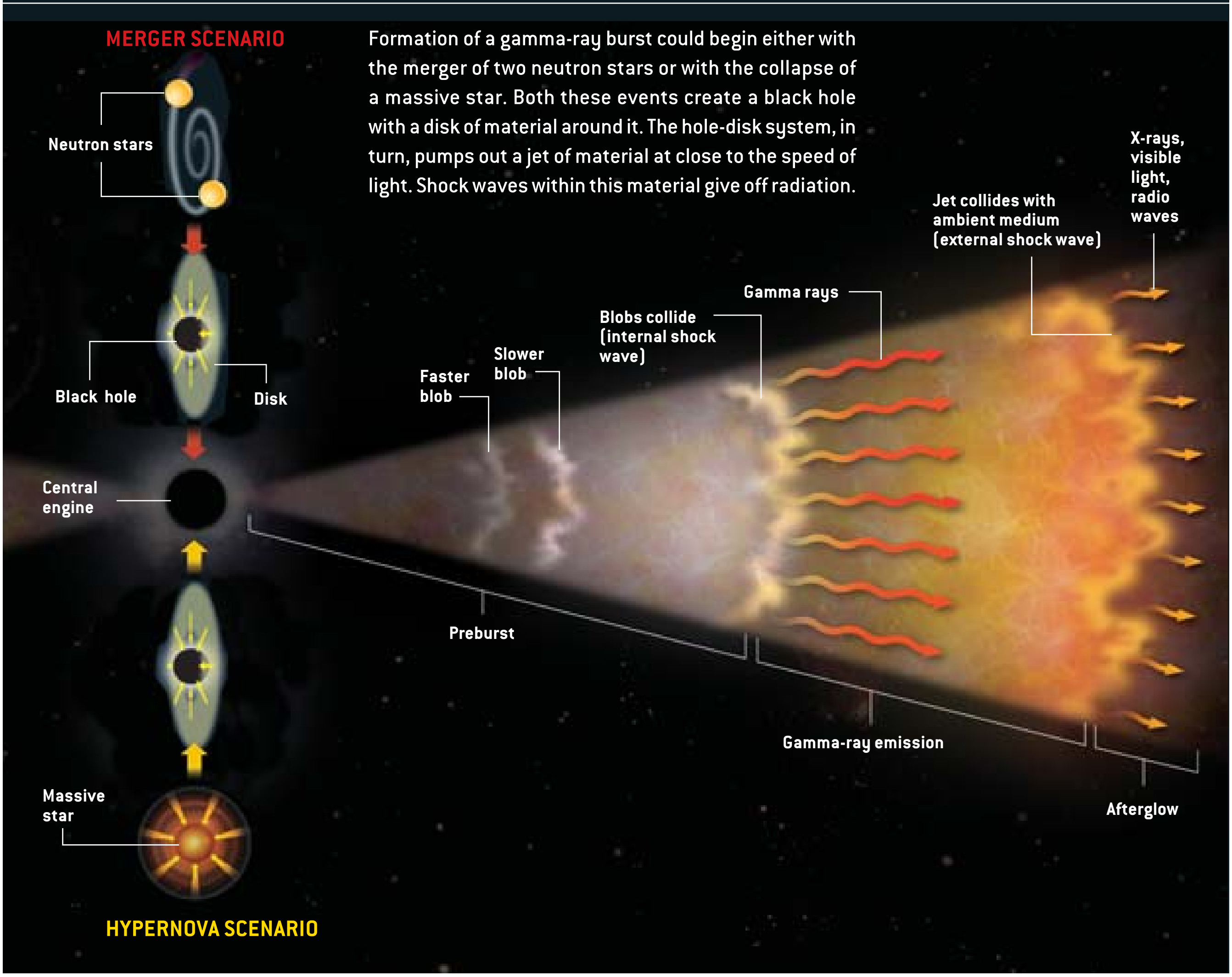
most likely the result of internal shock waves within the expanding fireball. Those shocks are set up when faster blobs in the expanding material overtake slower blobs. Because the fireball is expanding so close to the speed of light, the timescale witnessed by an external observer is vastly compressed, according to the principles of relativity. So the observer sees a burst of gamma rays that lasts only a few seconds, even if it took a day to produce. The fireball continues to expand, and eventually it encounters and sweeps up surrounding gas. Another shock wave forms, this time at the boundary between the fireball and the external medium, and persists as the fireball slows down. This external shock nicely accounts for the GRB afterglow emission and the gradual degradation of this emission from gamma rays to x-rays to visible light and, finally, to radio waves.

Although the fireball can transform the explosive energy into the observed radiation, what generates the energy to begin with? That is a separate problem, and astronomers have yet to reach a consensus. One family of models, referred to as hypernovae or collapsars, involves stars born with masses greater than about 20 to 30 times that of our sun. Simulations show that the central core of such a star eventually collapses to form a rapidly rotating black hole encircled by a disk of leftover material.

A second family of models invokes binary systems that consist of two compact objects, such as a pair of neutron stars (which are ultradense stellar corpses) or a neutron star paired with a black hole. The two objects spiral toward each other and merge into one. Just as in the hypernova scenario, the result is the formation of a single black hole surrounded by a disk.

Many celestial phenomena involve a hole-disk combination. What distinguishes this particular type of system is the sheer mass of the disk (which allows for a gargantuan release of energy) and the lack of a companion star to resupply the disk (which means that the energy release is a one-shot event). The black hole and disk have two large reservoirs of energy: the gravitational energy of the disk and the rotational energy of the hole. Exactly how these would be converted into gamma radiation is not fully understood. It is

BURSTING OUT



Formation of a gamma-ray burst could begin either with the merger of two neutron stars or with the collapse of a massive star. Both these events create a black hole with a disk of material around it. The hole-disk system, in turn, pumps out a jet of material at close to the speed of light. Shock waves within this material give off radiation.

JUAN VELASCO

possible that a magnetic field, 10^{15} times more intense than Earth's magnetic field, builds up during the formation of the disk. In so doing, it heats the disk to such high temperatures that it unleashes a fireball of gamma rays and plasma. The fireball is funneled into a pair of narrow jets that flow out along the rotational axis.

Because the GRB emission is equally well explained by both hypernovae and compact-object mergers, some other qualities of the bursts are needed to decide between these two scenarios. The association of GRBs with supernovae, for example, is a point in favor of hypernovae, which, after all, are essentially large supernovae. Furthermore, GRBs are usually found just where hypernovae would be expected to occur—namely, in areas of recent star formation within galaxies. A massive star blows up fairly soon (a few million years) after it is born, so its deathbed is close to its birthplace. In contrast, compact-star coalescence takes much longer (billions of years), and in the meantime the objects will drift all over the galaxy. If compact objects were the culprit, GRBs should not occur preferentially in star-forming regions.

Although hypernovae probably explain most GRBs, compact-star coalescence could still account for the poorly understood short-duration GRBs. Moreover, additional models for GRBs are still in the running. One scenario produces the fireball via the extraction of energy from an electrically charged black hole. This model suggests that both the immediate and the afterglow emissions are consequences of the fireball sweeping up the external medium. Astronomers have come a long way in understanding gamma-ray bursts, but they still do not know precisely what causes these explosions, and they know little about the rich variety and subclasses of bursts.

All these recent findings have shown that the field has the potential for answering some of the most fundamental questions in astronomy: How do stars end their lives? How and where are black holes formed? What is the nature of jet outflows from collapsed objects?

Blasts from the Past

ONE OUTSTANDING question concerns the dark, or "ghost," GRBs. Of the roughly 80 GRBs that have been localized and studied at wavelengths other than gamma rays, about 90 percent have been seen in x-rays. In contrast, only about 50 percent have been seen in visible light. Why do some bursts fail to shine in visible light?

One explanation is that these GRBs lie in regions of star formation, which tend to be filled with dust. Dust would block visible light but not x-rays. Another intriguing possibility is that the ghosts are GRBs that happen to be very far away. The relevant wavelengths of light produced by the burst would be absorbed by intergalactic gas. To test this hypothesis, measurement of the distance via x-ray spectra will be crucial. A third possibility is that ghosts are optically faint by nature. Currently the evidence favors the dust explanation. High-sensitivity optical and radio investigations have identified the probable host galaxies of two dark GRBs, and each lies at a fairly moderate distance.

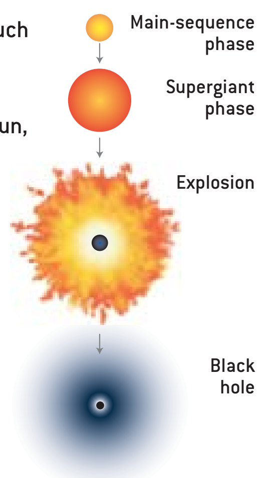
Another mystery concerns a class of events known as the x-ray-rich GRBs, or simply the x-ray flashes. Discovered by BeppoSAX, later confirmed by reanalysis of BATSE data and currently observed by HETE-2, these bursts represent 20 to 30 percent of GRBs. They give off more x-radiation than gamma radiation; indeed, extreme cases exhibit no detectable gamma radiation at all.

One explanation is that the fireball is loaded with a relatively large amount of baryonic matter such as protons, making for a "dirty fireball." These particles increase the inertia of the fireball, so that it moves more slowly and is less able to boost photons into the gamma-ray range. Alternatively, the x-ray flashes could be typical GRBs with jets that are pointing just out of our view, so that only

THE DESTINIES OF MASSIVE STARS

STARS SPEND MOST OF THEIR LIVES in the relatively unexciting main-sequence evolutionary phase, during which they casually convert hydrogen into helium in their cores via nuclear fusion. Our sun is in this phase. According to basic stellar theory, stars more massive than the sun shine more brightly and burn their fuel more quickly. A star 20 times as massive as the sun can keep going for only a thousandth as long.

As the hydrogen in the core of a star runs out, the core contracts, heats up and starts to fuse heavier elements, such as helium, oxygen and carbon. The star thus evolves into a giant and then, if sufficiently massive, a supergiant star. If the initial mass of the star is at least eight times that of the sun, the star successively fuses heavier and heavier elements in its interior until it produces iron. Iron fusion does not release energy—on the contrary, it uses up energy. So the star suddenly finds itself without any useful fuel.



The result is a sudden and catastrophic collapse. The core is thought to turn into a neutron star, a stellar corpse that packs at least 40 percent more mass than the sun into a ball with a radius of only 10 kilometers. The remainder of the star is violently ejected into space in a powerful supernova explosion.

There is a limit to how massive a neutron star can be—namely, two to three times as massive as the sun. If it is any heavier, theory predicts it will collapse into a black hole. It can be pushed over the line if enough matter falls onto it. It is also possible that a black hole can be formed directly during the collapse. Stars born with masses exceeding roughly 20 solar masses may be destined to become black holes. The creation of these holes provides a natural explanation for gamma-ray bursts. *—N.G., L.P. and P.J.T.L.*

CORNELIA BLIK

Classes of Gamma-Ray Bursts

BURST CLASS (SUBCLASS)	PERCENTAGE OF ALL BURSTS	TYPICAL DURATION OF INITIAL EMISSION (SECONDS)	INITIAL GAMMA-RAY EMISSION	AFTERGLOW X-RAY EMISSION	AFTERGLOW VISIBLE EMISSION	HYPOTHETICAL CENTRAL ENGINE	EXPLANATION FOR PECULIAR PROPERTIES
Long (normal)	25	20	✓	✓	✓	Energetic explosion of massive star	Not applicable
Long (ghosts or dark)	30	20	✓	✓	✗	Energetic explosion of massive star	Extremely distant, obscured by dust, or intrinsically faint
Long (x-ray-rich or x-ray flashes)	25	30	✗	✓	✗	Energetic explosion of massive star	Extremely distant or weighed down by extra particles
Short	20	0.3	✓	?	?	Merger of pair of compact objects	Does not occur in a star-forming region, so ambient gas is less dense and external shocks are weaker

the less collimated and less energetic x-rays reach us. Finally, most flashes might come from very distant galaxies—even more distant than the galaxies proposed to explain the ghost GRBs. Cosmic expansion would then shift the gamma rays into the x-ray range, and intergalactic gas would block any visible afterglow. In fact, most of these x-ray flashes do not have a detectable visible-light counterpart, a finding consistent with this scenario. If either x-ray flashes or ghost GRBs are located in extremely distant galaxies, they could illuminate an era in cosmic history that is otherwise almost invisible.

The next step for GRB astronomy is to flesh out the data on burst, afterglow and host-galaxy characteristics. Observers need to measure hundreds of bursts of all varieties: long and short, bright and faint, heavy in gamma rays or x-rays, bursts with visible-light afterglows and those without. Currently astronomers are obtaining burst positions from HETE-2 and the Interplanetary Network, a series of small gamma-ray detectors piggybacking on planetary spacecraft. The Swift mission, scheduled for launch in the fall of 2004, will offer multiwavelength observations of hundreds of GRBs and their afterglows, making automatic onboard x-ray and optical observations. A rapid response will determine whether the GRB has an x-ray or visible afterglow. The mission will be sensitive to short-duration bursts, which have barely been studied thus far.

Another goal is to probe extreme gamma-ray energies. GRB940217, for example, emitted high-energy gamma rays for more than an hour after the burst, as observed by the Energetic Gamma Ray Experiment Telescope instrument on the Compton Gamma Ray Observatory. Astronomers do not understand how such extensive and energetic afterglows can be produced. The Italian Space Agency's AGILE satellite, scheduled for launch in 2005, will observe GRBs at these high energies. The supersensitive Gamma-Ray Large Area Space Telescope mission, expected to launch in 2007, will also be key for studying this puzzling phenomenon.

Other missions, though not designed solely for GRB discovery, will also contribute. The International Gamma-Ray Astrophysics Laboratory, launched on October 17, 2002, is detecting about 10 GRBs a year. The Energetic X-ray Imaging Survey Telescope, to launch a decade from now, will have a sensitive gamma-ray instrument capable of detecting thousands of GRBs.

The field has just experienced a series of breakthrough years, with the discovery that GRBs are immense explosions occurring throughout the universe. Bursts provide us with an exciting opportunity to study new regimes of physics and to learn what the universe was like at the earliest epochs of star formation. Space- and ground-based observations over the coming years should allow us to uncover the detailed nature of these most remarkable beasts. Astronomers can no longer talk of bursts as utter mysteries, but that does not mean the puzzle is completely solved. SA

MORE TO EXPLORE

Observation of X-ray Lines from a Gamma-Ray Burst (GRB991216): Evidence of Moving Ejecta from the Progenitor. Luigi Piro et al. in *Science,* Vol. 290, pages 955–958; November 3, 2000. Preprint available at **arxiv.org/abs/astro-ph/0011337**

Gamma-Ray Bursts: Accumulating Afterglow Implications, Progenitor Clues, and Prospects. Peter Mészáros in *Science,* Vol. 291, pages 79–84; January 5, 2001. **arxiv.org/abs/astro-ph/0102255**

Blinded by the Light. Stan Woosley in *Nature,* Vol. 414, pages 853–854; December 20, 2001.

The Biggest Bangs: The Mystery of Gamma-Ray Bursts, the Most Violent Explosions in the Universe. Jonathan I. Katz. Oxford University Press, 2002.

Flash! The Hunt for the Biggest Explosions in the Universe. Govert Schilling. Cambridge University Press, 2002.